Culture, Information, and Screening Discrimination

Bradford Cornell; Ivo Welch

Culture, Information, and Screening Discrimination

Bradford Cornell and Ivo Welch

University of California, Los Angeles

We show that discrimination can occur even when it is common knowledge that underlying group characteristics do not differ and when employers do not prefer same-group candidates. When employers can judge job applicants' unknown qualities better when candidates belong to the same group and hire the best prospect from a large pool of applicants, the top applicant is likely to have the same background as the employer. The model has policy, empirical, and experimental implications. For example, the model predicts that "screening discrimination" is more likely to occur and persist in sectors in which underlying quality is important but difficult to observe, there are numerous applicants, interviewing (screening) is relatively cheap, and applicants have to acquire job-specific skills.

I. Introduction

This paper develops a simple model that can explain how discrimination, such as "racism"—defined as the tendency to hire or fraternize with people whose cultural backgrounds are similar to one's own—can develop spontaneously even when all individuals are rational, have no preference for people of their own type, and believe, correctly, that there are no average differences between people of various types. These assumptions distinguish the model from the work of Becker (1957), who assumes that individuals have racial preferences, and Akerlof (1984), who shows how racism, once it exists, can

be self-perpetuating but does not explain how it can arise spontaneously.

Our model is based on two premises. The first premise is that people are faced with the problem of screening other people in virtually every aspect of life. This sorting problem involves not only major decisions such as choosing friends, accepting members in social organizations, and hiring, but everyday decisions such as deciding whether the person walking toward you on the sidewalk means well or ill. Although the model applies to many kinds of human interaction, we concentrate on the labor market because discrimination in hiring is a particularly critical issue. We assume that by the time people enter the job market, they possess certain qualities that cannot be easily disclosed to employers. Among these qualities are not only measurable skills but also intangible characteristics that may be critically important in qualifying a candidate for a position (see, e.g., Arrow 1972; McCall 1972). For example, Arrow states that one difficulty employers face is determining which employees possess what he refers to as unobservable habits of action and thought that favor good performance in skilled jobs, including steadiness, punctuality, responsiveness, and initiative.[1] A related intangible characteristic that makes a person more productive is the willingness of an employee to "internalize a job" in the sense that he or she will put forth the effort to understand and complete tasks that cannot be explicitly delineated. Because these and related characteristics cannot be observed directly, employers must rely on indirect assessment procedures that measure them with error when making hiring decisions. To simplify our presentation, all these characteristics are subsumed in one unobservable variable, referred to as "character" or "quality."

The second premise is that people can distinguish between high- and low-character individuals more accurately when the people being sorted are of a similar cultural type. Cultural type here is interpreted broadly to include groups defined by language, religious belief, ethnic background, race, sex, sexual preference, neighborhood upbringing, schooling, or membership in social organizations. People who grow up under comparable circumstances will have a common framework for assessing each other's personal history.[2] Moreover, Scheflen (1971) and Mehrabian (1981) present evidence to show that people are able to interpret a host of culturally intermediated signals such as dress, manners and mannerisms, gestures, and style of speaking

---

[1] The contribution of our paper is that it extends Arrow's approach by showing how culture and information interact when people are being sorted.

[2] Personal history encompasses more than an applicant's employment record. It can include such information as memberships in social organizations, civic awards, personal interests, and other general indicators of character.

as well as spoken words. Furthermore, the ability to interpret these nonverbal cues has been found to have a significant cultural element (see, e.g., Ekman and Friesen 1975; Archer and Akert 1977; Brown 1986). For example, Japanese businessmen follow elaborate rules of personal interaction the observation of which is considered to be an indicator of character. However, this form of communication is difficult to interpret unless the interpreter is schooled in Japanese business culture. Lang (1986) provides an extensive survey of evidence about language barriers, where language is interpreted broadly to mean any form of communication, and their effect on segregation in the workplace. Similarly, it is likely to be easier to interpret someone else's personal references and see through possible deception when the recommendation comes from someone familiar or at least someone with a similar background. Consequently, the ability to assess an individual's personal history is enhanced when one is evaluating a person with a familiar cultural background. In sum, the critical assumption on which our analysis turns is that for a given assessment procedure the measurement error for a given evaluator is smaller when the people being considered have a similar cultural background.

As an illustration of the impact of culture as a key to communication and evaluation, Monroe Price, dean of Yeshiva University's Benjamin N. Cardozo School of Law in New York, cites a 30-year-old study by Yale University that captures the intuition of our model: "Lawyers responding 20 years ago, as some would today, argued that there is a natural tendency to discriminate in favor of hiring new associates coming from backgrounds similar to their own, 'not because they had any desire to do so, but because they tend to feel they knew such applicants better after a fleeting acquaintance than they did those from dissimilar backgrounds.' Virtually all firms said then they made personality a factor second only to school grades" (1984, p. 13).

Given our two assumptions—that screening is important and that it is easier for individuals to screen people of similar background—the model developed here shows that people tend to hire others of their own type even when they have no innate preference for similar people and even when they (correctly) believe that the distribution of quality among people of their own background is no different from the distribution of quality among people of other backgrounds. The intuition of the model can be conveyed by a simple example. Consider an employer faced with selecting the best applicant from a large group of applicants. Assume that his ability to judge candidates of the same background allows him to group such applicants but not applicants of another background into "better than average" and

"worse than average" categories, with each of them having roughly equal probabilities. If there is only one applicant from each background, the probability that the best applicant has the same background is 50 percent because the candidate from the same background is equally likely to be categorized as better or worse, whereas the candidate from the other background is always seen as average. However, with four applicants, two from each background, the probability that the best applicant has the opposite background of the employer equals the probability that both of the same-background candidates are judged to be worse than average. Consequently, the probability that the top candidate has a similar background increases from 50 to 75 percent. As the number of applicants rises further, the probability that the top candidate is of the same type rises rapidly toward one.

The cause for this "screening discrimination" is that the identical unconditional distribution of high- and low-quality individuals in the population is not important; employers make judgments based on their conditional assessments, and the conditional probability distribution has a greater variance the more accurate information an employer acquires. Therefore, when a single best candidate is chosen from a set of applicants, it is more likely that the top applicant comes from the set of applicants with the widest distribution. Models of such "rank-order tournaments" are developed in Lazear and Rosen (1981) and Rosen (1981). These authors also discuss the circumstances under which rank-order tournaments are more efficient than current marginal product compensation.

Our model is similar to that of Aigner and Cain (1977), extending Phelps (1972), who argue that lower wages can result not only from lower mean productivity but also from higher variance in inferred productivity. They find that less accurate testing or higher intrinsic quality variation can lead to lower wages for high-scoring blacks (relative to whites) and (relatively) higher wages for low-scoring blacks. We extend this model by arguing that it is only the high-scoring individuals that are rewarded in many tournament situations and by identifying why lower accuracy for blacks is not an "unusual result" (Aigner and Cain 1977, p. 179). Our model is also closely related to the model of Lang (1986), in which language barriers make successful cooperation among individuals from different groups costly. Lang's transaction cost argument can explain "natural" segregation among groups (similar to that in discrimination-taste-based models, e.g., in Becker [1957]), continued different wages across groups, and extra compensation for "bilingual" individuals. In contrast, we argue that segregation is a common result in a "filtering" situation in which an employer selects one of a large number of candidates, as in a

tournament.[3] In addition, segregation is not intrinsically advantageous and does not arise in our model when a larger number of applicants are hired. Finally, the literature on "matching" (e.g., finite and disjoint sets of individuals, say men and women) also addresses the same basic problem. As described by Roth and Sotomayor (1990), concepts of rational and equilibrium matching can be developed under sufficiently restrictive assumptions regarding the preferences of the individuals and the information they possess. The typical assumption is common knowledge. While Roth and Sotomayor relax this common-knowledge assumption somewhat, the spirit of their work remains quite different from ours. Whereas they make restrictive informational assumptions to explore robustness, we explore the implications of a specific type of heterogeneous information. In our model, not only is information incomplete, but individuals can be strategically deceptive. In fact, the key to our analysis is the assumption that the ability to detect deception is culturally limited. Whereas our setting limits the nature of the models that can be examined, it still produces challenging empirical predictions.[4]

Section II*A* proves that the effect of providing employers with more information is equivalent to increasing the variance of the inferred qualities. Although this increased variance alone is not sufficient to guarantee an increased probability of hiring same-background candidates when there are only a few applicants, with sufficiently many applicants the probability that an applicant of the same background is hired approaches one. Moreover, even in small samples, the increased variance tends to induce screening discrimination.

In Section II*B*, the model is generalized to an environment in which successful applicants in one generation become employers (interviewers) in the following generation. In a simple example, with as few as 40 candidates for five jobs, we show that it takes more than a million generations before the probability that an applicant from another background is hired exceeds 50 percent. An alternative generalization of the model is developed in Section II*C*. In this scenario,

---

[3] In addition, Lang focuses on situations in which bilingual skills can be acquired by minority members. Yet if an other-group candidate who can mimic the language of a same-group candidate is also able to mimic high quality, an employer may actually discount any information obtained in an interview from such a "bilingual" employee.

[4] More specifically, our model is not a full two-sided matching equilibrium of the sort analyzed in Roth and Sotomayor (1990). Our simpler partial-equilibrium model focuses on the issue at hand. Nevertheless, we speculate that our model and results can be made consistent with (i.e., embedded in) a full two-sided matching equilibrium. This would require, e.g., adding additional employers to the basic model. Section II*C* is close to the spirit of a full-equilibrium model in that it assumes the presence of a flexible outside market for low-wage labor.

applicants need to invest in job-specific skills before being screened for a job in a high-wage sector, which happens to be dominated by employers of one group. Under certain parameter restrictions, even in a long-run equilibrium, discrimination is perfectly persistent; that is, all applicants of other background work only in low-wage sector jobs. In Section II*D*, we argue that screening costs, like required early job-specific training, can lead to an equilibrium in which employers anticipate their higher likelihood of hiring candidates of the same background and thus optimally choose to spend their screening resources only on candidates of the same background. Section II*E* analyzes some reinforcing effects by which screening discrimination can eventually be internalized by both employers and employees and thereby evolves into discrimination in the Becker sense. Furthermore, this process can lead to *cascades* in which some employers, relying on the information that no other employers hire individuals of opposite background, optimally decide to make the same decision. Finally, Section III discusses the policy implications of our theory and suggests both experimentally and empirically testable hypotheses.

II. Screening Discrimination

Discrimination can occur in both hiring and wage-setting procedures, and the two may either substitute for or complement one another. Although the model developed here could be framed in terms of wage discrimination, we concentrate on hiring discrimination. We assume, therefore, that employers have to post a fixed wage for job positions and do not condition wages on the component of inferred quality that correlates with group background (perhaps because of antidiscrimination laws).

A. A Basic Model

To illustrate the workings of the model, we begin with an example in which employers hire a single employee from a pool of applicants purely on the basis of an inference of candidates' abilities. Assume that the underlying character quality, Q, of each candidate is a one-dimensional parameter, distributed uniformly from zero to one, that is, $f_Q(\cdot) = I_{[0, 1]}$, independent of group. Other employee characteristics and the incentive and monitoring system used by the employer are held constant so that employee output depends only on Q. Employers attempt to maximize the expected product of the prospective employee.

Consider first the employer's inference of the quality of a single applicant. An employer with no information would infer the expected

quality of every candidate to be ½. Screening allows an employer to gain additional information. We model the screening process as providing n conditionally independent signals s about each prospective employee, with each signal $s \in \{L, H\}$ (low, high). Signal and quality are assumed to be correlated in that the probability of observing the H (high) signal is Q. For example, if the underlying candidate quality is 1, the employer will receive an H signal with probability one; if the quality is 0, the employer will receive an H signal with probability zero; if the quality is ½, the employer will receive either an H signal or an L signal with equal probability. Given this setting with a uniform prior distribution and binomial signals, DeGroot (1970) shows that the ex ante probability of observing k signals of type H (or L) is $1/(n + 1)$ independent of k. The inferred quality given k signals of type H from n drawings is $(k + 1)/(n + 2)$.

With this information structure, if an employer observes one signal, he or she will infer the expected quality of the candidate, $\hat{Q}$, to be either ⅔ or ⅓, each with probability ½. Therefore, the ex ante variance of the employer's inferred (expected) quality of the candidate is $\frac{1}{2} \cdot (\frac{1}{3} - \frac{1}{2})^2 + \frac{1}{2} \cdot (\frac{2}{3} - \frac{1}{2})^2 = \frac{1}{36}$. More generally, let $\hat{Q} \equiv E(Q|\tilde{k}, n)$ be the quality inferred by the employer given k of n H signals, where the tilde over the k denotes the fact that the number of H signals is unknown to the outside observer. Therefore, the variance of the employer's inferred qualities of applicants is

$$\mathrm{var}_k(\hat{\tilde{Q}}) = \frac{n}{12(n + 2)}. \tag{1}$$

As the number of signals (n) increases (i.e., if the employer can screen better), the variance of inferred qualities increases. Signals have the effect of pulling the posterior beliefs away from the prior mean toward the true value. When the number of signals approaches infinity, the employer can infer the quality of the employee perfectly, and the distribution of inferred qualities approaches the underlying quality distribution itself (here the uniform distribution). The intuition that this example brings out is that the effect of providing employers with more information is equivalent to increasing the heterogeneity of the candidates' qualities. This conclusion is not an artifact of the uniform distribution in that it generalizes to all distributions with finite means and variance as stated in proposition 1.

PROPOSITION 1. Giving an employer more information increases the variance of the employer's inferred quality of one candidate, $\hat{Q}$.[5]

---

[5] It is important to emphasize that *less variance* in the ability of employers to judge has the effect of *more variance* in the distribution of the inferred qualities of applicants. The two must not be confused.

Proof. Let Q be the employee's unknown quality and $\hat{Q}$ be the employer's inferred expected quality of the employee. Let X_1 be a variable that summarizes the information set of employers *before* screening (available for both other-group and same-group applicants). Prior to screening, the employer infers the expected quality by computing $\hat{Q}|X_1$. This mean is updated only for same-group applicants by the information obtained from screening, denoted X_2. In the context of our model, other-group and same-group employees are alike, except for an additional piece of information that allows an employer to further update his inferred quality, only for same-group applicants. In other words, $E(Q|X_1, X_2) \neq E(Q|X_1)$ on a nonzero measure set. Since the realizations of the signals, X_1 and X_2, are not known, it is to be shown that the posterior distribution of inferred qualities, $\hat{Q}$, is wider when X_2 is available:

$$\operatorname{var}[E_{X_1, X_2}(Q|X_1, X_2)] > \operatorname{var}[E_{X_1}(Q|X_1)].$$

When variances are expanded, this is true if and only if

$$E[E_{X_1, X_2}(Q|X_1, X_2)^2] > E[E_{X_1}(Q|X_1)^2].$$

The latter term can be expanded to

$$E[E_{X_1}(Q|X_1)^2] = E\{E_{X_1}[E_{X_1, X_2}(Q|X_1, X_2)|X_1]^2\}.$$

Jensen's inequality now proves the proposition because

$$E\{E_{X_1}[E_{X_1, X_2}(Q|X_1, X_2)|X_1]^2\} < E\{E_{X_1}[E_{X_1, X_2}(Q|X_1, X_2)^2|X_1]\},$$

and the latter term is equal to $E\{E[Q|X_1, X_2]^2\}$.[6] Q.E.D.

To see the relevance of proposition 1, assume that the pool of applicants contains two types of people, "same-group" and "other-group." Assume that each candidate can be costlessly interviewed, so that the employer can screen every applicant and hire the applicant considered to be the best. Proposition 2 demonstrates that under these circumstances, with a fixed, positive proportion of same-group applicants in the applicant pool, employers are more likely to hire a same-group applicant when there is a large pool of same-group applicants to draw from. The intuition behind the proposition is straightforward. The best candidate is likely to emerge from the pool with the greatest variance of inferred qualities. The inferred variance is an increasing function of the informativeness of the signals. Because the signals are more informative for the same-group pool, the winning candidate is likely to emerge from that pool.

---

[6] By induction, the argument holds if X_2 represents a vector of information (obtained by screening).

PROPOSITION 2. Assume that signals obtained by screening are bounded above (i.e., there is a maximum inferred quality for both same-group and other-group candidates). In addition, assume that, in the screening process, employers observe an additional set of signals for same-group applicants, above and beyond that observed for other-group candidates, and that this signal is informative for all applicants. For a large enough pool of applicants, when the ratio of same-group and other-group applicants is held constant, the probability that the employer hires a same-group applicant is arbitrarily close to one.

Proof. Let $\hat{Q}^1_{max}$ be the maximum value of the employer's inferred quality ($\hat{Q}|X_1$) without additional information obtained by screening. Let $\hat{Q}^2_{max}$ be the maximum value of the employer's inferred quality ($\hat{Q}|X_1, X_2$), that is, with additional information obtained by screening. (The boundedness assumption assures that both $\hat{Q}^1_{max}$ and $\hat{Q}^2_{max}$ are finite numbers.) Because $\hat{Q}^1_{max}$ is an unbiased forecast, incoming screening information must not change the expected inferred quality $\hat{Q}^1_{max}$ *on average.* Consequently, for same-group candidates inferred to be of quality $\hat{Q}^1_{max}$, the additional (assumed to be informative) signal X_2 must sometimes decrease and sometimes increase the inferred quality relative to $\hat{Q}^1_{max}$. Therefore, the maximum inferred quality for same-group applicants exceeds that of other-group applicants: $\hat{Q}^2_{max} > \hat{Q}^1_{max}$.

Let p be the unconditional probability mass that a same-group quality applicant is judged to be of quality higher than $\hat{Q}^1_{max}$. The probability that at least one of n same-group candidates receives a high score unachievable by other-group candidates is $1 - (1 - p)^n$, which converges to one as n approaches infinity. Q.E.D.

This proposition can be illustrated with our earlier simple binomial example. Assume that other-group employees receive one signal and same-group employees receive one additional binomial signal. Assume further that there are 99 other-group applicants for each same-group applicant. With 1,000 applicants, there are 10 same-group applicants who receive an additional binomial signal. On average, a two-signal candidate receives HH with probability $p = 33$ percent. The probability that the best candidate is same-group is thus on the order of $1 - 0.67^{10} = 98$ percent. Doubling the number of applicants doubles the number of same-group applicants likely to receive the additional signal, pushing the probability that a same-group candidate is selected above 99.9 percent. Note that the two essential ingredients are the asymptotically increasing absolute number of same-group candidates and the (positive) probability that a same-group candidate is inferred to be of better quality than $\hat{Q}^1_{max}$. The number of other-group candidates is irrelevant.

For small applicant pools, the probability P that a same-group applicant is hired is computed as

$$P = \text{prob(hire same group)}$$

$$= \text{prob(EV of best same-group candidate}$$

$$> \text{EV of best other-group candidate)}$$

$$= \int_c \text{prob(EV of best same-group candidate} = c$$

$$\text{and EV of best other-group candidate} < c) \, dc$$

$$= \int_c \text{prob(EV of best same-group candidate} = c)$$

$$\cdot \text{prob(EV of best other-group candidate} < c) \, dc,$$

where EV denotes expected value. For example, reconsider the example above (uniform priors, binary signals) with two same-group and two other-group applicants. Furthermore, assume that employers' contact provides them with one signal on both same-group and other-group applicants, and their screening ability provides them with a second signal only on same-group applicants. The employer may thus infer the quality of each other-group applicant to be either $1/3$ or $2/3$, and, for two applicants, the probability that at least one of the two applicants is inferred to be of quality $2/3$ is 75 percent. For same-group applicants, the employer may infer the quality to be either $1/4$, $1/2$, or $3/4$, each with probability $1/3$. The top same-group applicant is inferred to be of quality $1/4$ with probability $(1/3)^2 = 1/9$, quality $1/2$ with probability $1/3$, and quality $3/4$ with probability $5/9$. Consequently, the probability that the employer chooses a same-group applicant is $1/9 \cdot 0\% + 1/3 \cdot 25\% + 5/9 \cdot 100\% = 23/36 \approx 64$ percent.

In more general cases, however, the computation of this probability involves solving a difficult, highly nonlinear, equation, especially if signals are discrete. The Appendix describes the necessary computations for two examples. First, it computes the probability under the simple distributional assumptions made above to illustrate the inferred quality variance increase (uniform prior, binary signals). Second, it computes an example in which the distributional assumptions are more general. The underlying distribution of qualities is assumed to be distributed unit normal, and the employer observes a (summarized) normally distributed signal with the mean of the (true) quality of each applicant.[7] The information of employers (the signal) has

[7] Under the central limit theorem, the normality assumption can be justified when both the qualities and the signals obtained are the sum of many independent observations.

variance σ_s^2 for same-group candidates and σ_o^2 for other-group candidates. The employer can screen same-group candidates better than other-group candidates, so $\sigma_s^2 < \sigma_o^2$. The Appendix further proves that the probability of hiring a same-group applicant strictly exceeds $\frac{1}{2}$ when there is more than one candidate each and that this probability decreases in σ_s^2 and increases in σ_o^2. Figure 1 plots the probability that a same-group candidate is hired, with an equal number of candidates of each type and a fixed ability to screen other-group applicants ($\sigma_o^2 = 1$). As expected, the figure shows that the probability of hiring a same-group applicant increases (1) when the number of applicants increases and (2) when the precision of the signal of the employer on same-group candidates increases, that is, when the employer has better screening abilities.

B. Many Employers and Generational Persistence

The previous subsection showed that if hiring is based on ex ante quality assessment, a discriminatory equilibrium results in which employers predominantly hire applicants of their own type. This subsec-
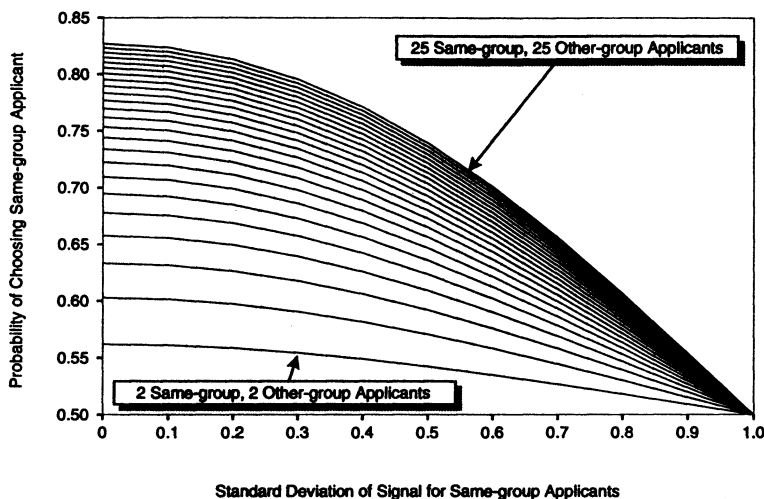
FIG. 1.—The probability that the top candidate has the same background is plotted as a function of the standard deviation of the employer's signal about (ability to screen) same-background candidates (more standard deviation implies poorer ability to screen) and the number of applicants (in steps of one). In this graph, there is an equal number of same-group and other-group applicants, the underlying population of qualities for both same-group and other-group applicants is distributed unit normal, and the ability of employers to screen other-group applicants is held constant at a standard deviation of 1.0. The figure shows that employers are more likely to hire same-group applicants (1) when they have more screening ability (lower signal standard deviation) for same-group applicants and (2) when the number of applicants increases.

tion investigates the stability of discrimination in an economy in which successful applicants become employers in the following generation. The key question is whether an equilibrium in which employers are predominantly of one type (the "status quo") can persist under reasonable assumptions. To examine this issue, we assume that employers belong to one group and compute the number of generations until the first other-group candidate is hired. For example, one could argue that slavery and other historical handicaps produced a system in which employers were exclusively nonblacks. The question we ask is how long it would take for the first black person to "break into the business" once racial biases in the Becker (1957) sense are eliminated.

To make the analysis tractable, we consider a framework in which employers have no information about other-group candidates except the unconditional distribution characteristics. For same-group candidates, who are drawn from an identical unconditional distribution, employers' screening ability provides them with information to make a fair judgment about whether a candidate is better than average or worse than average. With probability p, a same-group candidate is inferred by the employer to be better than the unconditional average, with probability $1 - p$ worse than average. For example, if the signals are symmetric (equal mass above and below the previous expected value) and continuous (to exclude cases in which the updated mean is identical to the prior mean), p would be $\frac{1}{2}$. Applicants inferred to exceed the unconditional mean are referred to as type H, and applicants inferred to be below the unconditional mean are called type L.

The market consists of m employers, each of whom plans to hire one person. Every employer observes the same signal for each of n applicants (to avoid the complications of a winner's curse). Half of the applicants are of the "same type" as the employers. Each employer hires the best applicant for the job. If there are at least m applicants inferred to be of high quality, they are hired before an other-group applicant is hired. In other words, the probability that all m employers hire type H applicants is the probability that the top m among n applicants are of type H. This probability, B, is one minus the cumulative binomial:

$$B = 1 - \sum_{i=0}^{m-1} \binom{n}{i} p^i (1 - p)^{n-i}.$$

For example, for five employers and 20 candidates and an equal probability that same-group candidates are type H or type L, the probability that the top five applicants are five of the same group is about 99.4 percent. For five employers and 40 candidates, B exceeds 99.999 percent.

Under the assumption that it is necessary to have experience in the industry to become an employer,[8] the probability that no other-group candidate is hired for t generations is B^t. Therefore, the expected number of generations before the first other-group candidate is hired with probability ξ is

$$\frac{\log(\xi)}{\log\left[1 - \sum_{i=0}^{m} \binom{n}{i} p^n (1-p)^{n-i}\right]}. \qquad (2)$$

For the five-employer/20-candidate scenario, it takes about 33 generations before the probability that a single other-group candidate is hired exceeds 50 percent. For the five-employer/40-candidate scenario, it takes about one million generations! This suggests that screening discrimination can be very persistent.

The preceding example illustrates that the model predicts that discrimination is more persistent when the applicant/employer ratio is high. Because high applicant/employer ratios are likely to occur in stable or declining industries and industries with relatively high wages in which competition for positions is severe, the model suggests that discrimination is more persistent in older industries such as banking and manufacturing than in rapidly growing industries such as computer software and biotech.

It is noteworthy that although discrimination can be long-lived, it does not persist forever. The average quality of rejected other-group applicants is higher than that of rejected same-group applicants. Because the top same-group applicants are siphoned off, the unconditional mean in the remaining pool falls.[9] If the industry allowed entry by other-group employers, even though they had not gone through the hiring procedure a generation earlier, the greater expected value in the pool of other-group applicants would assure them of excess rents in the form of higher-quality employees. This is the long-run force that pushes the eventual equilibrium in the direction of nondiscrimination.[10]

---

[8] This assumption differs from two-sided matching models, as discussed in Roth and Sotomayor (1990). Clearly, our model in this subsection is not a matching equilibrium; instead it describes the length of the adjustment process to reach the matching equilibrium.

[9] If the signal were continuous, same-group employers who are selecting applicants *after* the first few have committed to other employers would pick applicants from a posterior distribution that has a lower mean than those employers picking earlier.

[10] Note also that a same-group employer might hire an other-group applicant if she could contractually extract future rents. These rents result from the screening ability of an other-group employer to draw from the better pool in the future. We assume that this is not possible.

Furthermore, if there is a scarcity of applicants, discrimination breaks down quickly. For example, if there are two applicants per job, employers first select same-group, high-quality candidates, but they select other-group candidates before they select same-group, low-quality candidates. Similarly, applicants from the other group may break into a business in other ways, such as by lowering their reservation wage. Consequently, the model implies that discrimination is likely to be more persistent when there is little wage flexibility, for example, because of antidiscrimination laws that prohibit group-based wages.[11]

C. A Two-Sector Model

To this point, career choice has been held constant. However, if screening discrimination makes it difficult for some groups to break into professions, career decision making among other-group applicants will be affected. This, in turn, can alter the incentive within a minority population to develop the requisite skills necessary for success in those professions.

To model this possibility, consider a simple economy in which there are two industries. The high-wage sector is represented by a single firm, and the low-wage sector can expand or contract to absorb residual employees. As before, assume that no additional information other than the unconditional mean of other-group candidates is known for other-group candidates. Applicants do not know their own quality prior to training. Without loss of generality, same-group candidates are judged to be either "above mean" (type H, with probability p) or "below mean" (type L). Let there be s applicants from the same group as the single high-wage employer and o other-group applicants. Consider the decision of the high-wage employer, who hires the applicant inferred to be best. If even one of the same-group applicants is judged to be above average, the employer hires a same-group applicant. This probability that any same-group person is hired is $1 - (1 - p)^s$. Because all same-group applicants have an equal chance to be hired, the probability that a specific same-group individual is hired is $[1 - (1 - p)^s]/s$. An other-group applicant is hired into the position only if no same-group applicant is inferred to be above the unconditional mean. This probability is $(1 - p)^s$. The probability that a specific other-group individual is hired is $(1 - p)^s/o$. Now as-

---

[11] There are of course good reasons for these antidiscrimination laws, especially if they also ban hiring discrimination. If effective, such laws can restrain employers with market power and a "taste" for discrimination (Becker 1957).

sume that employees need to invest in training in order to qualify for a high-wage job. A same-group individual will become an applicant if

$$\frac{1 - (1 - p)^s}{s} W > C, \tag{3}$$

where W is the total present value of the increased wage in the high-wage sector and C is the up-front cost of acquiring necessary skills. In contrast, an other-group individual will become an applicant if

$$\frac{(1 - p)^s}{o} W > C. \tag{4}$$

From (3) and (4), no other-group applicants will ever apply for a position in the high-wage sector when

$$\frac{(1 - p)^s}{o} < \frac{C}{W} < \frac{1 - (1 - p)^s}{s}.$$

Knowing that in equilibrium other-group applicants will not participate, however, increases the probability that same-group applicants are hired from $[1 - (1 - p)^s]/s$ to $1/s$. Thus a perfectly discriminatory equilibrium exists if

$$\frac{(1 - p)^s}{o} < \frac{C}{W} < \frac{1}{s}.$$

Figure 2 illustrates the relevant parameter regions. In the two lighter gray areas, same-group applicants acquire the necessary training whereas other-group applicants do not. The example shows that screening discrimination can alter the incentives of minority applicants to acquire the necessary skills to qualify for high-wage jobs. Under such circumstances, discrimination can persist indefinitely when the high-wage sector is dominated by a particular group. Because minority applicants have lower incentives to acquire training, the tendency for discrimination to break down eventually is eliminated.

D. Ex Ante Screening versus Ex Post Evaluation

Thus far, the analysis has been based on the assumption that employers costlessly observe an exogenous number of signals. In this subsection, the model is generalized by assuming that screening is available via a costly interview. In this context, the basic condition that employers can more easily evaluate same-group applicants is interpreted to mean that employers have to expend added resources to obtain the same amount of information about an other-group applicant. If ap-
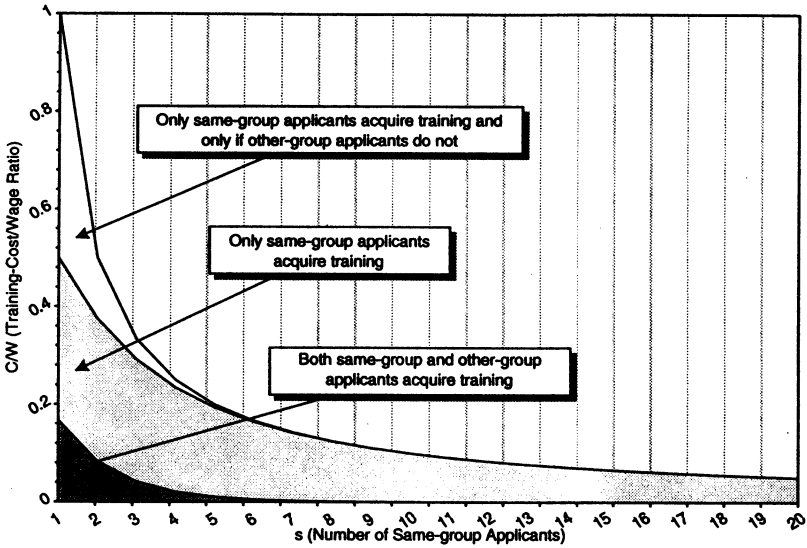
FIG. 2.—In the light gray areas, an equilibrium exists in which only same-group applicants acquire training in order to subsequently apply to the high-wage sector, whereas other-group applicants do not. The number of other-group applicants is held constant at $o = 3$. The probability that an applicant of the same group is perceived ("screened") to exceed the unconditional mean of the quality distribution is $p = \frac{1}{2}$.

plicants arrive for interviews without cost to the employer, the employer must decide which type of applicant to interview. If the number of applicants of each type is large, the employer will always choose to interview a same-group applicant, because for a given expenditure the employer can receive more accurate information about such applicants. Therefore, even a small difference in interviewing costs can magnify the apparent discrimination, because firms first interview only same-group applicants and resort to other-group applicants only when the former pool is exhausted.

When there are lower costs to screening same-group applicants, an equilibrium can arise in which only same-group candidates are interviewed and therefore hired. In some situations, however, the costs of screening, relative to the expected benefits, may be so high that employers choose to forgo screening altogether and instead rely on on-the-job evaluation. When the cost of on-the-job performance measurement—which equals the lost output sustained if a low-quality employee is hired plus the added expense of terminating and replacing that employee—is the same for all employees, relying on on-the-job evaluation eliminates discrimination. Consequently, the model predicts that screening discrimination is more prevalent in those in-

dustries in which ex ante interviews play a more important role than ex post on-the-job evaluation.

One implication of the preceding discussion is that discrimination declines if it becomes easier for employers to test employees by on-the-job trials. This suggests that legislation that makes firing employees more expensive, especially when the employee is a member of another group, has the unwanted effect of increasing reliance on ex ante interviews and increasing discrimination. Trial periods are also less effective when the potential for damage by a low-quality employee is high. Therefore, the model implies that discrimination will be more prevalent in industries in which employee error can lead to disastrous consequences, such as in capital-intensive industries and industries such as the medical and legal professions, where potential liability is large.

Another implication is that discrimination is likely to be exacerbated when it is difficult to monitor performance. For example, trial periods are unlikely to be effective when poor performance is difficult to discover before damage accumulates. Consequently, discrimination is likely to be greater at top levels of management than in basic jobs such as janitorial work, crop harvesting, trash collection, or basic construction tasks such as moving equipment and supplies.

E.   Reinforcing Effects

Thus far, employment screening discrimination has been analyzed in isolation. In realistic situations there are likely to be complex interactions. For example, customers, like employers, may be better at screening same-group service providers. Consequently, customers who can choose from a large number of firms are more likely to purchase from same-group service providers. For this reason, employers tend to hire applicants that have the same background as their customers. Therefore, a caste system in the sense of Akerlof (1984) can develop, in which it is not differential screening but third-party demand that reinforces the equilibrium.

An intriguing extension of the model involves general equilibrium considerations, or what Coleman (1990) has called micro to macro transition and the feedback from macro to micro. If the screening mechanisms analyzed here lead to segregation in the workplace, perceptions may be affected so that employers no longer believe that the unconditional distribution of character is the same in both populations. Specifically, cognitive dissonance theory as developed by Festinger (1957) implies that if a firm is segregated, employers and employees are more likely to develop theories that explain segregation as a result of the inferiority of people of the opposite type. This

perception can be enhanced by the fact that employers who act out of equilibrium by hiring an opposite type forfeit their screening advantage and, therefore, find that such employees are less productive on average.

Screening discrimination can be further amplified when employers infer the quality of applicants from the behavior of other employers. Bikhchandani, Hirshleifer, and Welch (1992) show that when decisions are discrete, as in hiring individuals from certain groups, "cascades" can quickly develop in which later employers rationally ignore their own information and act like employers before them. Thus one employer may conclude that if no other employer hired an other-group candidate, it is unwise to hire an other-group applicant, even when his own information suggests the opposite. As a result, a "power elite" may develop in which a homogeneous group of people occupy key positions in both business and social organizations. Unlike the power elites described by Mills (1956) and Domhoff (1967), however, this elite is held together not by class or racial conflict, but by constraints imposed by incomplete information.

III. Implications of the Model

Although the model offers a rich set of implications, in this section, we focus only on issues related to current American society. In this context, the other-group applicants are interpreted as "minorities." It is important to stress that in our model, a "minority" is defined as a group of people with a cultural background different from that of the dominant group. The minority characteristic per se is important in our model only insofar as it correlates with a socioeconomic or cultural background that makes mutual screening easier. Consequently, minorities may be defined by race, gender, religion, or even sexual preference or physical appearance.

A. Policy Implications

The most important policy implication of the model is that discrimination can develop not only from preferences to hire nonminority applicants or a belief that minority applicants are inferior, but also as a rational response to incomplete information. Our model suggests that overcoming discrimination is not only a matter of imposing costs on objectionable tastes or inappropriate beliefs, but also a matter of changing economically efficient behavior. Alternative policies for reducing screening discrimination can be broadly grouped into two categories: (1) reducing the information gap between the costs (or effectiveness) of screening minority versus nonminority candidates

and (2) reducing the benefit (net of cost) of ex ante screening, relative to the benefit (net of cost) of on-the-job performance measurement.[12]

With respect to the first point, reducing informational differences, to the extent that a profession can develop a common culture, screening discrimination can be ameliorated. For instance, academic economists have developed a unique culture that includes both a jargon and a set of conceptual tools that are largely orthogonal to ethnic groupings. Because ex ante assessments in academia have become based on this common culture to a large part, screening discrimination against minorities is reduced. This may explain, in part, the ethnic diversity of academic economics departments.

In contrast, the model casts doubt on the hope that discrimination can be reduced by teaching people to accept and to respect differences in culture, ethnic background, sex, and sexual orientation. Our analysis begins with the assumption that everyone accepts such differences without bias and yet screening discrimination persists. One error that is made by those who think that "accepting differences" can eliminate discrimination is treating each group of people as though it were homogeneous. By doing so, they overlook the sorting problem and its implications for understanding discrimination. Ironically, discrimination in our model arises precisely because members of the dominant group are unable to discriminate between members of the minority population. From this perspective, programs such as bilingual education may actually hinder assimilation into the currently prevalent business environment. Not only is language related to skill on the job, but use of language is also interpreted as an indicator of skill. As Mehrabian (1981) notes, people are judged not only by the content of their words but by their style of speech and inflection of their voice. All these means of communicating differ from language to language. Consequently, when people are not fluent in the native language, employers find it more difficult to assess their character ex ante, and a discriminatory equilibrium is more likely to result.

An alternative to forcing applicants to assimilate into the mainstream is to force mainstream employers to learn how to screen minority applicants more effectively. Thus one benefit of affirmative action programs is that the introduction of minority candidates to high-wage sectors helps corporations develop expertise in screening future minority candidates. Such a practice not only increases the "fairness" of the system but actually enhances its economic efficiency by improving utilization of the intrinsically identical skills of minority applicants. This conclusion, however, depends on the framework of

---

[12] Another method to reduce discrimination is to subsidize education to increase the mean in the minority pool or to exclude nonminority candidates from certain positions.

the analysis. The two-sector model developed in Section II*C* suggests that when affirmative action programs are applied indiscriminately, they may succeed in opening more positions to minority applicants in the non-high-wage sectors, which in turn can increase the incentives of minorities not to compete for high-wage sector jobs.

With respect to the second point, reducing the incentives to screen, minority job seekers benefit when the costs of ex ante screening are high relative to on-the-job performance measurement. This implies that lowering the costs of terminating an employee increases the incentive of employers to rely on direct on-the-job performance measurement and thereby lowers the likelihood that screening discrimination will persist. Unfortunately, current policy seems to be moving in the opposite direction with respect to termination costs. With increasing frequency, employers are being called on to prove that decisions to terminate minority employees are based on merit, not racial considerations. Because such proof is difficult to develop and because litigation is expensive, the willingness to rely on experimentation and direct observation is reduced and discriminatory equilibria become more likely. To ameliorate this problem, the government conceivably could directly promote information acquisition about minority applicants or provide subsidies or guarantees to employers that are hiring their first minority applicants. Finally, minority applicants might be able to overcome screening discrimination by lowering their reservation wage. For this reason, "fair-wage laws," designed to equate wages across ethnic groups, should be evaluated carefully.

In closing, we should note that policies designed to reduce screening discrimination may open a back door to individuals who harbor racist, ethnic, nationalist, sexist, or other preferences. For example, when racist employers have market power, permitting lower minority wages may allow such employers in effect to satisfy their tastes for discrimination. Similarly, the theory presented here carries the danger that it could mistakenly be used to justify discrimination based on informational asymmetry even when the motive is ethnic.[13] The critical empirical question, therefore, is whether discrimination is based on information or preferences.

B.   *Empirical Predictions and Possible Tests*

To review, the screening discrimination model makes several new predictions about the nature of discrimination. First, the primary result of our model is that discrimination arises when individuals

---

[13] Most likely, in any given case, discrimination is a combination of preferences, beliefs, and screening.

screen, ex ante, on the basis of background. This implies that the hiring decision is influenced not only by the similarity of the applicant's and the interviewer's race, sex, or other minority status but also by the similarity of their general background. By general background we mean such items as schooling, community participation, and club membership or even parental status, religious beliefs, or style of dress. Also, we would expect to see not only majority interviewers discriminate against minority applicants but also minority interviewers discriminate against majority applicants. Second, the model predicts that screening discrimination is more severe when there is a large number of applicants compared to job openings, such as in old, stable, or declining sectors. On the other hand, the model predicts that screening discrimination works *against* same-background candidates when there are few applicants for many jobs to fill. This occurs because the model implies a rank ordering of applicants, starting with high-inferred-quality nonminority candidates, followed by all minority candidates, and last by low-inferred-quality nonminority candidates.[14] Third, screening discrimination should occur primarily in sectors in which inferred quality is important and in which it is more efficient to screen ex ante than to measure on-the-job performance ex post.

A recent study by the Federal Reserve illustrates a possible test of the model. The study was prompted by 1990 data released under the Home Mortgage Disclosure Act showing substantially higher denial rates for black and Hispanic loan applicants. The data were analyzed extensively by Boston Fed economists Munnell et al. (1992). Their paper presented what they viewed as clear evidence of discrimination. Using logit regression models to hold constant factors including borrower credit history, borrower wealth, age and marital status, property type, existence of mortgage insurance, and loan-to-value ratios, Munnell et al. report that race is a significant determinant of the probability that a loan application will be accepted.

Our model suggests that the Munnell et al. study, despite its comprehensive nature, omitted a key variable. As Paré (1993) reports, if there is an art to banking, it lies in making a character loan. He quotes one banker as saying that after examining all the financial data, "You simply have to look into someone's eyes and perceive that he is indeed earnest and capable and passionate about his particular project, and agree to gamble on him" (p. 74). This quote captures the essence of our analysis. In order to see something meaningful when one looks

---

[14] Self-selection complicates an empirical test. If low-inferred-quality nonminority candidates, aware of their characteristics, decide not to spend resources applying, the predicted rank ordering of applicants would break down.

into someone's eyes, there must be a common cultural base. There-
fore, the Boston Fed analysis omitted one key variable: the back-
ground of the lending officer. Our model predicts that minority lend-
ing officers with extensive experience in minority neighborhoods are
likely to approve *more* minority loan applications. In addition, the
model predicts that the relative acceptance rate of minority applica-
tions falls when loan rationing increases. Finally, the model implies
that legislation that makes it difficult for banks to quickly foreclose on
problem loans would reduce the relative acceptance rate of minority
applicants at nonminority banks.[15]

C. Experimental Implications

The key assumption on which the screening discrimination model is
based is that people are better at evaluating same-group individuals
and that this enhanced ability leads to statistical discrimination when
selecting from a large sample. This suggests a series of questions
that could be examined experimentally: (1) Are experimental subjects
better at judging people of similar background? (2) Do experimental
subjects believe that they are better at judging people of similar back-
ground? (3) How do experimental subjects update the inferred vari-
ance and accuracy of their assessments? (4) Do experimental subjects
discriminate in favor of people of similar background on the basis of
their updated conditional probabilities? (5) Do experimental subjects
discriminate more when screening from a large number of same-
group and other-group applicants for a limited number of positions?

The work that comes closest to testing the assumptions of the model
includes studies of the employment interview. Reviews of this litera-
ture by Arvey and Campion (1982) and Harris (1989) summarize
studies of the impact of demographics (race, gender, age, and the
like) on the interview process. The central question examined in those
studies, however, is whether interviewers tend to be biased in favor
of or against specific groups. The question of whether interviewers
are able to make more accurate assessments of an applicant's skill
when the applicant has a similar cultural and racial background has
not yet been addressed.

IV. Summary and Conclusion

This paper has developed a model in which two well-known ideas
interact. The first idea, drawn from the economics literature, is that

[15] It would be interesting to examine whether small, well-collateralized or insured
loans (likely to be repaid) are subject to less discrimination. Our model would be
inapplicable if discrimination was found to be no less severe under these circumstances.

people face the problem of sorting other people in virtually every walk of life. Whether one is hiring a new employee or choosing a new club member, the task is to distinguish those who are likely to be successful from those who are not. The second idea, drawn from the literature in social psychology and sociology, is that culture plays a central role in interpersonal communication and evaluation. This occurs because people communicate not only through words but through a vast array of cultural cues. In addition, interpretation of an individual's personal history requires understanding the cultural context in which that individual lives.

Putting the two ideas together leads to the conclusion that individuals favor associating with people of their own type, even when they have no preference for similar people, because they can more accurately distinguish good and bad individuals in a population of similar people. If there is no offsetting benefit to compensate for the fact that it is more difficult to sort people whose culture is unfamiliar, a discriminatory equilibrium can arise. This discrimination, however, is *not* based on ethnic preference, on a belief regarding the superiority of certain cultures, or on a failure to accept cultures other than one's own. It results simply from the fact that unfamiliarity makes it more difficult to make accurate assessments and that rationed allocations such as good jobs, friendships, and marriages go toward candidates inferred to be of the highest quality, which in a large pool of potential candidates tend to be people of similar background.

Appendix

Computations to Derive the Probability P of Hiring a Same-Group Applicant

A. The Discrete Example with Uniform Prior and Binomial Signals

This section illustrates the complexities involved in computing P, the probability of hiring a same-group applicant. We consider the distributional assumptions in the simple example used in the text to illustrate the variance increase: the employer has a uniform prior on the distribution of quality and receives binary conditionally independently and identically distributed signals (signal H with probability Q). Assume that there are s same-group applicants and o other-group applicants. For same-group applicants, the employer receives n_s signals; for other-group applicants, the employer receives n_o signals, with $n_o < n_s$. In other words, the employer receives additional signals by his improved screening ability for same-group applicants.

We can write the probability that the employer's inferred quality of this employee is Q_0 as

$$\text{prob}[E(Q|\check{k}, n) = Q_0] = \text{prob}\left[\frac{k+1}{n+2} = Q_0\right] = \text{prob}[k = Q_0(n+2) - 1] \quad (A1)$$

for Q_0 in $1/(n + 2)$, $2/(n + 2)$, . . . , $(n + 1)/(n + 2)$. The probability that the expected value of the *best* among s same-group candidates is Q_0 is the probability corresponding to a particular index i of H signals:

$$\text{prob}[k_{\max} = Q_0(n + 2) - 1] = \left(\frac{i + 1}{n_s + 1}\right)^s - \left(\frac{i}{n_s + 1}\right)^s. \qquad \text{(A2)}$$

For example, for $n = 5$ signals, the probability that the expected value is $1/7$ is equivalent to the probability that the employer receives zero H signals, which is $1/6$; the probability that the expected value is $2/7$ is $1/6$, and so forth. If there are $s = 2$ same-group applicants, the probability that the top same-group applicant has an expected value of $1/7$ is $\text{prob}[k_{\max} = 1/7 \cdot (5 + 2) - 1] = (1/6)^2 - (0/6)^2 = (1/6)^2$.

The inferred qualities of one same-group candidate and one other-group candidate are equal if $(i + 1)/(n_o + 2) = (j + 1)/(n_s + 2)$, where i is the number of H signals for other-group applicants and j is the number of H signals for same-group applicants. The probability that the inferred quality of the best other-group applicant is below the quality of the best same-group applicant if the same-group applicant received a score of i H signals is

$$\text{prob}\{\text{EV}_o | n_o \text{ signals} \le [\text{EV}_s | i H \text{ signals in } n_s \text{ signals}]\} \qquad \text{(A3)}$$

$$= \left\{\frac{\text{Int}\left[\dfrac{(i + 1)(n_o + 2)}{n_s + 2} - 1\right]}{n_o + 1}\right\}^o.$$

To compute the probability of hiring a same-group applicant, one needs to compute the inference over all possible inferred qualities or, equivalently, over all possible signal realizations. Together, the expression for the probability of hiring a same-group applicant is[16]

$$P = \sum_{i=0}^{n_s} \left[\left(\frac{i + 1}{n_s + 1}\right)^s - \left(\frac{i}{n_s + 1}\right)^s\right]\left\{\frac{\text{Int}\left[\dfrac{(i + 1)(n_o + 2)}{n_s + 2} - 1\right]}{n_o + 1}\right\}^o \qquad \text{(A4)}$$

$$= (n_s + 1)^{-s}(n_o + 1)^{-o}\left\{\sum_{i=0}^{n_s} [(i + 1)^s - i^s]\text{Int}\left[\frac{(i + 1)(n_o + 2)}{n_s + 2} - 1\right]^o\right\}.$$

It can be shown that the probability of hiring an applicant of the same group increases with the number of same-group applicants (s) and decreases with the number of other-group applicants (o). However, we cannot algebraically show that this expression increases with n_s and decreases with n_o, although both intuition and simulations show this to be the case.

---

[16] This expression assumes that an employer hires a same-group applicant if the inferred quality of the best same-group candidate equals that of the best other-group candidate. If, on indifference, an other-group candidate is hired, the i in the "Int" expressions needs to be changed to $i + 1$.

B.   A Continuous Example with Normal Priors and Normal Signals

Figure 1 in the text graphs the probability of hiring a same-group applicant for continuous signal distributions.[17] In figure 1, we assume that the distribution of qualities in the underlying population, Q, is distributed Gaussian normal with mean zero and unit variance, and employers observe a Gaussian normal signal, $\bar{x}$, with mean Q. For same-group candidates, the signal has variance σ_s^2 and for other-group candidates variance σ_o^2, $\sigma_s^2 < \sigma_o^2$. Let s again be the number of same-group applicants and o be the number of other-group applicants. DeGroot (1970, p. 169) shows that an employer's updating rule creates a posterior distribution of inferred qualities for one candidate that is normal with a mean

$$\hat{Q} = \frac{\bar{x}/\sigma^2}{1 + \sigma^2}. \tag{A5}$$

Rearranging and taking expectations with respect to $\bar{x}$, we see that proposition 1 holds. The employer's inferred quality, $\hat{Q}$, about one employee is distributed

$$\hat{Q} \sim N\left(0, \frac{1}{\sigma^2 + 1}\right), \tag{A6}$$

which states that the variance of inferred qualities decreases with the variance of the information obtained through screening. One could now compute the actual probability by substituting (A6) into the equation in note 17. Unfortunately, this expression cannot be easily differentiated and signed because it depends on rather complicated weighting factors in the integral. Instead we have to rely on an indirect proof.

PROPOSITION 3. Assume that the distribution of inferred qualities of same-group candidates is $N(0, \sigma_a^2)$ and that of other-group candidates is $N(0, \sigma_b^2)$, $\sigma_a^2 > \sigma_b^2$. Then the probability that the top candidate among n same-group and n other-group candidates ($n \geq 2$) is drawn from the same-group population exceeds ½ and strictly increases in the ratio σ_a^2/σ_b^2.

To prove this proposition, we need to rely on the single-crossing property of the two normal distributions. Let S be the (wider) cumulative $N(0, \sigma_a^2)$ distribution and O be the $N(0, \sigma_b^2)$ distribution. The term σ_a^2 is the variance of inferred qualities (not of the signal) and thus is equivalent to $1/(1 + \sigma_s^2)$. Equivalently, $\sigma_b^2 = 1/(1 + \sigma_o^2)$. Not only do $S(x)$ and $O(x)$ have equal means (of zero), but the normal distribution function also assures that, for $x > 0$, $O(x) > S(x)$ and, for $x < 0$, $O(x) < S(x)$ (single crossing at zero). In the context of the model, $S(x)$ is the distribution of inferred qualities for same-

---

[17] For a continuous distribution, the probability can be rewritten as

$$P = \int_c \left\{ \left[\frac{\partial}{\partial c} \prod_{i=1}^s F_s(c) \right] F_o(c)^o \right\} dc = \int_c [sF_s(c)^{s-1} f_s(c) F_o(c)^o] dc,$$

where $F_s(\ )$ is the (ex ante) cumulative distribution of inferred qualities for a single applicant of group s, as discussed earlier.

group employees and $O(x)$ is the distribution of inferred qualities for other-group employees.

Consider the function $h(x) = S^{-1}(O(x))$. It is straightforward to verify that, for $x > 0$, $h(x) \geq x$, and vice versa for $x < 0$. Specifically, for our normal distributions, $h(x) = (\sigma_s/\sigma_o)x$. Let o_i and p_i be independent draws from distribution O, and let $s_i = h(p_i)$. Then the s_i are a sample from distribution S and are independent from o_i. Let $\overline{O}$ be the largest o_i, $\overline{P}$ be the largest of the p_i, and $\overline{S}$ be the largest s_i. It follows that $\overline{S} = h(\overline{P})$. We are interested in the probability that $\overline{S} > \overline{O}$. First note that the event space can be enumerated into the following set of cases:

$(0<\overline{O}<\overline{S}<\overline{P})$ $(0<\overline{O}<\overline{P}<\overline{S})$ $(0<\overline{P}<\overline{O}<\overline{S})$ $(0<\overline{P}<\overline{S}<\overline{O})$ $(0<\overline{S}<\overline{O}<\overline{P})$ $(0<\overline{S}<\overline{P}<\overline{O})$

$(\overline{O}<0<\overline{S}<\overline{P})$ $(\overline{O}<0<\overline{P}<\overline{S})$ $(\overline{O}<\overline{S}<0<\overline{P})$ $(\overline{O}<\overline{S}<\overline{P}<0)$ $(\overline{O}<\overline{P}<0<\overline{S})$ $(\overline{O}<\overline{P}<\overline{S}<0)$

$(\overline{S}<0<\overline{O}<\overline{P})$ $(\overline{S}<0<\overline{P}<\overline{O})$ $(\overline{S}<\overline{O}<0<\overline{P})$ $(\overline{S}<\overline{O}<\overline{P}<0)$ $(\overline{S}<\overline{P}<0<\overline{O})$ $(\overline{S}<\overline{P}<\overline{O}<0)$

$(\overline{P}<0<\overline{O}<\overline{S})$ $(\overline{P}<0<\overline{S}<\overline{O})$ $(\overline{P}<\overline{O}<0<\overline{S})$ $(\overline{P}<\overline{O}<\overline{S}<0)$ $(\overline{P}<\overline{S}<0<\overline{O})$ $(\overline{P}<\overline{S}<\overline{O}<0)$

Now note that because $\overline{S} = h(\overline{P})$, if $0 < \overline{P}$, then $\overline{P} < \overline{S}$. Therefore, the following events never happen: $(0 < \overline{O} < \overline{S} < \overline{P})$, $(0 < \overline{S} < \overline{O} < \overline{P})$, $(0 < \overline{S} < \overline{P} < \overline{O})$, $(\overline{O} < 0 < \overline{S} < \overline{P})$, $(\overline{O} < \overline{S} < 0 < \overline{P})$, $(\overline{S} < 0 < \overline{O} < \overline{P})$, $(\overline{S} < 0 < \overline{P} < \overline{O})$, and $(\overline{S} < \overline{O} < 0 < \overline{P})$. Analogously, if $\overline{P} < 0$, then $\overline{S} < \overline{P}$, and $P(\overline{O} < \overline{P} < 0 < \overline{S})$, $P(\overline{O} < \overline{P} < \overline{S} < 0)$, $P(\overline{P} < 0 < \overline{O} < \overline{S})$, $P(\overline{P} < 0 < \overline{S} < \overline{O})$, $P(\overline{P} < \overline{O} < 0 < \overline{S})$, $P(\overline{P} < \overline{O} < \overline{S} < 0)$, $P(\overline{P} < \overline{S} < 0 < \overline{O})$, and $P(\overline{P} < \overline{S} < \overline{O} < 0)$ never occur. We want to compute $P(\overline{O} < \overline{S})$. Collecting terms from the set above in which $\overline{O} < \overline{S}$ and omitting zero terms, we find that

$$P(\overline{O} < \overline{S}) = P(0 < \overline{O} < \overline{P} < \overline{S}) + P(0 < \overline{P} < \overline{O} < \overline{S})$$
$$+ P(\overline{O} < 0 < \overline{P} < \overline{S}) + P(\overline{O} < \overline{S} < \overline{P} < 0)$$
$$= P(0 < \overline{O} < \overline{P} < h(\overline{P})) + P(0 < \overline{P} < \overline{O} < h(\overline{P}))$$
$$+ P(\overline{O} < 0 < \overline{P} < h(\overline{P})) + P(\overline{O} < h(\overline{P}) < \overline{P} < 0).$$

But note that if $\overline{P} < 0$, $h(\overline{P}) < \overline{P}$, so we can rewrite $P(0 < \overline{O} < \overline{P} < h(\overline{P}))$ as $P(0 < \overline{O} < \overline{P})$ and $P(\overline{O} < 0 < \overline{P} < h(\overline{P}))$ as $P(\overline{O} < 0 < \overline{P})$. Furthermore, note that $P(0 < \overline{P} < \overline{O} < h(\overline{P}))$ is $P(0 < \overline{P} < \overline{O}) - P(\overline{O} > h(\overline{P}) > 0)$, and $P(\overline{O} < h(\overline{P}) < \overline{P} < 0)$ is $P(\overline{O} < h(\overline{P}) < 0)$. Therefore, we can rewrite our probability as

$$P(\overline{O} < \overline{S}) = P(0 < \overline{O} < \overline{P}) + P(\overline{O} < 0 < \overline{P}) + P(0 < \overline{P} < \overline{O})$$
$$+ [P(\overline{O} < h(\overline{P}) < 0) - P(\overline{O} > h(\overline{P}) > 0)].$$

Intuitively, this equation states that when $h(\cdot)$ changes, the probability that the top candidate comes from the same group increases with $P(\overline{O} < h(\overline{P}) < 0) - P(\overline{O} > h(\overline{P}) > 0)$. We now show that the variance ratio in $h(\cdot)$ has the effect of shifting out $\overline{P}$. If σ_a/σ_b increases, more probability mass is lost in the second term $P(\overline{O} > h(\overline{P}) > 0)$ than is gained by the second term $P(\overline{O} < h(\overline{P}) < 0)$. Figure A1 illustrates these two areas. Consider any fixed possible realizations of $\overline{P}$. When there are two same-group candidates, a symmetric outward shift loses more (shaded) area on the right (in the positive section) than it does on the left. Thus $P(\overline{O} < \overline{S})$ increases with $h(\cdot)$ when $n = 2$. The figure also visualizes why there is no change in probability when there is only
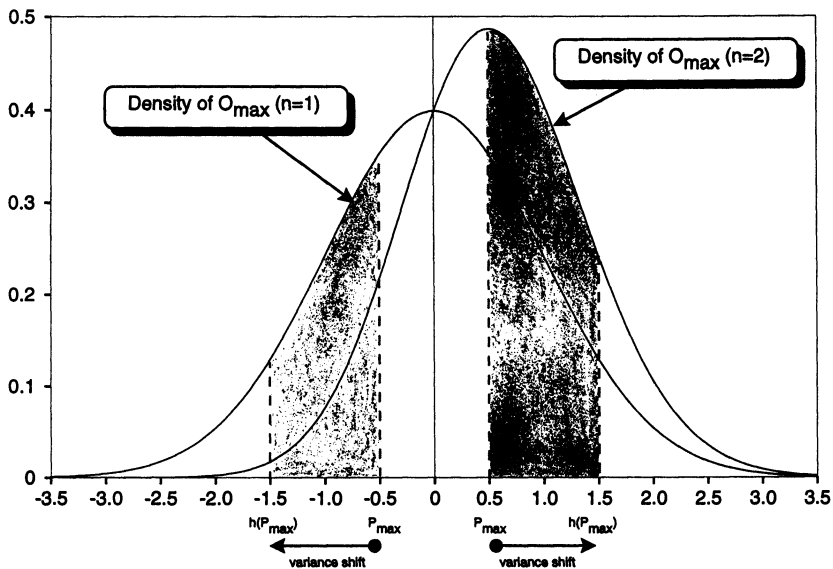
FIG. A1.—This figure plots the density function for the top other-group candidate for one and two other-group candidates. The question is whether shifting cutoffs toward the tails (by using $h(x)$ rather than x) reduces the difference of the two areas in the tail. When $n = 1$, the net loss of area from increasing $h()$ is identical because the density is symmetric. When $n = 2$, however, the area loss in the below 0 region from increasing $h()$ is always less than the area gain in the above region from increasing $h()$.

$n = 1$ candidate: a symmetric shift shaves off equal areas under the normal curve. Algebraically, if we can show that for, all positive P_0,

$$\frac{\partial[P(\overline{O} < -P_0 < 0) - P(\overline{O} > P_0 > 0)]}{\partial P_0} < 0,$$

then we can conclude that an increase in σ_a/σ_b increases $P(\overline{O} < \overline{S})$. By inspection of figure A1, the net change in area can be expressed not only by the change in shaded areas from $\overline{P}$ to $h(\overline{P})$ but also by the change in the tail area to the sides of $h(\overline{P})$. The left tail area is $F(-P_0)$ and the right tail area is $1 - F(P_0)$. We need to subtract the two areas and examine the change in this area difference with respect to P_0; that is, we need to compute

$$\frac{\partial\{F^n(-P_0) - [1 - F^n(P_0)]\}}{\partial P_0} = nF(P_0)^{n-1}F'(P_0) - F(-P_0)^{n-1}F'(-P_0),$$

where F denotes the cumulative distribution function of one single draw from the inferred quality distribution of other-group candidates (same as O), and n is the number of candidates. Exploiting the symmetry of the normal distribution, $F'(-P_0) = F'(P_0) = f(P_0)$, we get

$$\frac{\partial\{F^n(-P_0) - [1 - F^n(P_0)]\}}{\partial P_0} = nf(P_0)[F(P_0)^{n-1} - F(-P_0)^{n-1}].$$

As expected, when $n = 1$, there is no net area change due to the symmetry of the normal distribution. However, when $n > 2$, because $F(P_0) > F(-P_0)$, the expression is positive. In sum, we have shown that, for any given draw from $h(\overline{P})$ (i.e., any possible inferred quality of the best other-group candidate), when $h(x)$ increases (i.e., when the relative variance ratio of the inferred qualities increases), the probability that a same-group candidate is chosen increases. Thus a higher variance ratio of inferred qualities also strictly increases the unconditional probability that a same-group candidate is chosen. Q.E.D.

Figure 1 in the text plots the probability of picking a same-group candidate as a function of σ_s^2 and n. As with any other good proof, the figure does not contradict our proof.

C. An Example That the Variance Increase Alone Is Not Sufficient

We conclude with an example that shows that employers with more screening ability are not always more likely to hire same-group applicants in small applicant pools for certain distributions. As we explain below, this example depends on an unusual screening signal that shifts most of the mass of the inferred quality distribution toward zero, while at the same time increasing the variance of this distribution.

Assume that the employers judge other-group applicants to be of quality $+1$ or -1 with equal probability. For same-group applicants, the employer receives one extra signal: If the candidate prescreening was judged to be of quality $+1$ and the signal is A, the expected quality jumps from $+1$ to $+25$. However, signal A is observed with only 4 percent probability. With 96 percent probability, signal B is observed, and the expected quality drops from $+1$ to 0. (As required, the extra signal does not alter the expected values $[.04 \cdot 25 + .92 \cdot 0 = 1]$.) Symmetrically, if before prescreening the candidate was judged to be of quality -1, with the additional screening information, the expected quality can jump from -1 to -25 with 4 percent probability, and with 96 percent probability, the expected quality moves from -1 to 0. The resulting distribution of expected qualities for same-group applicants is therefore -25 with probability 4 percent, 0 with probability 92 percent, and $+25$ with probability 4 percent. (Proposition 1 is easy to verify: the variance of the inferred quality distribution is higher with the screening signal, increasing from $.5 \cdot 1^2 \cdot 2 = 1$ to $4\% \cdot 25^2 \cdot 2 = 50$.)

Now compute the probability of hiring an other-group candidate. First, note that screening does not influence the likelihood of hiring a same-group candidate if there is one candidate each. The probability that the employer chooses the other-group candidate is the probability that the inferred quality of the best other-group candidate is larger than the inferred quality of the best same-group candidate. With one same-group and other-group candidate each, this is the probability that the best other-group candidate is -1 (50 percent) and the best same-group candidate is -25 (4 percent), plus the probability that the best other-group candidate is $+1$ (50 percent) and the best same-group candidate is -25 or 0 ($4\% + 92\% = 96$ percent). This sums to 50 percent, an equal probability that the hired candidate has either background.

With two same-group and two other-group candidates, however, the probability that an other-group candidate is chosen is the probability that the best other-group candidate is -1 (25 percent) and the best same-group candidate is -25 (0.16 percent), plus the probability that the best other-group candidate is $+1$ (75 percent) and the best same-group candidate is -25 or 0 (96% $\cdot$ 96%). This sums to a 70 percent probability of hiring the other-group candidate!

In other words, increasing the information (through screening ability) by the employer for a subgroup of applicants may *decrease* the probability of hiring an employee from this subgroup as long as there are two employees. (When the number of applicants is increased, the probability of hiring a same-group employee eventually exceeds the probability of hiring an other-group employee.) We conclude that firms with better screening ability on same-group applicants are more likely to hire such in small applicant pools only under distributional assumptions that rule out large asymmetric shifts in the inferred quality distribution.

References

Aigner, Dennis J., and Cain, Glen G. "Statistical Theories of Discrimination in Labor Markets." *Indus. and Labor Relations Rev.* 30 (January 1977): 175–87.

Akerlof, George A. "The Economics of Caste and of the Rat Race and Other Woeful Tales." In *An Economic Theorist's Book of Tales: Essays That Entertain the Consequences of New Assumptions in Economic Theory.* Cambridge: Cambridge Univ. Press, 1984.

Archer, Dennis, and Akert, Robert M. "Words and Everything Else: Verbal and Nonverbal Cues to Social Interpretation." *J. Personality and Soc. Psychology* 35 (June 1977): 443–49.

Arrow, Kenneth J. "Models of Job Discrimination." In *Racial Discrimination in Economic Life,* edited by Anthony H. Pascal. Lexington, Mass.: Heath, 1972.

Arvey, Richard D., and Campion, James E. "The Employment Interview: A Summary and Review of Recent Research." *Personnel Psychology* 35 (Summer 1982): 281–322.

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

Bikhchandani, Sushil; Hirshleifer, David; and Welch, Ivo. "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades." *J.P.E.* 100 (October 1992): 992–1026.

Brown, Roger. *Social Psychology.* 2d ed. New York: Free Press, 1986.

Coleman, James S. *Foundations of Social Theory.* Cambridge, Mass.: Harvard Univ. Press, 1990.

DeGroot, Morris H. *Optimal Statistical Decisions.* New York: McGraw-Hill, 1970.

Domhoff, G. William. *Who Rules America?* Englewood Cliffs, N.J.: Prentice-Hall, 1967.

Festinger, Leon. *A Theory of Cognitive Dissonance.* Stanford, Calif.: Stanford Univ. Press, 1957.

Harris, Michael M. "Reconsidering the Interview: A Review of Recent Literature and Suggestions for Future Research." *Personnel Psychology* 42 (Winter 1989): 691–725.

Johnson, Harold G.; Ekman, Paul; and Friesen, Wallace V. "Communicative Body Movements: American Emblems." *Semiotica* 15, no. 4 (1975): 335–53.

Lang, Kevin. "A Language Theory of Discrimination." *Q.J.E.* 101 (May 1986): 363–82.

Lazear, Edward P., and Rosen, Sherwin. "Rank-Order Tournaments as Optimum Labor Contracts." *J.P.E.* 89 (October 1981): 841–64.

McCall, John J. "The Simple Mathematics of Information, Job Search, and Prejudice." In *Racial Discrimination in Economic Life,* edited by Anthony H. Pascal. Lexington, Mass.: Heath, 1972.

Mehrabian, Albert. *Silent Messages: Implicit Communication of Emotions and Attitudes.* 2d ed. Belmont, Calif.: Wadsworth, 1981.

Mills, C. Wright. *The Power Elite.* New York: Oxford Univ. Press, 1956.

Munnell, Alicia H.; Browne, L. E.; McEneaney, J.; and Tootell, G. M. B. "Mortgage Lending in Boston: Interpreting the HMDA Data." Manuscript. Boston: Fed. Reserve Bank, 1992.

Paré, Terence P. "Why Banks Are Still Stingy." *Fortune* 127 (January 25, 1993): 73–75.

Phelps, Edmund S. "The Statistical Theory of Racism and Sexism." *A.E.R.* 62 (September 1972): 659–61.

Price, Monroe. "As Women Fight Discrimination, History Repeats Itself." *Nat. Law J.* (July 30, 1984).

Rosen, Sherwin. "The Economics of Superstars." *A.E.R.* 71 (December 1981): 845–58.

Roth, Alvin E., and Sotomayor, Marilda A. Oliveira. *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis.* Cambridge: Cambridge Univ. Press, 1990.

Scheflen, Albert E. *Body Language and the Social Order: Communication as Behavioral Control.* Englewood Cliffs, N.J.: Prentice-Hall, 1971.