Prejudice and Competitive Signaling

Sue H. Mialon*

Abstract

In a signaling game between an employer and applicants, there is prejudice if applicants are pre-judged based on an index, such as race, without reference to their qualifying signals. We investigate what causes the employer to ignore the applicants' informative signals in favor of the uninformative index. Prejudice arises when competition between the applicants erodes their signaling incentive and reduces the quality of signals significantly. To minimize the decrease in quality, the employer may pre-screen the applicants based on the index, removing the impact of competition. In order to reduce prejudice, we advocate policies that enhance the quality of signals.

JEL Classification Numbers: J71, J78, D82, K31

Keywords: Competitive Strategic Signaling, Prejudice, Discrimination.

*Sue H. Mialon, Department of Economics, Emory University, Atlanta, GA 30322 (e-mail: smialon@emory.edu, phone: 404-712-8169). I am grateful to Maria Arbatskaya, Daniel S. Hamermesh, R. Preston McAfee, Hugo Mialon, Kaz Miyagiwa, Malathi Velamuri, and Seung Han Yoo for their helpful comments.

1 Introduction

Prejudice is "a preconceived judgment or opinion, an adverse opinion or learning formed without just grounds or before sufficient knowledge" (Merriam Webster dictionary). Hence, prejudice implies decision-making without having all relevant information. Typically, this requires a situation in which the relevant information is private and not readily available. Signaling theory predicts that, in such a situation, a person (an applicant) who has valuable, private information would have an incentive to reveal the hidden information by an investment in signaling methods. What has not received much attention, however, is that successful signaling also requires the receiver of the information (an employer) to have an incentive to utilize the signaled information. Although signaling models assume that the employer always fully utilizes the signaled information, the existence of prejudice implies that this may not be the case in reality.

In this context, this paper develops a model of prejudice as an outcome of unsuccessful signaling, not because of a lack of signaling incentives by the applicant, but because of a lack of interest by the employer in utilizing the signals. This paper investigates when the employer becomes less interested in fully utilizing the information (henceforth, full information) and discusses the effectiveness of related policies.

The model considers a signaling game between an employer and n applicants. The applicants are easily distinguishable by an irrelevant but observable characteristic, such as race, gender, or physical beauty, which is termed an index, following Spence (1973). On the other hand, the applicants' quality is private information. They can signal their hidden quality to the employer by investing in an effort.

Consider an example in which two applicants, a male and a female, are considered for a job. In order to determine if they are qualified, the employer must review their application thoroughly and process their signals. However, instead of acquiring all of the information and signals that are available from each applicant, the employer may decide to examine only the male candidate's information and not the information of the female candidate. In this case, there is *prejudice*, not because gender was a factor in the decision, but because some useful information about the female candidate's qualification was ignored as a result of gender-based pre-screening. Whether the female candidate was qualified, or even more qualified than the male candidate, would never be known.

Hence, *prejudice* differs from *discrimination* in that prejudice refers to differentiating the two applicants by gender *before* the employer learns anything about them, whereas discrimination refers to treating them differently on the basis of gender even *after* the employer has investigated all of the available information and found that both have the same qualifying signals.

Why would the employer prefer to forgo potentially useful information based on uninformative gender (index) instead of acquiring full information? To the employer, such a pre-screening based on an index is preferable only if more information does not help in identifying high quality candidates. Then, when is full information inefficient? In our benchmark case, which considers a typical model of signaling in which each type of applicant invests in signals independently, we find that the employer has no reason to discard any valuable information from the applicants in favor of less informative index. Hence, such a framework is inadequate to explain prejudice.

However, the game changes if there is a shortage of return from successful signaling, which we consider in Section 4. In the example of two applicants, the shortage occurs if there is only one position available for the two. Then, there is a chance that the applicants' signaling efforts will be wasted even without prejudice, because there is insufficient room for all of the qualified applicants. An applicant's chance of being chosen depends on whether the other applicant has a qualifying signal. This makes the two applicants' signaling decisions interdependent. Since receiving a return from their signaling efforts becomes uncertain, the applicants' incentive to signal decreases. Even fair competition under full information can be as discouraging as prejudice. The only difference is that, under prejudice, the effort is wasted if the applicant does not possess the favorable characteristic, whereas under competition for limited positions, the effort is wasted if the applicant does not win the competition. As the applicants' signaling efforts decrease, the value of full information to the employer diminishes.

The value of information decreases further if the signals cannot differentiate the applicants sufficiently. For example, suppose that every applicant makes a signaling effort and all applicants obtain equally-qualifying signal. Then, even after reviewing all information, the employer would have to rely on the index for a selection, because there are not enough positions to hire all of them and the signals are useless in differentiating the applicants. We call such an index-based decision comparative discrimination to distinguish it from prejudice and statistical discrimination.[1] The

[1] The differences between comparative discrimination and statistical discrimination are explained

possibility of comparative discrimination lowers applicants' signaling incentives and reduces the value of signals to the employer. The decrease in the value of information can destroy the value of competitive signaling that runs on full information. As the employer expects the quality of information under full information to deteriorate substantially, he chooses prejudice.

The main role of prejudice is to *pre-screen* the applicants in order to reduce the number of applicants to consider and to decrease the amount of information to process. Since the selection is independent of the applicants' effort, it weakens the value-destroying competition among the applicants. That is, when competition reduces the applicants' signaling incentives, prejudice becomes an alternative in order to alleviate the negative impact of competition.

The benefits of competitive signaling are from allocating signaling incentives to *all* applicants. In contrast, prejudice always results in the most biased allocation of the incentives. In this paper, prejudice is inefficient when an overall greater value of signaling can be obtained by keeping opportunities open for every applicant. However, we find that competitive signaling can be more inefficient than prejudice, if it distributes incentives to all applicants only to ensure that no one receives sufficient incentive to make a meaningful effort. This may happen, for example, if signaling is highly costly. Also, competitive signaling is inefficient if the signals are more correlated with the index, such as race or gender, than with the applicants' quality. For example, suppose that it is prohibitively costly for minorities to signal, but not for others. Then, minorities with high ability will never have a sufficient incentive to signal due to economic hardship even if there is no prejudice. Thus, in this case, there is no essential difference between competitive signaling and prejudice.

In order to maintain the efficiency of full information under competition, we emphasize the importance of policies that improve the quality of signals and make the signals closely correlated with the hidden quality but not with the applicants' current economic status, race, or gender. For instance, the Head Start program or the No Child Left Behind policy may help to reduce the racial gap due to an unfavorable signaling environment for minorities. If these policies can make signaling equally accessible for both races, independently of their economic status, any difference in their signaling outcomes would be attributed to a difference in the hidden quality. In that case, a competition-oriented admission policy would work efficiently.

---

in Section 4.2.1.

The remainder of this paper is organized as follows. Section 2 briefly summarizes the contributions of this paper to the literature. Section 3 presents a benchmark case in which there is no competition among applicants. In Section 4, we introduce a strategic signaling model with a possible shortage of the return for signaling efforts. Section 5 derives equilibrium and provides comparative statics. In Section 6, we discuss the policy implications of the results. Section 7 presents our conclusions.

2 Contributions to the Literature

This paper is the first to model prejudice, in distinction from discrimination, as a purposeful act of ignoring potentially valuable information about applicants in favor of a pre-judgment based on an uninformative index. This paper also analyzes how the expectation of discrimination interacts with prejudice.

On discrimination, there is an abundance of theoretical literature. While some theories, such as in Becker, (1957) describe discrimination as a matter of preference, many others spurred by the seminal papers of Phelps (1972) and Arrow (1973) explains group inequality as a result of rational choice based on various reasons that motivate the use of group index to infer some relevant quality of applicants in the presence of incomplete information about the unobservable quality.

In models of statistical discrimination, group inequality can arise in equilibrium even if there is no exogenous difference between groups. Coate and Loury (1993) explains that the inequality is a result of group-wide coordination onto different equilibria, i.e., coordination failure, whereas Moro and Norman (2004) explains that it is a result of group-wide specialization over different types of jobs. Moro and Norman (2004) are the first to explicitly model discrimination as a "relative" advantage that induces a greater payoff for one group at the expense of another. In their general equilibrium model, the production technology requires two complementary inputs. If too many workers invest in the skills for one input, the marginal product of each input adjusts accordingly. Discrimination is an outcome that two groups specialize in different tasks. Mialon and Yoo (2014) introduce interdependent signaling decisions of applicants to explain discrimination. They find that employers may prefer discriminatory hiring, because it reduces the overall risks of hiring. Fang and Moro (2011) provide an excellent survey of the literature on statistical discrimination.[2]

[2] There are many other studies with interesting approaches to explain discrimination, although

The framework of fixed wage in this paper plays a similar role of "posted wage" in Lang et al.(2008). This commonly-observed labor market practice makes employers unable to adjust their posted wage ex post upon observing many more candidates with a qualifying signal than the available number of positions. Thus, under this framework of posted wages, the competition among workers are expected to be persistent. Lang et al.(2008) focuses on how posted wages amplies slight racial preferences to produce significant racial discrimination and segregation in labor markets.

In order to explain prejudice, the present paper develops the concept of inter-group interactions in Moro and Norman (2004) in the model of strategic and interdependent signaling decisions by the applicants. In this aspect, this paper is closely related to Mialon and Yoo (2014) as they analyze discrimination in the framework of interdependent signaling as well. However, the present paper focuses on showing how intense competition destroys applicants' signaling incentives and leads to prejudice in equilibrium by making the signals worthless, whereas Mialon and Yoo (2014) focuses on the employers' incentive to be engaged in a discriminatory hiring practice.

3 Model

Consider a signaling game between an employer R (the receiver of signals) and n applicants (the senders of signals), $n \geq 2$. An applicant earns $W > 0$ if hired, and nothing, otherwise.[3] Each applicant is characterized by an unobservable ability $A \in \{H, L\}$ and an observable characteristic $\theta \in \Theta = \{1, 2, ...n\}$. We assume that there is one applicant for each θ.[4] An applicant of type θ has a high H ability with

they are not closely related to this paper. For example, Mailath et al (2000) model discrimination as an outcome of search frictions. Chaudhuri and Sethi (2008) discuss the relationship between integration and discrimination. Blume (2005) considers a dynamic model of statistical discrimination to explain how a discriminatory equilibrium is selected. He explains that a discriminatory equilibrium may occur as a result of firms' learning of worker investment decisions. Fryer (2007) also shows an interesting dynamic effect of statistical discrimination. He finds that a discriminated worker who overcomes initial discrimination can actually benefit from the discrimination.

Another strand of literature on discrimination focuses on explaining what makes the group inequality persistent. See, for examples, Dulauf (1996), Benabou (1996), Mookerjee et al (2010), and Bowles et al. (2014).

[3] We take $W > 0$ as exogenously given because prejudice occurs in many non-market situations as well as in markets. Hence, the payoff W from being selected for a scarce position is often a non-market outcome. For example, in the case of college admissions, W is not a market price.

[4] While all of n applicants differ intrinsically, in equilibrium, the number of characteristics θ that are actually used to differentiate the applicants can be much less than n, as shown in Section 5. The current setup allows the maximum number of *differentiable* characteristics. If necessary, the

a probability of $q > 0$. Applicants can make an effort (e) to signal their ability. The effort generates a signal of high ability σ_H with a probability $e \in [0, 1]$ and a signal of low ability σ_L with a probability $1 - e$. Let $C_{\theta A}(e) \in [0, \infty)$ and $mc_{\theta A}(e)$ be the cost and the marginal cost, respectively, of investing in e for an applicant of type θA. $C_{\theta A}(0) = 0$ and for $e > 0$, $C_{\theta L}(e) > C_{\theta H}(e)$. $C_{\theta A}(e)$ is twice differentiable for $e \in (0, 1)$. We assume that $mc_{\theta A}(e) = C'_{\theta A}(e) > 0$, $mc'_{\theta A}(e) = C''_{\theta A}(e) > 0$, and $mc_{\theta L}(e) > mc_{\theta H}(e)$. Without loss of generality, we also assume that θ is ordered by the level of $mc_{\theta L}(e)/mc_{\theta H}(e)$, i.e., $mc_{1L}(e)/mc_{1H}(e) \geq mc_{2L}(e)/mc_{2H}(e) \geq ... \geq mc_{nL}(e)/mc_{nH}(e)$ for $e > 0$.

The payoff for the employer R depends only on the applicant's ability A. R receives a payoff of $V > 0$ for hiring an $H-$type applicant, but receives 0 for selecting a $L-$type applicant. In order to use signals σ_H or σ_L, R must review and extract the information from the applications,[5] whereas θ is readily observable and available. In some cases, R may want to pre-select only a subset Θ_P of applicants for a review based on the observed θ only, $\Theta_P \subset \Theta, \Theta_P \neq \Theta$. In that case, R observes the signals only from the pre-selected type $\theta \in \Theta_P$. In the benchmark case below, we examine whether R might be interested in pre-selecting applicants based on θ in a model of no competition among the n types of θ. Then, in Section 4, we introduce the case of insufficient positions that results in an interaction among the applicants.

Throughout the paper, regardless of whether or not there is an interaction among the applicants, the timing of the game between the employer and the applicants is as follows. At stage 0, nature chooses the type θ and the ability A for each applicant and determines k available positions for the employer. If $k \geq n$, there is no competition. This case is considered in the benchmark model. Then, the main model considers the case when $k < n$. This case $k < n$ represents a situation in which applicants foresee a non-negligible chance of failing to get a position even if they are qualified. At stage 1, the employer determines whether to consider the entire pool of applicants (T) or to review only a subset (S) based on the θ. Without knowing the employer's decision on T or S,[6] applicants decide how much effort e to make to signal their ability. At

employer R is able to find as many characteristics as he needs for discrimination.

[5] This review process may involve a cost of s. In the main analysis, we assume that $s = 0$ or insignificant, for simplicity. We discuss the case when $s > 0$ in Section 6.2.

[6] The choice between T and S in hiring is hidden from the public since employment discrimination and prejudice based on indices such as race and gender is illegal. Even in the case of a non-hiring decision, social norms against prejudice and discrimination often make the information about the decision on T or S unavailable.

stage 2, the signals are realized for a given effort e. If R had chosen S to pre-select a subset Θ_P at stage 1, only the signals from the pre-selected applicants in Θ_P will be observed. If R had chosen T, all applicants' signals will be observed. R determines whether to hire an applicant with a signal σ_H, whereas we assume that R never hires an applicant with the signal σ_L.[7]

3.1 Benchmark

Suppose that there are an unlimited number of positions in the sense that $k \geq n$. In this case, R is willing to select as many applicants as possible after observing the qualifying signal σ_H for any θ.[8] Applicants need not be concerned about how many of other applicants may have the same qualifying signal σ_H, which makes their signaling decisions independent. This case allows us to identify the effect of competition when we introduce the case of limited positions ($k < n$) in the next section.

Let $I_T = 1$ denote the probability that the employer chooses to review the entire applications (T). If S is chosen instead ($I_T = 0$), R pre-selects n^* applicants based on θ, $n^* < n$. Let p_θ be the probability that type θ is pre-selected as one of the n^* applicants. Since $n^* < n$, it has to be that $p_\theta < 1$ for some θ. For an applicant of type θA, the expected payoff from investing in e is $E(e|\theta, A) = e[I_T + (1 - I_T)p_\theta)]W - C_{\theta A}(e)$. The optimal effort level $e_{\theta A} > 0$ for type θA satisfies

$$x_\theta \equiv [I_T + (1 - I_T)p_\theta]W = mc_{\theta A}(e_{\theta A}). \tag{1}$$

x_θ shows signaling incentives for θ. Since $C_{\theta L} > C_{\theta H}$, $mc_{\theta L}(e) > mc_{\theta H}(e)$, and $mc'_{\theta A} > 0$ for $e > 0$, we find that $e_{\theta H} = mc^{-1}_{\theta H}(x_\theta) > mc^{-1}_{\theta L}(x_\theta) = e_{\theta L}$. That is, $H-$type applicants's efforts are higher for any $x_\theta > 0$. If $x_\theta = 0$, $e_{\theta H} = e_{\theta L} = 0$.

[7] This is a behavioral assumption that we make in order to simplify the analysis. An applicant is required to have σ_H to be considered for hiring, although σ_H does not guarantee hiring.

This assumption does not affect the qualitative results of this paper. If any applicant with σ_L can be hired as well, it will overall reduce applicants' incentive to signal given that signaling is costly. In this paper, the main result of prejudice stems from insufficient signaling efforts by the applicants and ineffective signaling. This implies that allowing the possibility of hiring for σ_L makes signaling less informative and less valuable to the employer. Thus, it will only strengthen the result of this paper about equilibrium prejudice.

[8] Note that, in this model, if the qualifying signal σ_H of some θ were not to be used for hiring, R can choose not to acquire the signal in the first place by excluding those θ in the review process by choosing S, rather than not using σ_H after acquiring it. That is, in the present framework, R prioritizes the decision between T or S in determining the value of the signal σ_H.

For given $e_{\theta H}$ and $e_{\theta L}$, the employer R's expected payoff from using full information (T) is

$$E(T) = \sum_{\theta} \underbrace{q e_{\theta H}(V - W) - (1-q) e_{\theta L} W}_{U_\theta} \tag{2}$$

and the expected payoff from using only a subset S is $E(S) = \sum_{\theta} p_\theta U_\theta$, respectively, $p_\theta < 1$ for some θ. U_θ represents the value of signals from type θ. (1) and (2) show that in this benchmark case, U_θ is independent of the efforts chosen by other type θ', $\theta \neq \theta'$. As long as $U_\theta > 0$ for θ, R prefers to use the information from θ. Using (1), we can write that $U_\theta(x_\theta) = q mc_{\theta H}^{-1}(x_\theta)(V - W) - (1-q) mc_{\theta L}^{-1}(x_\theta) W$.

Assumption 1 $\frac{q(V-W)}{(1-q)W} > \frac{mc_{\theta L}^{-1\prime}(W)}{mc_{\theta H}^{-1}(W)}$.

Assumption 2 $\frac{mc_{\theta L}^{-1\prime}(a_1)}{mc_{\theta H}^{-1}(a_1)} > \frac{mc_{\theta L}^{-1\prime}(a_2)}{mc_{\theta H}^{-1}(a_2)}$ *for* $a_2 > a_1$.

Assumption 1 ensures that at the maximum level of signaling incentive, i.e., at $x_\theta = W$, the effort level is sufficiently high that $U_\theta(W) > 0$ for *all* θ. Assumption 2 implies that it is more costly to make an effort for $L-$type than for $H-$type and thus an increase in x_θ increases $e_{\theta H}$ more than it increases $e_{\theta L}$. These assumptions ensure that finding signaling efforts from any type θ is intrinsically valuable to the employer. Then, not exploring all of the information, S, means forgoing a chance to discover valuable information. Thus, as long as the applicants make enough effort, R prefers to use all of the available signaling information from the applicants.

Lemma 1 *Let* x_0 *is the level of incentive that satisfies* $U_\theta(mc_{\theta A}^{-1}(x_0)) = 0$. *For* $k \geq n$, *if the applicants' belief of* p_θ *is high enough that* $p_\theta \geq x_0/W$, *in equilibrium, they make a sufficient effort and the employer acquires all of the signaled information.*

Proof. Proofs are provided in the Appendix. ■

Lemma 1 states that, when there is no shortage of positions, if prejudice occurs, it can only be due to self-fulfilling expectations of a low p_θ, just as traditional statistical discrimination models predict. Clearly, this case does not explain prejudice properly. First of all, given that $k \geq n$, the positions are always available as long as the applicant makes an effort to obtain a qualifying signal. Thus, any equilibrium involving an expectation of $p_\theta < x_0/W$ is not stable. In addition, prejudice against a type θ' applicant particularly requires that type θ' expects $p_{\theta'} < p_\theta$, whereas the other type θ

expects a high enough p_θ, $\theta' \neq \theta$. However, there is no reason why type θ' applicant should believe that only $p_{\theta'}$ is particularly low.

Prejudice (and discrimination) is always a relative concept. Offering a type θ applicant an opportunity to prove his qualification becomes prejudice, only if the same opportunity is not available to a type θ' applicant for no proper reason. This implies that prejudice is R's decision to allocate the opportunities asymmetrically among the n types of applicants, which requires *comparison of the applicants.* Then, the framework in this benchmark is inadequate to explain prejudice, because the employer's decision for each applicant is independent. Therefore, in the next section, we consider a new model in which the employer needs to consider all types simultaneously *in comparison.*

4 Strategic Signaling

Suppose now that $k < n$. For the employer R, in selecting the best k candidates, the value of each applicant's signal is now relatively determined in comparison with other applicants' signals. In order to be selected, an applicant not only need a qualifying signal, but also her signal must be of higher quality than that of her competitors. Thus, applicants' signaling decisions become interdependent. As k declines, competition among the applicants intensifies.

In this set-up, we define that there is prejudice in equilibrium when R strategically chooses S based on θ to forgo reviewing the information of some applicants. When S is chosen, R decides n^* of applicants to be included in a subset Θ_P for a review, $n^* < n$. If $n^* \leq k$, the number of pre-selected θ in the subset Θ_P is smaller than the number of available positions, and thus, any pre-selected θ applicant who has σ_H will be selected. Thus, $p_\theta = 1$ for all $\theta \in \Theta_P$. On the other hand, if $n^* > k$, then, $p_\theta < 1$ for some $\theta \in \Theta_P$.[9]

[9] This means that when S is chosen with $n^* > k$, R has a plan to choose each θ with a *pre-determined* probability $p_\theta < 1$ before he observes each of θ applicant's signal. This setup simplifies our analysis. Alternatively, if we allow R to make a decision after observing the signals from n^* applicants, given that $n^* > k$, it engenders (limited) competition among the pre-selected applicants, which is similar to the case of full information. That is, such a setup makes what follows from the choice of S a lot similar to T. For the main analysis, we use the simpler framework in which S involves no expectation of competition, so that we can easily identify the effect of competition under full information T. However, the main results from the simpler framework can be easily extended to a more general case of pre-selection. We discuss this general framework of pre-selection in Section 6.1. We show that prejudice is expected to be more prevalent under the general type of pre-selection

On the other hand, if R chooses T to review all, and if more than k applicants turn out to have a qualifying signal σ_H, there is competition among the qualified applicants. Let M be the set of the m qualified applicants, $n \geq m > k$. Then, R selects an applicant $\theta \in M$ with a probability $\pi_{\theta m} \in [0, 1]$ based on θ. We define that there is *discrimination* when $\pi_{\theta m} \neq \pi_{\theta' m}$, for any $\theta, \theta' \in M$. If $\pi_{\theta m} > k/m$ for some $\theta \in M$, there is *discrimination* in favor of type θ.

Therefore, prejudice differs from discrimination in the level of information that R has, although in both situations, R refers to the index θ in selecting applicants. There is prejudice when R does not wish to learn more about an applicant's quality if the type is θ', and is only interested in the quality of the type θ applicant, i.e., $\theta \in \Theta_P$, and $\theta' \notin \Theta_P$. Discrimination, on the other hand, occurs only after R acquires all of the information from all applicants and when R treats two applicants who have the same qualifying signals differently.

Example 1. Consider a case of $k = 1$ position and $n = 3$ *ex ante* identical applicants, $\theta \in \{1, 2, 3\}$. Suppose that candidates of type 1 and 3 have a qualifying signal σ_H. If R acquires all of the signals from the three candidates, R will observe that the two candidates ($m = 2$) have σ_H. Without discrimination, R selects each of the two with a probability of 1/2. Alternatively, R may select the type 1 applicant with $\pi_{1m} = 1$ and disregard type 3 (i.e., $\pi_{3m} = 0$). In the latter, although there is no prejudice, there is discrimination against type 3. On the other hand, prejudice occurs if R pre-selects only $n^* = 1$ applicant, say type 1, to review with $p_1 = 1$ and $p_2 = p_3 = 0$. Consequently, R will never be able to discover if type 3 also has σ_H.

4.1 Applicants' problem

For example, consider the problem of applicant of type $1H$ when $n = 3$ and $k = 1$. If R chooses T to review all ($I_T = 1$), $1H$ with a qualifying signal faces competition when $m > k = 1$ and expects to be selected with a π_{1m} probability, whereas she expects to be selected for sure if she is the only one with σ_H. If R chooses S ($I_T = 0$), $1H$ expects to be selected with a $p_1$probability. Let Ψ_2 and Ψ_3 denote the probabilities that the type 2 and 3 applicants have a qualifying signal σ_H, respectively. Then, $\Psi_\theta = qe_{\theta H} + (1 - q)e_{\theta L}$, $\theta = 2, 3$. For given I_T, p_1,and π_{1m}, the expected payoff for a

that allows limited competition after S.

type $1H$ applicant is

$$E(e_{1H}) = e_{1H}W\left[\begin{array}{c} I_T\{\Psi_2\Psi_3\pi_{13} + (\Psi_2(1-\Psi_3) + (1-\Psi_2)\Psi_3)\pi_{12} \\ +(1-(\Psi_2+\Psi_3-\Psi_2\Psi_3))\} + (1-I_T)p_1 \end{array}\right] - C_{1H}(e_{1H}). \quad (3)$$

Generalizing this, we can see that for any given m applicants with a qualifying signal σ_H, it is either (i) $m > k$, or (ii) $m \leq k$. When $m > k$, an applicant remains to face competition even after obtaining σ_H. On the other hand, if $m \leq k$, there are at most $k-1$ other applicants with the qualifying signal σ_H, and thus, the applicant is guaranteed for selection as long as she has σ_H.

When $m > k$, let $\Psi_{-\theta Cm}$ denote the probability that a type θ applicant with σ_H faces competition with $m-1$ other applicants. In the above example of $k=1$, the applicant $1H$ faces competition either if all 3 applicants have σ_H ($m=3$) or two of them have σ_H ($m=2$). If $m=3$, $\Psi_{-1Cm} = \Psi_2\Psi_3$, and when $m=2$, $\Psi_{-1Cm} = \Psi_2 + \Psi_3 - 2\Psi_2\Psi_3$. Then, the expected payoff for a type θA applicant is

$$E(e_{\theta A}) = e_{\theta A}W\left[I_T\left\{\sum_{m=k+1}^{n}(\Psi_{-\theta Cm}\cdot\pi_{\theta m}) + \left(1-\sum_{m=k+1}^{n}\Psi_{-\theta Cm}\right)\right\} + (1-I_T)p_\theta\right] - C_{\theta A}(e_{\theta A}). \quad (4)$$

Note that the probability of competition $\sum_{m=k+1}^{n}\Psi_{-\theta Cm}$ increases as k decreases, other things being equal. If $k=3$, for example, $\sum_{m=k+1}^{n}\Psi_{-\theta Cm}$ is a probability that there are at least three other applicants with σ_H. If $k=2$, it is a probability that there are at least two other applicants with σ_H.

The optimal effort level $e^*_{\theta A} > 0$ satisfies

$$x_{\theta C} \equiv \left[I_T\underbrace{\left\{\sum_{m=k+1}^{n}(\Psi_{-\theta Cm}\cdot\pi_{\theta m}) + \left(1-\sum_{m=k+1}^{n}\Psi_{-\theta Cm}\right)\right\}}_{w_\theta} + (1-I_T)p_\theta\right]W = mc_{\theta A}(e^*_{\theta A}), \quad (5)$$

for $x_{\theta C} > 0$. If $x_{\theta C} = 0$, $e^*_{\theta H} = e^*_{\theta L} = 0$. Since $\Psi_{-\theta Cm}$ depends on the signaling efforts $e_{-\theta A}$ by other competing applicants $-\theta$, (5) can be rewritten as

$$\begin{aligned} x_{\theta C} &\equiv [I_T w_\theta(e_{-\theta C}) + (1-I_T)p_\theta]\,W \text{ and} \quad (6)\\ e^*_{\theta A} &= mc^{-1}_{\theta A}(x_{\theta C}(e_{-\theta C})). \end{aligned}$$

Comparing (1) and (6), we can show the impact of expected competition on each applicant's signaling incentive $x_{\theta C}$. From (1), when each applicant chose signaling efforts independently, the signaling incentive was $x_\theta = [I_T + (1 - I_T)p_\theta]W$. The signaling incentive is now weighted by $w_\theta \leq 1$ in the case of facing competition ($I_T = 1$), considering the fact that the selection is not guaranteed for applicant θ ($\pi_{\theta m} \leq 1$) even after obtaining the qualifying signal. To compare x_θ and $x_{\theta C}$, suppose that p_θ is the same in both cases. Since $\pi_{\theta m} \leq 1$, $w_\theta \leq 1$, we find that $x_{\theta C} \leq x_\theta$ always, with a strict inequality for at least one θ with $\pi_{\theta m} < 1$. Therefore, other things being equal, the applicants' incentive to signal is lower in the case of competitive signaling.

Proposition 1 *Other things being equal, competition for signaling among the applicants reduces their incentive to signal.*

When $k < n$, one applicant's signaling effort imposes a negative externality on the other applicants. An applicant θ with σ_H is guaranteed for selection only if there are at most $k - 1$ other applicants with σ_H. However, if more than $k - 1$ applicants have σ_H, the applicant can receive the return from her signaling only with a $\pi_{\theta m} \leq 1$ probability. As a lower k increases the probability of competition and makes it more difficult to recoup the return from signaling, applicants' incentive to signal decreases (i.e., a lower $x_{\theta C}$).

On the other hand, (6) shows that such a competitive incentive disappears if R does not use full information ($I_T = 0$). The selection is guaranteed as long as an applicant is one of the pre-selected n^*. Type θ applicant expects $x_{\theta C} = p_\theta W$ with a p_θ chance of pre-selection. As a result, an applicant's best response becomes independent of the efforts of the other applicants, eliminating the effect of strategic signaling. This reveals that one of the most important functions of the pre-selection S is to remove the impact of competition among the applicants. Thus, if competition among the applicants destroys their incentive to signal and the value of signaling, R has an incentive to choose S in order to eliminate the impact of competition. We show this in Section 5.

4.2 Employer's problem

4.2.1 Comparative Discrimination

First, consider R's decision concerning the optimal $\pi^o_{\theta m}$ after acquiring all applicants' information, if more than k applicants turn out to have σ_H. For example, suppose that $n = 3, k = 2$, and R observes that all three applicants $\theta = 1, 2, 3$ have σ_H. R's payoff from selecting type 1 is $E(1|\sigma_H,\sigma_H,\sigma_H) = [\{q^3 e_{1H}e_{2H}e_{3H} + q^2(1-q)(e_{1H}e_{2H}e_{3L} + e_{1H}e_{2L}e_{3H}) + (1-q)^2 q(e_{1H}e_{2L}e_{3L})\}/\Delta_{HHH}]V - W$, where $\Delta_{HHH} = \Psi_1 \times \Psi_2 \times \Psi_3$ and $\Psi_\theta = qe_{\theta H} + (1-q)e_{\theta L}$. Then, $E(1|\sigma_H,\sigma_H,\sigma_H) > E(2|\sigma_H,\sigma_H,\sigma_H)$ if

$$\frac{q(1-q)}{\Delta_{HHH}}\left[\Psi_3(e_{1H}e_{2L} - e_{2H}e_{1L})\right] > 0. \tag{7}$$

That is, type 1 is preferable to type 2 if R believes that $e_{1L}/e_{1H} < e_{2L}/e_{2H}$.[10] The condition implies that type 1's signal σ_H is more likely from an $H-$type than is type 2's signal σ_H. Thus, when many applicants have equally qualifying signals, R hires the applicants who have a lower probability of signaling errors. If R expects $e_{1L}/e_{1H} < e_{2L}/e_{2H} < e_{3L}/e_{3H}$, for example, type 1 and type 2 applicants will be chosen for the $k = 2$ positions.

In general, when m applicants have σ_H and $m > k$, a type θ applicant is preferable to type θ' if $e_{\theta L}/e_{\theta H} < e_{\theta' L}/e_{\theta' H}$, for $\theta \neq \theta'$. Since $mc_{1L}/mc_{1H} \geq mc_{2L}/mc_{2H} \geq ... \geq mc_{nL}/mc_{nH}$, other things being equal, the employer's rational belief about the level of $e_{\theta L}/e_{\theta H}$ is ordered as $e_{1L}/e_{1H} \leq e_{2L}/e_{2H} \leq ... \leq e_{nL}/e_{nH}$. Then, among the m qualified applicants in the set M, there exists a kth applicant whose e_L/e_H is ranked at k. Let M_k be that applicant and $(e_L/e_H)_{M_k}$ be the level of expected effort ratio of the applicant. Then, $e_{iL}/e_{iH} \leq (e_L/e_H)_{M_k} \leq e_{jL}/e_{jH}$, for i, $j \in M$. If there is no tie with the M_k, R selects all $i \in M$ and M_k with a probability of 1. If there is a tie with the M_k, R randomizes among the applicants that are tied.

Lemma 2 *Suppose that $m > k$ applicants have σ_H. The employer's optimal strategy $\pi^o_{\theta m}$ is characterized as follows.*

$$\begin{array}{ll} \pi^o_{\theta m} = 1 & \text{if } \frac{e_{\theta L}}{e_{\theta H}} < (\frac{e_L}{e_H})_{M_k} \\ \pi^o_{\theta m} \in [0,1] & \text{if } \frac{e_{\theta L}}{e_{\theta H}} = (\frac{e_L}{e_H})_{M_k} \quad \text{for } \theta \in M. \\ \pi^o_{\theta m} = 0 & \text{if } \frac{e_{\theta L}}{e_{\theta H}} > (\frac{e_L}{e_H})_{M_k} \end{array} \tag{8}$$

[10] Here, $e_{1H} > 0$, $e_{2H} > 0$ since σ_H is observed from both applicants.

Example 2. Suppose $n = 4, k = 2$, and $\theta = 1, 3, 4$ have σ_H (i.e., $m = 3$). Then, $(e_L/e_H)_{M_2} = e_{3L}/e_{3H}$. If $e_{1L}/e_{1H} < e_{3L}/e_{3H} < e_{4L}/e_{4H}$, then the employer selects 1 and 3, i.e., $\pi^o_{13} = \pi^o_{33} = 1$, $\pi^o_{43} = 0$. If $e_{1L}/e_{1H} < e_{3L}/e_{3H} = e_{4L}/e_{4H}$, then the employer selects 1 and randomizes between 3 and 4, i.e., $\pi^o_{13} = 1$, $\pi^o_{33} + \pi^o_{43} = 1$.

What determines R's belief over $\frac{e_{\theta L}}{e_{\theta H}}$ is the innate difference in the cost of signaling for each θ applicant. A type θ with a greater difference in signaling ability between $L-$type and $H-$type is more likely to have a lower ratio. For instance, suppose that $H-$type females face much higher costs than do $H-$type males in signaling because of a social environment that encourages females' gender role as a housewife and discourages their labor market participation. Then, males' signal σ_H is likely to be a better indication of the $H-$type than females' signal σ_H. That is, the employer should expect a lower $\frac{e_L}{e_H}$ for males than females, other things being equal. Thus, when both male and female applicants have equally qualifying signals, the employer would favor males.

Instead, suppose that $\frac{e_{\theta L}}{e_{\theta H}} = \frac{e_{kL}}{e_{kH}}$ for all θ. In this case, $\pi^o_{\theta m} \in [0, 1]$. There is comparative discrimination if $\pi^o_{\theta m} \neq k/m$ for some θ. Comparative discrimination differs from statistical discrimination in that it requires all of the competing applicants to be *compared* in the *relative* strength of their signals and to have *equally qualifying signals*, whereas in the case of statistical discrimination, applicants are not compared and often do not have equally qualifying signals.

With this optimal strategy $\pi^o_{\theta m}$, at stage 1, R's expected payoff from acquiring the signaling information from all types is

$$E(T) = \sum_{\theta} \left\{ \underbrace{\sum_{m=k+1}^{n} \Psi_{-\theta C m} \pi^o_{\theta m} + (1 - \sum_{m=k+1}^{n} \Psi_{-\theta C m})}_{w_\theta} \right\} U_\theta = \sum_{\theta} w_\theta U_\theta \qquad (9)$$

where $U_\theta = q e_{\theta H}(V - W) - (1 - q) e_{\theta L} W$. (9) shows how competition affects the employer's problem. Compared to (2) under no competition, the value of type $\theta's$ signal U_θ is now weighted by w_θ. Since $k < n$, even if all signals are reviewed, some of the qualified θ cannot be hired ($\pi^o_{\theta m} < 1$). Thus, $w_\theta < 1$ for some θ. As k decreases, the weight w_θ decreases, since the chance of competition $\sum_{m=k+1}^{n} \Psi_{-\theta C m}$ increases. Thus, competition reduces the value of signals for the employer.

4.2.2 Prejudice

Now consider the case that R only selects a subset Θ_P to review at Stage 1. For given choices of n^* and p_θ to select Θ_P, the employer's payoff for acquiring the information only from Θ_P is

$$E(S) = \sum_\theta p_\theta U_\theta, \tag{10}$$

where $p_\theta = 0$ for $\theta \notin \Theta_P$. Compared to (9), (10) shows that in the case of prejudice, p_θ replaces w_θ for the weight of U_θ. $E(S)$ increases as the n^* of the selected applicants in Θ_P have a higher U_θ, whereas a higher n^* lowers the weight p_θ on U_θ. Hence, R will choose p_θ to maximize the allocation of the weight on the highest U_θ while using n^* to maximize the benefits of including as many of applicants with a high U_θ as possible. Given that $mc_{1L}/mc_{1H} \geq mc_{2L}/mc_{2H} \geq ... \geq mc_{nL}/mc_{nH}$, other things being equal, R's belief is $e_{1L}/e_{1H} \leq e_{2L}/e_{2H} \leq ... \leq e_{nL}/e_{nH}$. Since $U_\theta = qe_{\theta H}(V - W) - (1 - q)e_{\theta L}W$ is decreasing in $e_{\theta L}/e_{\theta H}$, the employer expects that $U_1 \geq U_2 \geq ... \geq U_n$. Hence, R chooses $\theta = 1, 2, ..n^*$ with $p_\theta > 0$ and eliminates the rest of type θ' applicants by allocating $p_{\theta'} = 0$, for $\theta' = n^* + 1, n^* + 2, ..n$.

Now consider the choice of n^*. If there is no tie for U_θ, $U_{k-1} > U_k > U_{k+1}$, and thus, it is always best to choose the first k applicants (i.e., $n^* = k$) by allocating $p_\theta = 1$ for $\theta = 1, 2, .., k$ and $p_{\theta'} = 0$ for $\theta' > k$. If there is a tie with the kth applicants, R may randomize among the applicants who are tied.

Lemma 3 *Suppose that a pre-screening based on θ takes place. The employer selects $n^* \geq k$ applicants as follows.*

$$\begin{array}{ll} p^o_\theta = 1 & \textit{if } \frac{e_{\theta L}}{e_{\theta H}} < \frac{e_{kL}}{e_{kH}} \\ p^o_\theta \in [0, 1] & \textit{if } \frac{e_{\theta L}}{e_{\theta H}} = \frac{e_{kL}}{e_{kH}} \\ p^o_\theta = 0 & \textit{if } \frac{e_{\theta L}}{e_{\theta H}} > \frac{e_{kL}}{e_{kH}} \end{array} \tag{11}$$

Example 3. Suppose that $n = 3$, $\theta = 1$, 2, 3, $k = 2$, and $mc_{3H}(e) = mc_{2H}(e) > mc_{1H}(e)$ while the cost for L-type is the same across all types. Then, for any given x, $e_{3H} = e_{2H} = mc^{-1}_{2H}(x) < mc^{-1}_{1H}(x) = e_{1H}$, and thus, $\frac{e_{3L}(x)}{e_{3H}(x)} = \frac{e_{2L}(x)}{e_{2H}(x)} > \frac{e_{1L}(x)}{e_{1H}(x)}$. This means that $U_1 > U_2 = U_3$ for a given x. Then, R pre-selects type 1 with $p^o_1 = 1$, and randomizes between types 2 and 3 with $p^o_2 + p^o_3 = 1$.

5 Equilibrium

5.1 *Ex Ante* Identical Applicants

When applicants are ex ante identical in the sense that $\frac{mc_{\theta L}}{mc_{\theta H}}(e) = \frac{mc_{\theta' L}}{mc_{\theta' H}}(e) = \frac{mc_L}{mc_H}(e)$ for all $\theta \neq \theta'$, other things being equal, it is expected that $e_{\theta L}/e_{\theta H} = e_{\theta' L}/e_{\theta' H}$. In this case, based on Lemma 2, the equilibrium depends on the applicants' expectations over $\pi^o_{\theta m}$. We first consider the case that the applicants expect no discrimination.

5.1.1 When a non-discriminatory $\pi^o_{\theta m}$ is expected

Suppose that applicants expect equal treatment after R reviews all of the applications (T). Under the expectation of $\pi^o_{\theta m} = \pi^o_{\theta' m} = k/m$, all applicants have an equal incentive to signal and expect the same probability of encountering competition with other applicants, $\Psi_{-\theta Cm} = \Psi_{-\theta' Cm} = \Psi_{-Cm}$. Let x^*_E be the level of the incentive under T for all θ. Then, from (5), for all θ,

$$
\begin{aligned}
x^*_E &= w_E W, \text{and} && (12) \\
w_E &= 1 - \sum_{m=k+1}^{n} \Psi_{-Cm}\left(1 - \frac{k}{m}\right) < 1. && (13)
\end{aligned}
$$

Let $e^*_E = (e_H(x^*_E),\, e_L(x^*_E))$ denote the vector of effort levels by the applicants at x^*_E. Then, at e^*_E, R's expected payoff from full information T is

$$
\begin{aligned}
E(T, \pi^o_{\theta m} &= k/m,\, e^*_E) = n w_E U_E, && (14) \\
\text{where } U_E &= q e_H(x^*_E)(V - W) - (1 - q) e_L(x^*_E) W. && (15)
\end{aligned}
$$

On the other hand, $R's$ payoff of using partial information from the pre-selected, S, and randomizing among the first n^* applicants with p^o_θ, $k \leq n^* < n$, and $p^o_{\theta'} = 0$ for other θ' is

$$E(S, p^o_\theta, e^*_E) = kU_E. \quad (16)$$

Suppose now that applicants expect S. For a pre-selected type $\theta \in \Theta_P$, the probability of being selected is p^o_θ and the incentive to signal is $x^p_{\theta C} = p^o_\theta W$. All other types of applicants $\theta' \notin \Theta_P$ do not make an effort since $x^p_{\theta' C} = 0$ for θ'. Let $e_P = (e_{\theta H}(x^p_{\theta C}),\, e_{\theta L}(x^p_{\theta C}),\, e_{\theta' H}(0) = e_{\theta' L}(0) = 0)$ be the vector of applicants' effort levels. At e_P, R's expected payoff from hiring type θ only is,

$$E(S, p^o_\theta, e_P) = kU_p, \tag{17}$$
$$\text{where } U_P = qe_{\theta H}(x^p_{\theta C})(V-W) - (1-q)e_{\theta L}(x^p_{\theta C})W. \tag{18}$$

If R chooses T, given that only type $\theta \in \Theta_P$ makes efforts, competition occurs only if $n^* > k$ and only within the n^* of θ applicants. Then, for $n^* \geq k$,

$$E(T, \pi^o_{\theta m} = k/m) = n^* w_\theta(e_P)U_p, \tag{19}$$

where $w_\theta(e_P) = 1 - \sum_{m=k+1}^{n^*} \Psi_{-Cm}(1-k/m) < 1$ for $n^* > k$ and $w_\theta(e_P) = 1$ if $n^* = k$. Proposition 2 summarizes the equilibrium prejudice in this case.

Proposition 2 *Suppose all applicants are ex ante identical and they expect equal treatment under full information.*

1. $n^* = k$, $p^o_\theta = 1$, $p^o_{\theta'} = 0$, and $x^p_{\theta C} = W > x^*_E > x^p_{\theta' C} = 0$, for $\theta \in \Theta_P$, $\theta' \notin \Theta_P$.
2. *If* $U_E < 0$, *the unique equilibrium involves prejudice and only the application from type* $\theta \in \Theta_P$ *is considered for a review. The equilibrium value of signals under prejudice is* kU^*_p, where $U^*_p = qmc_H^{-1}(W)(V-W) - (1-q)mc_L^{-1}(W)W$.
3. *If* $U_E > 0$, *multiple equilibria exist. In the equilibrium of full information, all applicants are reviewed and the applicants with* σ_H are *hired with a probability of* k/m if $m > k$, or a probability of 1, otherwise. *Prejudice may occur in equilibrium as a result of self-fulfilling expectation of prejudice.*

Example 4. Consider $n = 3$ applicants, $\theta = 1, 2, 3$, when $k = 2$. Expecting that $\pi^o_{1m} = \pi^o_{2m} = \pi^o_{3m} = 2/3$ if $m = 3$, each applicant makes the same level of effort. Each has the same probability Ψ of generating a signal σ_H, and expects $x^*_E = (1 - 1/3\Psi^2)W < W$. Then, $E(T) = 3(1 - 1/3\Psi^2)U_E(x^*_E)$, whereas with $n^* = 2$, $E(S) = 2U_E(x^*_E)$. Since $3(1 - 1/3\Psi^2) > 2$ for $\Psi^2 < 1$, as long as $U_E(x^*_E) > 0$, T can be chosen in equilibrium. However, if $U_E(x^*_E) < 0$, T is never chosen in equilibrium, and prejudice occurs. Under S, $p^o_1 = p^o_2 = 1$, $p^o_3 = 0$, and $x^p_{1C} = x^p_{2C} = W$ while $x^p_{3C} = 0$, which results in $E(S, x^p_{1C} = x^p_{2C} = W) = U_1(W) + U_2(W) > 0$.

For the type θ who expects to be favored under prejudice, the incentive for signaling is higher under prejudice, i.e. $x^p_{\theta C} = W > x^*_E$. Prejudiced reviews optimize the signaling outcomes from the pre-selected applicants. However, since only few applicants are pre-selected based on θ (which is unrelated to the ability), ex ante, the employer has to bear the risk of selecting unqualified applicants among θ and missing the chance of finding the $H-$types among the unselected. The risk can be avoided by delaying the selection after observing more informative signal σ_H under full information T. However, the information is obtained at the expense of its quality because open competition for everyone increases the uncertainty in getting the position for everyone and lowers their signaling incentives. Hence, the employer benefits from T only if the expected quality of the information is high enough ($U_E > 0$).

Proposition 2 shows that prejudice is an outcome of low expected quality of signals under T. What reduces U_E facilitates prejudice. For example, intense competition or inefficient signaling technology facilitate prejudice by lowering U_E. When only few positions are available, an equal opportunity for everyone means a very slim chance of getting the position for every applicant. This may result in an insufficient effort from any applicant, and make the signal not so informative to the employer. This is likely to occur if signaling is very costly.

Proposition 3 *Suppose that applicants are ex ante identical and they expect equal treatment under full information. Other things being equal, prejudice is more likely to occur as k, V, or q decreases and as the cost of signaling increases.*

A decrease in k increases competition and lowers the chance of being selected upon facing competition. Thus, it lowers signaling incentives x^*_E (or w_E). Since U_E is an increasing function of the incentive x^*_E (w_E), a decrease in k lowers the value of signals U_E and thus more likely to induce prejudice. Similarly, a low V promotes prejudice. A low V implies that the value of skills that $H-$type offers is not much greater than the value of skills that $L-$type offers. Hence, prejudice is more likely to be prevalent in an occupation that does not require specific skills.

On the other hand, even if full information occurs in equilibrium, it is not clear whether the expected value of signals under T in equilibrium would be higher than that of S. It could be that $E(T, \pi^o_{\theta m} = k/m,\ e^*_E) = nw_E U_E(x^*_E) < E(S, p^o_\theta = 1,$ $e_P) = kU_p(W)$, if $U_E(x^*_E)$ is not high enough since $U_E(x^*_E) < U_p(W)$.

The results in Propositions 2 and 3 imply that equal opportunity policies need to be implemented with a caution. A policy that promotes equal opportunities basically enforces full information T in this framework by setting the expectation of unprejudiced review of applications. Such a policy will be efficient only if spreading out the signaling incentives to everyone does not destroy the value of signaling, i.e. $U_E > 0$. If $U_E < 0$, in order to make the policy efficient, there have to be other policies that enhance the value of U_E as well, such as policies that lower the cost of signaling.

5.1.2 When discrimination is expected

Because there are only k positions available, applicants may expect that, even if they have the same qualifying signals, they may not be treated equally. Suppose that after observing equal signals, R makes selections according to his preferences based on the order of θ. Let i and j be the favored and the discriminated applicants, respectively, $i \leq k < j$. The i types expect that $\pi^o_{im} > \pi^o_{jm}$ under T, for $i, j \in M$. This implies that other things being equal, $x_{iC} > x_{jC}$ and thus, $\frac{e_{iL}}{e_{iH}} < \frac{e_{jL}}{e_{jH}}$. Then, all j types reduce their efforts. This reduces type i's probability of facing competition, $\sum_{m=k+1}^{n} \Psi_{-iCm}$, and thus increases their signaling efforts further. Thus, $\pi^o_{im} = 1 \geq \pi^o_{jm}$. All i types expect to be selected for sure if they obtain σ_H, despite competition.

In contrast, j types with σ_H will be selected in competition against other types only if not many applicants have the qualifying signal ($m \leq k$) or not enough of the qualifying signals are from more preferred types $\theta \leq j-1$ when $m > k$. Let $I_{jmk} = 1$ be an index for the cases in which j has a qualifying signal and no more than $k-1$ of the m signals are from the types who are more preferred than j, when $m > k$. In those cases, $j \leq M_k$ and, thus, the applicant j is selected for sure. Otherwise, j is never chosen when $I_{jmk} = 0$. Since j has a greater chance to be selected than $j+1$ upon facing competition, the weight w_j depends on the preference order j. Therefore, for all $i \leq k < j$,

$$\begin{aligned} x_{iC} &= W > x_{jC} = w_j W, \\ \text{where } w_j &= \sum_{m=k+1}^{n} \Psi_{-jCm} I_{jmk} + \left(1 - \sum_{m=k+1}^{n} \Psi_{-jCm}\right) < 1. \end{aligned} \tag{20}$$

Let e^d_{jH} and e^d_{jL} be the levels of effort by j and $e^*_{UE} = (e_{iH},\, e_{iL},\, e^d_{jH},\, e^d_{jL})$ be the vector of the effort levels by i and j applicants when the applicants expect full information T. Then, $e^d_{jH} > 0$ and $e^d_{jL} > 0$ since $x_{jC} > 0$ due to the chance of no competition $(1 - \sum_{m=k+1}^{n} \Psi_{-jCm} > 0)$.

$$\begin{aligned} E(T, \pi^o_{im} &= 1, \pi^o_{jm} = I_{jmk}, e^*_{UE}) = kU^*_p + \sum_{j=k+1}^{n} w_j U_j(e^d_{jH}, e^d_{jL}) \\ &\gtrless kU^*_p = E(S, p^o_i = 1, p^o_j = 0, e^*_{UE}). \end{aligned} \tag{21}$$

Note that the value of signal from the favored type i is the same as U^*_p in Proposition 2. This is because type $i's$ signaling incentive x_{iC} is the same as the incentive of the pre-selected θ under S, i.e., $x_{iC} = W = x^p_{\theta C}$. Type i's incentive $x_{iC} = W$ is independent of others' efforts and free from competition. Thus, discrimination makes the favored type i applicants immune to competition by treating them like the pre-selected under prejudice. Yet, by keeping the opportunies open to type j, full information with discrimination generates extra values of signals $\sum_{j=k+1}^{n} w_j U_j(e^d_{jH}, e^d_{jL})$. The question is whether this value is high enough, i.e., $\sum_{j=k+1}^{n} w_j U_j(e^d_{jH}, e^d_{jL}) > 0$. (21) shows that the full information T cannot occur in equilibrium if the value of signals from the discriminated type j applicants is not high enough.

On the other hand, when the applicants expect prejudiced pre-selection S, j types do not expect to be reviewed at all since $p^o_j = 0$, implying that $x^p_{jC} = 0$. Thus, they make no effort. Type i applicants' incentives for signaling will be the same $x_{iC} = W$ as under T. The employer expects $e_P = (e_{iH}(W), e_{iL}(W), e_{jH} = e_{jL} = 0)$, and

$$\begin{aligned} E(T, \pi^o_{im} = 1, \pi^o_{jm} = I_{jmk}, e_P) &= kU^*_p \\ &= E(S, p^o_i = 1, p^o_j = 0, e_P). \end{aligned} \tag{22}$$

Proposition 4 *Suppose that all applicants are ex ante identical while they expect discriminatory $\pi^o_{\theta m}$ under full information.*

1. *If $\sum_{j=k+1}^{n} w_j U_j(e^d_{jH}, e^d_{jL}) < 0$, the unique equilibrium results in prejudice against type j and only type i applicants are considered.*

2. *If $\sum_{j=k+1}^{n} w_j U_j(e^d_{jH}, e^d_{jL}) > 0$, multiple equilibria exist. The equilibrium value of signals is higher under full information T than under prejudice.*

If discrimination is expected under T, prejudice occurs in equilibrium when the value of signals from the discriminated j is low, $\sum_{j=k+1}^{n} w_j U_j(e^d_{jH}, e^d_{jL}) < 0$. This is when the signaling incentive for type j, x_{jC} (or w_j), is too low. From (20), if $I_{jmk} = 1$, $x_{jC} = W$, $U_j(e^d_{jH}, e^d_{jL}) > 0$, and the signal from the discriminated j is of value. However, type j can be selected out of m qualifiers ($I_{jmk} = 1$) only if fewer than k are from the favored $\theta \leq j-1$. This means that as j gets close to n, j's chance becomes quite slim. This is particularly true if k is small, in which case most of the applicants belong to the unfavored j and they expect no selection in most cases. Such an expectation lowers $j's$ signaling incentive, making full information T worthless.

Prejudice in this case means excluding the applicants to be discriminated from the outset. The employer excludes j because they do not make sufficient efforts due to the expectation of discrimination. However, from (21) and (22), it is clear that if T is achievable in equilibrium, the *equilibrium* payoff of T is higher than that of S. This is because T permits more use of informative signals from applicants than S and there is always a chance that some of the favored i type applicants do not produce a qualifying signal. Thus, having type j applicants motivated to make sufficient level of efforts enhances efficiency. In order to eliminate prejudice, discriminated group $j's$ signaling incentives need to increase sufficiently.

An interesting question is whether promoting nondiscrimination (i.e., equal $\pi^o_{\theta m}$) reduces prejudice. The main difference between the cases of unequal $\pi^o_{\theta m}$ and equal $\pi^o_{\theta m}$ is in how the risk of no-selection is distributed across applicants. Under discrimination, the favored group i receives a higher incentive $x_{iC} = W$ and the discriminated group j receives a lower incentive $x_{jC} = w_j W < W$. In contrast, under non-discrimination, all applicants (i or j) have an equal incentive $x^*_E = w_E W < W$. Nondiscrimination makes all applicants equally share the risk of non-selection, whereas unequal treatment imposes the risk only onto the discriminated group j. Nondiscrimination makes every applicant's signal equally weak, whereas discrimination weakens the signals from j only. Since it is the discriminated group $j's$ effort level that determines prejudice, if x^*_E is high enough that $\sum_j w_E U_E(x^*_E) > \sum_j w_j U_j(e^d_{jH}, e^d_{jL})$, then the discriminated j's signals have higher quality under nondiscrimination. In this case, non-discriminatory policy effectively reduces prejudice.

However, it is not clear whether $x^*_E > x_{jC}$ on average. Other things being equal, under discrimination, a j close to k has a greater chance of selection $x_{jC} > x^*_E$, whereas for a j close to n, there is almost no chance of selection, $x_{jC} < x^*_E$. Which one dominates will depend on k and n. Also, the effect of spreading the incentives on competition Ψ_{-jCm} depends on the signaling cost function and other parameters such as W and q. Hence, the effect of promoting nondiscrimination (equal $\pi^o_{\theta m}$) on prejudice is ambiguous. Nevertheless, we can conclude that nondiscriminatory policy can have a meaningful effect in reducing prejudice only if the equal distribution of risks does not destroy the value of signals.

Corollary 1 *Nondiscriminatory policy can reduce prejudice only if $U_E > 0$. If $U_E < 0$, nondiscriminatory policy facilitates prejudice.*

5.2 Heterogeneous Applicants

Suppose that $\frac{mc_{iL}}{mc_{iH}}(e) > \frac{mc_{jL}}{mc_{jH}}(e)$ for $i = 1, 2, ...k$ and $j = k+1, k+2, ..n$, for any given e. The outcome in this case is similar to that in the case of discrimination in Section 5.1.2. The only difference in this case is that for the same level of signaling incentive, U_θ is different for $\theta = i$ and $\theta = j$, since their signaling costs are different. Under full information T, when σ_H was observed from more than k applicants, the employer favors i group over j group, and thus, $\pi^o_{im} = 1$, and $\pi^o_{jm} = I_{jmk}$. Let $e^h = (e^h_{iH}, e^h_{iL}, e^h_{jH}, e^h_{jL})$ be the vector of effort levels by the applicants. Then,

$$\begin{aligned} E(T, \pi^o_{im} &= 1, \pi^o_{jm} = I_{jmk}, e^h) \\ &= \sum_{i=1}^{k} U_i(e^h_{iH}, e^h_{iL}) + \sum_{j=k+1}^{n} w^h_j U_j(e^h_{jH}, e^h_{jL}), \end{aligned} \tag{23}$$

where w^h_j is w_j evaluated at e^h. This is lower than $E(S, p^o_i = 1, p^o_j = 0, e^h) = \sum_{i=1}^{k} U_i(e^h_{iH}, e^h_{iL})$ if $\sum_{j=k+1}^{n} w^h_j U_j(e^h_{jH}, e^h_{jL}) < 0$. On the other hand,

$$E(S, p^o_i = 1, p^o_j = 0, e^h_p) = \sum_{i=1}^{k} U_i(e^h_{iH}, e^h_{iL}) = E(T, \pi^o_{im} = 1, \pi^o_{jm} = 0, e^h_p), \tag{24}$$

where $e^h_p = (e^h_{iH}, e^h_{iL}, e^h_{jHP} = e^h_{jLP} = 0)$ is the vector of effort levels under the expectation of S. Prejudice occurs in equilibrium when $\sum_{j=k+1}^{n} w^h_j U_j(e^h_{jH}, e^h_{jL}) < 0$.

Compared to the case of identical costs, heterogenous costs amplify the effect of unequal treatment. Given the cost advantage, type i expects unequally favorable treatment, and receives a greater incentive to signal than type j. In addition, the more efficient signaling technology available to i makes it easier for type i to increase signaling efforts. Combined with the higher incentive for signaling, this effect strengthens the quality of type $i's$ signal. Similarly, the cost disadvantage for type j lowers the quality of their signal further.

Suppose that the heterogeneity in the costs of signaling takes the form of a mean-preserving spread of the cost for identical applicants. Then, on average, the signaling technology of type i applicants is more efficient and the technology of type j applicants is less efficient than that of the applicants with identical costs. In particular, this implies that in order to make the same level of signaling efforts, the type j applicants with an inefficient cost technology have to incur a higher cost than what they would have had with identical costs. Thus, the type j's effort is lower when the costs are different. As a result, the value of signals from the discriminated type j is lower in the case of heterogeneous costs than in the case of identical costs, i.e., $\sum_{j=k+1}^{n} w_j^h U_j(e_{jH}^h, e_{jL}^h) < \sum_{j=k+1}^{n} w_j U_j(e_{jH}^d, e_{jL}^d)$. Hence, prejudice is more likely to occur when the costs of signaling differ across the applicants.

Simply put, if signaling itself is not easy for some applicants, quality of their signals is low and that makes prejudice more prevalent. In this case, prejudice may not easy to reduce unless the level of support for the disadvantaged measures up proportionally to the level of disadvantage they face. That is, effective anti-discrimination policies may need to be non-neutral in this case. We close this section with a brief discussion on this policy implication using a simple example in the following.

Example 5. Suppose that there are $n = 3$ applicants, $\theta = 1, 2, 3$, $k = 2$, and $e_{3L}(x)/e_{3H}(x) > e_{2L}(x)/e_{2H}(x) > e_{1L}(x)/e_{1H}(x)$. Consider a policy that generally lowers the cost for all types by $(1 - \delta)$. Although the ratio $e_{\theta L}(x)/e_{\theta H}(x)$ remains the same, as a result of a reduction in the cost, effort levels increase. This increases U_θ uniformly. $R's$ optimal strategy is to hire types 1 and 2 whenever they have σ_H and $U_1(W) > U_2(W) > 0$ at $x_1 = x_2 = W$. Type 3 applicant with σ_H is hired under T only if at least one of the two does not have σ_H. Thus, type 3 expects that $x_3 = (1 - \Psi_1\Psi_2)W$ under T. If $U_3(x_3) < 0$, $E(S, e^h) = U_1(W) + U_2(W) > E(T, e^h) = U_1(W) + U_2(W) + U_3(x_3)$. Thus, lowering the cost for all by $(1 - \delta)$ will not eliminate prejudice. Alternatively, consider a policy to reduce the cost of type 3 applicant disproportionately so that $e_{3L}(x)/e_{3H}(x) \approx e_{2L}(x)/e_{2H}(x) > e_{1L}(x)/e_{1H}(x)$. Even if type 3 is discriminated in favor of type 2, it is likely that at $x_3 = (1 - \Psi_1\Psi_2)W$, $U_3(x_3) > 0$ since the cost for 3 is lower and thus her effort is greater than before. In this case, it can be that $E(T, e^h) > E(S, e^h)$, and thus, prejudice can be avoided.

This suggests that a race-conscious anti-discrimination policy can be more effective than a race-neutral anti-discrimination policy in fighting prejudice if minorities face much more difficulty in signaling.[11] As Aristotle argued, treating "like cases alike" is required for equality and "identical treatment is not equal treatment,[..] if individuals are not similarly situated."[12]

6 Discussion

6.1 Alternative Form of Prejudice

In the main analysis, we imposed a simple framework of selection under S. R commits to a selection of type θ with a probability of p_θ prior to observing the signal. As a result, even if $n^* > k$, the applicants do not expect competition if S is chosen.

However, a more elaborate scheme under S is possible. Consider choosing a subset Θ_P with a plan to competitively select the finalists after observing the signals of the preselected candidates in Θ_P, when $n^* > k$. That is, after selecting n^* applicants first, R observes the signals of n^* applicants. If more than k have σ_H, R selects type θ with a δ_θ probability. Therefore, the n^* pre-selected applicants may still face competition under S if many applicants have a qualifying signal σ_H.

This set-up is similar to the case of T, except that any competition will be only within a smaller set of n^* pre-selected applicants. Such a scheme is more realistic. When we incorporate such a partial competition mechanism under S, type $1H$ applicant's effort is now determined by

$$\begin{aligned} x_{1C}^* &\equiv [I_T\{\Psi_{-1C}\pi_1 + (1-\Psi_{-1C})\} + (1-I_T)p_1\{\Psi_{-1P}\delta_1 + (1-\Psi_{-1P})\}]W \\ &= mc_{1H}(e_{1H}^*), \end{aligned} \tag{25}$$

where Ψ_{-1P} is the probability that 1 faces competition under S. This is smaller than Ψ_{-1C} if effort levels are the same.

[11] As a result of affirmative action bans for higher education, in California, Michigan, and Texas, any race-based decision is now prohibited in the college admission process. There have been doubts whether color-blind affirmative action promotes efficiency. Antonovics and Backes (2013) report that, under a system that emphasized grades and test scores, the rate of under-representation would have been twice as great, if the admission standard of the University of California system had not been adjusted to incorporate race-relevant factors. This raises doubts as to whether competitive signals, like test scores, are more correlated with the quality of the applicants than with their race.

[12] "Minority Report," *The Economist* (Oct. 23, 2013)

In general, reducing the chance of competition under S enhances the signaling incentive of pre-selected applicants under S, which increases the value of signals U_θ under S. This increases R's incentive to choose S. That is, prejudiced reviews will become more attractive with this more sophisticated scheme of S. Hence, even if competitive signaling is not very inefficient, prejudice may be observed frequently. In this sense, what we find in our main analysis is the lower bound for the parameter ranges where prejudice arises.

6.2 Effectiveness of Competitive Signaling

Our analysis of what facilitates prejudice shows that the gist of the problem lies in whether competitive signaling system is efficient. It may not always be efficient. Excessive competition among the applicants may destroy their signaling incentive.

In the analysis, we have assumed that there is no cost of processing information, $s = 0$, for simplicity. In reality, there is a cost. When $s > 0$, reviewing only a subset of the applicants generates extra cost savings. Thus, the set of parameters for which prejudice is attractive to R will become larger. Then, in order to prevent prejudice, competitive signaling system needs to be much more efficient.

A competitive system is effective if it can offer sufficient opportunities to signal their ability to the disadvantaged applicants who would not have the same opportunities otherwise. To get the most of the competitive system, it requires the signals to be closely correlated with a hidden quality of the applicants, rather than irrelevant factors like economic hardship. If economic hardship significantly limits the possibility to signal for high-quality underprivileged applicants (j types), they will not be able to make sufficient signaling efforts even under competitive signaling, resulting in $U_j < 0$. In this case, prejudice will occur simply because the underprivileged applicants cannot afford signaling, and not because they are unqualified.

Hence, to reduce prejudice, we advocate policies that enhance the quality of information under competitive signaling and ensure the signals to be free from the influence of other irrelevant factors. The Head Start program or the No-Child-Left-Behind policy would be good examples. As these policies make the signaling environment more homogenous for all applicants, more underprivileged applicants can participate in signaling, and any observed signaling outcomes will correctly represent the hidden quality for all.

7 Conclusion

This paper models prejudice in a signaling game between an employer and n applicants who differ in an irrelevant but observable characteristic. Prejudice occurs when the employer chooses to ignore some of the informative signals of the applicants based on their irrelevant characteristic. In equilibrium, prejudice arises when the value of signals is not high enough since the applicants do not make enough signaling efforts as a result of insufficient incentives to signal. Competition for limited positions often becomes the reason to generate the insufficient incentives. When all applicants are identical and there is no discrimination, prejudice arises because an equal allocation of signaling incentives across many applicants cannot make any applicant's signal valuable. In this context, we find that equal opportunity policies to ensure competitive signaling may not be always efficient and can actually facilitate prejudice.

Limitied number of available positions may generate an expectation of discrimination even when there is no reason for discrimination. When discrimination of ex ante identical applicants is expected, prejudice arises because the discriminated groups' signals have no value as a result of their insufficient efforts. Similarly, when applicants face different costs to signal, prejudice arises when the signals from the disadvantage applicants are not worthwhile to review since they do not make enough efforts as signaling is too costly for them.

Overall, an effective policy to reduce prejudice is to make signaling more affordable for the applicants because such a policy improves the quality of information from signaling and makes the employer more interested in using full information. To improve the quality of information in competitive signaling, we discuss the implications of the policies such as the Head Start program and the No Child Left Behind policy.

The presumption of the inefficiency in prejudice in this paper is due to the fact that prejudice uses criteria that are unrelated to ability for selection, while it is ability that matters. Hence, even though we provide some explanations on how prejudice can improve the quality of information by concentrating the signaling incentives onto some favored groups, the main point of this paper is that such an improvement is possible only when the competitive system is severely inefficient.

Moreover, we expect that any gain from prejudice exists only in the short-term when the applicants have sufficient incentive for signaling. If there no longer is an incentive due to prejudice, prejudice will determine the course. Then, the capacity of the market will be permanently set at the current k, because no one else except the favored groups makes an effort. In that case, it is doubtful whether the favored group will continue to make the same level of effort, knowing that positions are guaranteed for any level of effort as other discriminated groups no longer make any effort. Hence, the fundamental inefficiency of prejudice is in a dynamic context. Such a long-term impact of prejudice is an important topic to investigate in future studies in order to understand the inefficiency of prejudice.

8 Appendix

1. Proof of Lemma 1

For any given $p_\theta \in [0,1)$, if $I_T = 1$, $x_\theta = W$, $e_{\theta A} = mc_{\theta A}^{-1}(W)$, and thus $U_\theta(e_{\theta A}) > 0$. Thus, $\sum_{\theta=1}^{n} U_\theta(e_{\theta A}(W)) > \sum_{\theta=1}^{n} p_\theta U_\theta(e_{\theta A}(W))$, confirming that $I_T = 1$. Under the expectation of $I_T = 0$, $x_\theta = p_\theta W$. Let x_0 be the level of incentive at which $U_\theta(e_{\theta A}(x_0)) = 0$. Then, for $p_\theta W \geq x_0 \Leftrightarrow p_\theta \geq x_0/W$, $E(S, p_\theta W) = \sum_{\theta=1}^{n} p_\theta U_\theta(e_{\theta A}(p_\theta W) \leq E(T, p_\theta W) = \sum_{\theta=1}^{n} p_\theta U_\theta(e_{\theta A}(p_\theta W))$. Thus, if $p_\theta \geq x_0/W$, applicants expect T, which increases their efforts to $e_{\theta A}(W)$, and the employer chooses T in equilibrium. If some θ expects $p_\theta < x_0/W$, however, lower incentives for efforts reduces efforts and the value of signal from θ, and thus, $E(T, p_\theta W) < E(S, p_\theta W)$. Hence, the self-fulfilling expectation of a low p_θ leads to an equilibrium at which $I_T = 0$.

2. Proof of Proposition 2

First, $nw_E > k \Leftrightarrow 1 - k/n > \sum_{m=k+1}^{n} \Psi_{-Cm}(1 - k/m)$, since $1 - k/n \geq (1 - k/m)$ and $1 > \sum_{m=k+1}^{n} \Psi_{-Cm}$. Then, from (14) and (16), it is always that $E(T, \pi_{\theta m}^o = k/m, e_E^*) > E(S, p_\theta^o = k/n^*, e_E^*)$ as long as $U_E > 0$. Thus, if $U_E > 0$, R reviews all applications T with a plan to select k applicants based on $\pi_{\theta m}^o = \pi_{\theta' m}^o = k/m$ if $k < m$. If $U_E < 0$, however, R never chooses T even if applicants expect T. Thus, the equilibrium necessarily involves prejudice. When $U_E > 0$, in equilibrium, R chooses full information T in expectation of e_E^* and the applicants make efforts e_E^* under the expectation of T and equal treatment $\pi_{\theta m}^o = k/m$.

Prejudice occurs in equilibrium when R turns to S, regardless of whether or not $U_E > 0$. The applicants must correctly expect the choice of S. From (17) and (19), if $n^* > k$, S is always better since $w_\theta(e_p) < 1$. On the other hand, if $n^* = k$,

$w_\theta(e_p) = 1$, $\pi^o_{\theta m} = 1$, and thus, $E(T, \pi^o_{\theta m} = k/m, e_P) = E(S, p^o_\theta = k/n^*, e_P)$. Since U_p is an increasing function of $x^p_{\theta C}$, and $x^p_{\theta C}$ increases as $p^o_\theta \to 1$, under equal treatment, the optimal n^* is k. This results in $x^p_{\theta C} = W > x^*_E$ and $e_P = (e_{\theta H}(W), e_{\theta L}(W), e_{\theta' H}(0) = e_{\theta' L}(0) = 0)$. At e_P, $E(S, p^o_\theta = k/n^*, e_P) = E(T, \pi^o_{\theta m} = k/m, e_P)$. Thus, S is chosen in equilibrium and the efforts are realized at e_P.

3. Proof of Proposition 3

Let $x_0 > 0$ be the level of signaling incentive at which $U_E(x_0) = 0$. That is, $U_E(x_0) = qmc_H^{-1}(x_0)(V - W) - (1 - q)mc_L^{-1}(x_0)W = 0$. Since $U_\theta(W) > 0$, this implies that $x_0 < W$. Similarly, we can define an effort ratio $e_0 \equiv e_{0L}/e_{0H} = mc_L^{-1}(x_0)/mc_H^{-1}(x_0) > 0$ that induces $U_E(x_0) = 0$. Then, $e_0 = q(V - W)/(1 - q)W$. The employer would find the signaling information from applicant θ valuable only if $e_\theta < e_0$, or $x_\theta > x_0$.

(1) A lower V lowers $q(V - W)/(1 - q)W$. Thus, as V decreases, it is more likely that $e_0 < e_{\theta L}(x^*_E)/e_{\theta H}(x^*_E)$ and $U_E(x^*_E) < 0$, other things being equal. Therefore, prejudice is more likely to occur for a low V.

(2) A lower q affects the signaling incentive $x^*_E = [1 - \sum_{m=k+1}^{n} \Psi_{-Cm}(1 - k/m)]W$ as well as $q(V - W)/(1 - q)W$. The probability of competition $\sum_{m=k+1}^{n} \Psi_{-Cm}$ depends on how likely other applicant θ' will have σ_H. The probability of observing σ_H from applicant θ', $\Psi_{\theta'} = qe_{\theta' H} + (1 - q)e_{\theta' L}$, is an increasing function of q. Thus, a decrease in q increases the incentive to signal x^*_E by decreasing the probability of competition and lowers $q(V - W)/(1 - q)W$. Let e_{L0} be the level at which $e_{L0} = q_L(V - W)/(1 - q_L)W$ for a lower $q_L < q$. Then, $e_{L0} = mc_L^{-1}(x_{L0})/mc_H^{-1}(x_{L0}) = q_L(V - W)/(1 - q_L)W < e_0$. Since $\frac{e_{\theta L}}{e_{\theta H}}$ decreases as x increases, it must be that $x_{L0} > x_0$ for $q_L < q$. That is, for a low q, a higher level of incentive x_{L0} is required to get $U_E(x_{L0}) = 0$. Therefore, other things being equal. prejudice is more likely to occur for a low q.

(3) If signaling is very costly, in order to attain e_0, a high incentive is required, i.e., $x_0 \approx W$. Then, distributing the incentives across all applicants can easily lower x^*_E to the level below x_0, resulting in $e_L(x^*_E)/e_H(x^*_E) > e_0$. Thus, R does not find it worthwhile to examine all information. In this case, reducing the cost is an effective remedy as it will lower x_0, which is more likely to induce $x^*_E > x_0$, and $U_E > 0$.

(4) $U_E(x^*_E)$ is increasing in $x^*_E = w_E W$. From (13), $w_E = [1 - \sum_{m=k+1}^{n} \Psi_{-Cm}(1 - k/m)]$. To show the effect of k, suppose that w_E is differentiable with respect to k. Then, $dw_E/dk = \sum_{m=k+1}^{n} \Psi_{-Cm}(1/m) - \sum_{m=k+1}^{n}(1 - k/m)\left[d\Psi_{-Cm}/dk + (\Psi'_{-Cm})dw_E/dk\right]$. The first term $\sum_{m=k+1}^{n} \Psi_{-Cm}(1/m) > 0$ is the direct effect of increasing k. This is pos-

itive since an increase in k increases the signaling incentive by increasing the chance of selection. The second term $d\Psi_{-Cm}/dk < 0$ shows that increasing k directly lowers the probability of facing competition. The third term $(d\Psi_{-Cm}/dw_E)dw_E/dk > 0$ is the indirect effect. As the incentive for signaling increases, a higher level of effort increases the chance of having a qualifying signals from competing applicants, and thus, there is a higher chance of facing competition. Then, $dw_E/dk = \frac{\sum_{m=k+1}^{n} \Psi_{-Cm}(1/m) - \sum_{m=k+1}^{n}(1-k/m)[d\Psi_{-Cm}/dk]}{1+\sum_{m=k+1}^{n}(1-k/m)(\Psi'_{-Cm})} > 0$. Thus, an increase in k increases w_E and x_E^*. The same argument can be applied to the case when w_E is not differentiable.

4. Proof of Proposition 4

(1) Suppose $\sum_{j=k+1}^{n} w_j U_j(e_{jH}^d, e_{jL}^d) < 0$. From (21), $E(T, \pi_{im}^o = 1, \pi_{jm}^o = I_{jmk}, e_{UE}^*) < E(S, p_{im}^o = 1, p_{jm}^o = 0, e_{UE}^*)$. Hence, the employer deviates to S, which makes only e_{UP} sustainable in equilibrium.

(2) If $\sum_{j=k+1}^{n} w_j U_j(e_{jH}^d, e_{jL}^d) > 0$, under the expectation of T, e_{UE}^* are chosen by the applicants and expecting e_{UE}^* the employer has incentive to choose T. On the other hand, if the applicants choose e_{UP} as a result of expecting S, the employer may choose S in equilibrium, and thus the equilibrium is self-fulfilling. Since $\sum_{j=k+1}^{n} w_j U_j(e_{jH}^d, e_{jL}^d) > 0$, ex ante, $E(T, \pi_{im}^o = 1, \pi_{jm}^o = I_{jmk}, e_{UE}^*) > E(S, p_{im}^o = 1, p_{jm}^o = 0, e_P)$.

5. Proof of Corollary 1

1. When $U_E > 0$. Suppose that $\sum_{j=k+1}^{n} w_j U_j(e_{jH}^d, e_{jL}^d) < 0$. Under the expectation of discrimination, prejudice is the unique equilibrium. Nondiscriminatory policy establishes the expectation of equal $\pi_{\theta m}^o$. Since $U_E > 0$, the equilibrium may induce full information. Hence, non-discriminatory policy reduces prejudice.

If $\sum_{j=k+1}^{n} w_j U_j(e_{jH}^d, e_{jL}^d) > 0$, on the other hand, full information can arises in equilibrium. Since it is the expectation of full information that leads to the equilibrium, the same full equilibrium is achieved under the expectation of non-discrimination when $U_E > 0$. In the case of prejudiced outcome, nondiscriminatory policy does not make any difference.

2. When $U_E < 0$. Suppose that $\sum_{j=k+1}^{n} w_j U_j(e_{jH}^d, e_{jL}^d) < 0$. Under the expectation of discrimination, prejudice is the unique equilibrium. Non-discriminatory policy does not alter the outcome since $U_E < 0$. On the other hand, if $\sum_{j=k+1}^{n} w_j U_j(e_{jH}^d, e_{jL}^d) > 0$, full information can arises in equilibrium. Since prejudice is the unique outcome under equal $\pi_{\theta m}^o$ given that $U_E < 0$, nondiscriminatory policy promotes prejudice.

References

[1] Antonovics, K. and B. Backes (2014), "The Effect of Banning Affirmative Action on College Admissions Policies and Student Quality," *Journal of Human Resources*, 49(2), pp.295-322.

[2] Arrow, K.J. (1973), *The Theory of Discrimination*, Discrimination in Labor Markets, edited by Ashenfelter, O. and A. Rees, Princeton University Press, pp.3–33.

[3] Becker, G.S. (1957), *The Economics of Discrimination*, Chicago: The University of Chicago Press.

[4] Benabou, R. (1996), "Equity and Efficiency in Human Capital Investment: The Local Connection," *Review of Economic Studies*, 63, pp.237-264.

[5] Bowles, S., G. Loury, and R. Sethi (2014), "Group Inequality," *Journal of European Economic Association*, 12(1), pp.129–152.

[6] Blume, L. (2005), "Learning and statistical discrimination," *American Economic Review,* 95 (2), pp.118-121.

[7] Chaudhuri, S. and R. Sethi (2008), "Statistical Discrimination with Peer Effects: Can Integration Eliminate Negative Stereotypes?" *Review of Economic Studies*, 75, pp.579-596.

[8] Coate, S. and G.C. Loury (1993), "Will affirmative-action policies eliminate negative stereo-types?" *American Economic Review,* 83, pp.1220-1240.

[9] Durlauf, S.N. (1996), "A Theory of Persistent Income Inequality," *Journal of Economic Growth*, 1, pp.75-93.

[10] Eeckhoudt, L. and C. Gollier (1995), "Demand for risky assets and the monotone probability ratio order", *Journal of Risk and Uncertainty*, 11, pp.113-122.

[11] Fang, H. and A. Moro (2011), "Theories of Statistical Discrimination and Affirmative Action: A Survey," *Handbook of Social Economics*, Vol. 1A, edited by Jess Benhabib, Matthew O. Jackson and Alberto Bisin, North-Holland, pp.133-200.

[12] Fryer, R. (2007), "Belief-Flipping in a Dynamic Model of Statistical Discrimination," *Journal of Public Economics*, 91(5), pp.1151-1166.

[13] Lang, K., M. Manove, and W.T. Dickens, (2005), "Racial Discrimination in labor markets with posted wage offers," *American Economic Review*, 95, pp.1327-1340.

[14] Mailath, G., Samuelson, L., and A. Shaked (2000), "Endogenous Inequality in Integrated Labor Markets with Two-sided Search," *American Economic Review*, 90(1), 46-72.

[15] Mialon, S. and S.H. Yoo (2014), "Employers' Preference for Discrimination," Emory University Economics Working Paper Series 14-02.

[16] Mookherjee, D., S. Napel, and D. Ray, (2010), "Aspirations, Segregation, and Occupational Choice," *Journal of the European Economic Associations*, 8, pp.139-168.

[17] Moro, A., and P. Norman (2003), "Affirmative Action in a Competitive Economy," *Journal of Public Economics, 87,* pp.567–594.

[18] Moro, A., and P. Norman (2004), "A General Equilibrium Model of Statistical Discrimination," *Journal of Economic Theory*, 114 (1), pp.1–30.

[19] Phelps, E. (1972), "The Statistical Theory of Racism and Sexism," *American Economic Review*, 62, pp.659-661.

[20] Spence, M. (1974), "Job Market Signaling," *Quarterly Journal of Economics*, 87, pp.355–374.

[21] The Society for Human Resource Management (SHRM) (2011), "Executive Brief: What Factors Influence Cost-Per-Hire?" www.shrm.org/research/benchmarks/.