I Z A

IZA DP No. 8445

The Timing of Preference and Prejudice in Sequential Hiring Games

Glen R. Waddell
Logan M. Lee

August 2014

The Timing of Preference and Prejudice in Sequential Hiring Games

Glen R. Waddell
*University of Oregon
and IZA*

Logan M. Lee
University of Oregon

ABSTRACT

The Timing of Preference and Prejudice in Sequential Hiring Games[*]

We model a hiring process in which the candidate is evaluated sequentially by two agents of the firm who each observe an independent signal of the candidate's productivity. We introduce the potential for taste-based discrimination and characterize how one agent's private valuation of the candidate influences the other agent's hiring practices. This influence is often in an offsetting direction and is partially corrective. Yet, this offsetting response can also be large enough that even a high-productivity candidate who is privately favoured by one agent, as may be the case in efforts to increase gender or racial diversity, is less likely to be hired even when the other agent has no preference over private, non-productive attributes.

Corresponding author:

Glen R. Waddell
Department of Economics
University of Oregon
Eugene, OR 97403-1285
USA
E-mail: waddell@uoregon.edu

1 Introduction

Gender and race gaps in wages and employment persist in U.S. and global labor markets. While experimental evidence supports taste-based racial discrimination as a direct contributor (Bertrand and Mullainathan, 2004; Carlsson and Rooth, 2011; Castillo, Petie, Torero and Vesterlund, 2013), incomplete information can also give rise to statistical discrimination (Altonji and Pierret, 2001; Farber and Gibbons, 1996; Aigner and Cain, 1977). We consider a mechanism in the intersection of these areas.

In particular, we consider agents' incentives within a sequential evaluation of a job candidate when the agents of the firm potentially realize private benefits or costs associated with an observable but non-productive attribute of the candidate. Holding the sequential nature constant—an initial screening followed by further consideration if the initial screening goes well—we vary *where* in the sequence and to what degree the candidate's non-productive attribute is valued. Among our results, we will show that pro-diversity hiring policies—"top-down" directives, in particular—may be limited in their ability to narrow gaps in outcomes across race or gender, or even contribute to increasing wage and employment gaps. Thus, in this setting, even preference *for* some non-productive attribute in a job candidate can be to the candidate's detriment.

The setting we consider is rich enough to capture the relevant tradeoffs yet sufficiently straightforward that we can speak effectively to policy. We abstract away from the role of committees, for example, and consider only individual agents, two in number, and acting in sequence on behalf of the firm or institution. We assume that the candidate is considered by the second agent (have in mind the firm's owner, for example, although one could imagine university administrator also fitting well) only when the first agent (a division manager, for example, or a department chair) has determined that the candidate is worthy of forwarding in the search. In that way, the process we model captures the typical "up or out" nature of

1

job searches.[1]

Becker (1957) first introduced an economic model of discrimination in which employers had a taste for discrimination, insofar as there was a disamenity to employing minority workers who would have to compensate employers by being more productive at a given wage or being willing to accept a lower wage for identical productivity. Elements of this intuition will remain in our model, although the implications will now depend on where in the sequence such a disamenity is introduced—whether it is introduced "early" or "late." Elements of the longer literature will also be evident in what follows as we reconsider the role of private valuations amid uncertainty around worker productivity(Arrow, 1971; Phelps, 1972; McCall, 1972; Arrow, 1973; Spence, 1973).[2]

In terms of actionable policy, we will speak directly to the implications of directed searches—where private values are arguably a stronger motivating factor at the top of the firm's hierarchy. We will refer to these preferences as "top-down," and demonstrate that in such environments, early decision makers will often take positions that offset the anticipated preferences of later decision makers. In the limit, when the late-arriving preference *for* the personal attribute is large, this "offsetting" effect is sufficient to leave even the

---

[1]Green and Laffont (1987) model a two-person decision problem but assume away a hierarchy of agents. Similarly, Luo (2002) considers collective decision making in a two-person model where agents collaboratively to make decisions.

[2]In other related work, Eriksson and Lagerström (2012) use a resume study in Norway to show candidates who have non-Nordic names, are unemployed, or older receive significantly fewer firm contacts. Kuhn and Shen (2013) find that job postings in China that explicitly seek a certain gender, while suggestive that firms have preferences for particular job-gender matches, only play a significant role in hiring decisions for positions that require relatively little skill. Jacquemet and Yannelis (2012) discuss whether observed bias is due to discrimination against a particular group or favouritism for another group. Other explanations for gender and race gaps include firms benefitting from increased productivity when workforces are homogenous (Breit and Horowitz, 1995), and in-group-favouritism effects (Lewis and Sherman, 2003). Pinkston (2005) introduces the role for differentials in signal variance (e.g., black men have noisier signals of ability than white men) into a model of statistical discrimination. Ewens, Tomlin and Wang (2012) consider separating statistical discrimination from taste-based discrimination and find support for statistical discrimination in rental markets. For a review of the evolution of empirical work on discrimination, see Guryan and Charles (2013).

high-productivity candidates from the privately preferred group worse off; facing a *lower* probability of employment. For example, where leadership values female candidates, highly productive female applicants are harmed by early decision makers protecting their interest against the anticipation of favourable treatment in subsequent rounds. In no way is this due to a disutility associated with hiring a candidate with a particular attribute (e.g., we do not need the first agent to dislike female candidates to find that female candidates can be made worse off when favoured by the second agent) but is solely due to agents having incomplete information of candidates' true abilities and the requisite tradeoffs being made at the margin when the early mover anticipates a candidate-favouring bias being introduced by subsequent decision makers. Thus, one might fear that policies designed to encourage the hiring of workers who increase workforce diversity can promote even the opposite outcome if agents of the firm (particularly those acting early in hiring decisions) do not share equally in those interests.

This tension between the first and second decision makers is fundamental. As such, we consider comparative statics around these margins, varying the private values introduced by the first and second agents as we consider the implications on employment and workforce productivity. As private values influence the relative probabilities with which candidates of different abilities are hired, we will also discuss the distributional consequences for subsequent promotion games.

In Section 2 we introduce the model we have in mind, solving the sequential consideration of agents backwards. Throughout, we consider private values of either sign although cases in which candidates are "favoured" somewhere in the hiring process may be the more relevant to policy, especially where we demonstrate that this can be to their detriment. We do this in two settings.

In Section 3 we consider a setting in which the second agent in the sequence is somewhat "naive" in forming his expectations of the first agent's action—not expecting that the first agent may respond to the second agent's private incentives. For example, university lead-

ership may reveal that they favour female or minority candidates at the margin and fully expect that departments will not work to oppose these interests. Yet, as long as there is the potential for departments to value those non-productive attributes differently, interests can be in conflict. In particular, we discuss the model's implications in light of the asymmetries in how early and late decision makers can influence outcomes when agents are moving in sequence, including subsequent promotion games and the role of incentive pay.

In Section 4 we consider a setting in which Agent 2 is "savvy" regarding Agent 1's incentives, and fully anticipates this in his own optimization routine. While we tend to think that those in leadership positions (university deans, for example) may fall short of fully anticipating how others (department committees) might respond to "top-down" directives, we offer additional intuition by considering outcomes spanning these settings. In this setting, we consider whether the second decision maker can incentivize the first toward better aligning their private valuations of the non-productive attribute.

In Section 5 we offer concluding remarks.

2   Theory

2.1   The setup

We are intent on considering the implications of agents having private values associated with some non-productive attribute of a job candidate as they undertake the hiring responsibilities for the firm. In so doing, we consider a two-stage hiring game in order to speak to the implications of these private values being introduced to the hiring process at different stages. By assumption, Agent 1 considers the candidate first and either rejects the candidate or forwards the candidate to Agent 2 for further consideration. If forwarded, Agent 2 can then reject or hire the candidate. Within such a hierarchy, we then consider private valuations: "bottom-up" preferences (e.g., grass roots efforts to increase racial diversity among co-workers), or "top-down" preferences (e.g., a university administrator's preference

4

to increase the presence of female faculty in STEM fields), or combinations thereof.[3]

As a candidate's productivity is not verifiable, both agents only know that with probability $\alpha \in (0, 1)$ a given candidate is highly productive and would therefore be a "good" for the firm. We quantify the upside to hiring such a candidate as an increase in the firm's value from V_0 to V_g. With probability $(1 - \alpha)$ the candidate's productivity is such that hiring the candidate would be "bad" for the firm and would decrease the firm's value from V_0 to V_b. In such a case, the firm is always best served by rejecting the candidate, in which case the firm's value would remain at the status-quo level, V_0. Without loss of generality, we assume that $V_0 = 0$.

It is uninteresting to consider compensation schemes that do not tie remuneration to agents' actions. That said, these weights are determined outside the model and we simply parameterize these relationships in Agent 1 receiving $\tau_1 \in (0, 1)$ of the value to the firm and Agent 2 receiving $\tau_2 \in (0, 1)$, such that $\tau_1 + \tau_2 \leq 1$. As agents are moving in strict sequence, consistent with a hierarchy, it may be reasonable to further anticipate that $\tau_1 \leq \tau_2$.[4]

We introduce the potential for discrimination and favouritism by allowing for some non-productive but verifiable attribute of the candidate to be privately valued by either or both agents. Given the sequence of actions, we notate any private benefits accruing to Agent 1 from hiring the candidate as B_1, and any private benefits accruing to Agent 2 as B_2. To maintain interest and relevance, we will limit agents' private values to those that yield interior solutions.[5] That is, we will limit private values to those that do not have the agents' first-order conditions collapse to "always reject" or "always accept." The model can be solved backwards.

---

[3]STEM: Science, Technology, Engineering, and Mathematics.

[4]For some context regarding the use of incentive pay broadly, see Murphy (2013).

[5]Assuming that $\tau_1 V_b \leq B_1 \leq \tau_1 V_g$, and $\tau_2 V_b \leq B_2 \leq \tau_2 V_g$ effectively limits the set of values where an agent has these dominant strategies to just those where $B_i = \tau_i V_b$ or $B_i = \tau_i V_g$, respectively. More generally, the range of private values over which interesting interactions occur depends on the payoff levels to agents relative to these private values. That is, in the symmetric case, where $B_i < \tau_i V_b$, Agent i will adopt an "always-reject" strategy. Likewise, where $B_i > \tau_i V_g$, Agent i will adopt an "always-accept" strategy.

2.2 Agent 2's problem

When the candidate is forwarded to Agent 2 for final consideration, Agent 2 draws an independent signal of the candidate's productivity. The signal, s_2, is drawn from $N(\mu_b, \sigma_b)$ if the candidate is a "bad" type, and from $N(\mu_g, \sigma_g)$ if the candidate is a "good" type, where $\mu_b < \mu_g$. $F_b(\cdot)$ is the CDF of $N(\mu_b, \sigma_b)$ and $F_g(\cdot)$ is the CDF of $N(\mu_g, \sigma_g)$.[6] With such a setup, Agent 2's decision rule can then be summarized in the choice of a reservation signal, $\hat{s}_2$. If the realized signal, s_2, is higher than the chosen reservation signal, $\hat{s}_2$, the candidate is hired. If $s_2 < \hat{s}_2$, the candidate is rejected and no hire is made.

Formally, Agent 2's objective equation can be written as,

$$
\begin{aligned}
\underset{\hat{s}_2}{\text{Max}}\, V_2(\hat{s}_2) \;=\; & \alpha[F_g(\mathbb{E}_2[\hat{s}_1]) + (1 - F_g(\mathbb{E}_2[\hat{s}_1]))F_g(\hat{s}_2)]\tau_2 V_0 \\
& + \alpha(1 - F_g(\mathbb{E}_2[\hat{s}_1]))(1 - F_g(\hat{s}_2))(\tau_2 V_g + B_2) \\
& + (1 - \alpha)[F_b(\mathbb{E}_2[\hat{s}_1]) + (1 - F_b(\mathbb{E}_2[\hat{s}_1])F_b(\hat{s}_2)]\tau_2 V_0 \\
& + (1 - \alpha)(1 - F_b(\mathbb{E}_2[\hat{s}_1]))(1 - F_b(\hat{s}_2))(\tau_2 V_b + B_2).
\end{aligned}
\tag{1}
$$

As Agent 2 only considers the candidate upon her having successfully navigated Agent 1's evaluation, the probability Agent 2 puts on the candidate being highly productive is updated from the population parameter, α, to reflect Agent 1's evaluation (i.e., that s_1 must have been no smaller than $\hat{s}_1$). Each term in (1) therefore represents the probability weighted outcomes of the hiring game—the candidate is either a good candidate but not hired (Agent 2 realizes $\tau_2 V_0$), good and hired ($\tau_2 V_g + B_2$), bad and not hired ($\tau_2 V_0$), or bad and hired ($\tau_2 V_b + B_2$). While the true conditional probability depends on Agent 1's reservation signal, $\hat{s}_1$, what matters to characterizing Agent 2's choice is his belief about

---

[6]Lang and Manove (2011) suggest that employers find it more difficult to evaluate the productivity of black candidates than white candidates. This would imply that non-productive attributes may be correlated with signal noise. Our model can easily encompass this potential by allowing σ_b and σ_g to vary with the candidate's non-productive attribute.

what Agent 1's reservation signal was in the first stage, which we capture as $\mathbb{E}_2[\hat{s}_1]$.[7]

Given (1), Agent 2's choice of $\hat{s}_2$ solves the first-order condition,

$$\frac{\alpha(1 - F_g(\mathbb{E}_2[\hat{s}_1]))f_g(\hat{s}_2)}{(1 - \alpha)(1 - F_b(\mathbb{E}_2[\hat{s}_1]))f_b(\hat{s}_2)} = \frac{\tau_2 V_0 - (\tau_2 V_b + B_2)}{(\tau_2 V_g + B_2) - \tau_2 V_0}. \tag{2}$$

That is, in equilibrium Agent 2's optimal reservation signal, $\hat{s}_2^\star$, equates the ratio of probabilities of committing type-I and type-II errors (i.e., $\alpha(1 - F_g(\mathbb{E}_2[\hat{s}_1]))f_g(\hat{s}_2)$, and $(1 - \alpha)(1 - F_b(\mathbb{E}_2[\hat{s}_1]))f_b(\hat{s}_2)$, respectively) with the ratio of costs (i.e., $(\tau_2 V_g + B_2) - \tau_2 V_0$, and $\tau_2 V_0 - (\tau_2 V_b + B_2)$).

2.3 Agent 1's problem

In the first stage, Agent 1 draws an independent signal, s_1, of the candidate's productivity to be compared to a chosen reservation signal, $\hat{s}_1$. As above, the candidate's signal of productivity, s_1, is drawn from $N(\mu_b, \sigma_b)$ if the candidate is a "bad" type and from $N(\mu_g, \sigma_g)$ if the candidate is a "good" type. If $s_1 < \hat{s}_1$, the candidate's file is immediately abandoned and no hire is made—Agent 2 never sees the candidate and the resulting firm value is V_0. If $s_1 \geq \hat{s}_1$, the candidate is then subjected to consideration by Agent 2, as described in Equation (2).

Where $R_2(\mathbb{E}_2[\hat{s}_1])$ captures Agent 2's choice of $\hat{s}_2$ given his expectation of $\hat{s}_1$, Agent 1's objective equation can be written,

$$\begin{aligned}
\underset{\hat{s}_1}{\text{Max}}\, V_1(\hat{s}_1) &= \alpha[F_g(\hat{s}_1) + (1 - F_g(\hat{s}_1))F_g(R_2)]\tau_1 V_0 \\
&\quad + \alpha(1 - F_g(\hat{s}_1))(1 - F_g(R_2))(\tau_1 V_g + B_1) \\
&\quad + (1 - \alpha)[F_b(\hat{s}_1) + (1 - F_b(\hat{s}_1))F_b(R_2)]\tau_1 V_0 \\
&\quad + (1 - \alpha)(1 - F_b(\hat{s}_1))(1 - F_b(R_2))(\tau_1 V_b + B_1).
\end{aligned} \tag{3}$$

---

[7]Agent 2's expectation of the probability a good candidate cleared Agent 1's reservation is therefore $1 - F_g(\mathbb{E}_2[\hat{s}_1])$, while the expectation of the probability a bad candidate cleared Agent 1's reservation signal is $1 - F_b(\mathbb{E}_2[\hat{s}_1])$.

where we capture in B_1 any private value Agent 1 associates with the candidate's non-productive attribute. In general, Agent 1 chooses $\hat{s}_1$ subject to the first-order condition,

$$\frac{\alpha f_g(\hat{s}_1)(1 - F_g(R_2)) + \alpha(1 - F_g(\hat{s}_1))f_g(R_2)(\partial R_2/\partial \hat{s}_1)}{(1 - \alpha)f_b(\hat{s}_1)(1 - F_b(R_2)) + (1 - \alpha)(1 - F_b(\hat{s}_1))f_b(R_2)(\partial R_2/\partial \hat{s}_1)} = \frac{\tau_1 V_0 - (\tau_1 V_b + B_1)}{(\tau_1 V_g + B_1) - \tau_1 V_0}. \tag{4}$$

As above, Agent 1 chooses his optimal reservation signal, $\hat{s}_1^\star$, to equate the ratio of probabilities of committing type-I and type-II errors with the ratio of costs.[8]

3 When Agent 2 is naive

3.1 Agent behavior

In this section, we begin with the consideration of strictly "top-down" preferences (i.e., $B_2 \neq 0$ while $B_1 = 0$), which is consistent with Agent 1 being interested only in the productivity of the candidate while Agent 2 has private objectives associated with hiring, such as to increase the representation of certain races or gender of worker (i.e., $B_2 > 0$).

We model Agent 2's naiveté by setting his expectation of Agent 1's reservation signal, $\mathbb{E}_2[\hat{s}_1]$, equal to what Agent 1 would choose in the absence of any private values (i.e., as if $B_2 = 0$). In particular, this is akin to Agent 2 not anticipating that Agent 1 will consider B_2 when choosing $\hat{s}_1$. When $\mathbb{E}_2[\hat{s}_1] = \hat{s}_1^\star|_{B_2=0}$, Agent 2's first-order condition in (2) simplifies to

$$\frac{\alpha(1 - F_g(\hat{s}_1^\star|_{B_2=0}))f_g(\hat{s}_2)}{(1 - \alpha)(1 - F_b(\hat{s}_1^\star|_{B_2=0}))f_b(\hat{s}_2)} = \frac{\tau_2 V_0 - (\tau_2 V_b + B_2)}{(\tau_2 V_g + B_2) - \tau_2 V_0}, \tag{5}$$

and $\hat{s}_2^\star$ depends on the expectation of Agent 1's reservation signal, here set to $\hat{s}_1^\star|_{B_2=0}$, constant in B_2. That $\mathbb{E}_2[\hat{s}_1] = \hat{s}_1^\star|_{B_2=0}$ also implies that $\partial R_2(\mathbb{E}_2[\hat{s}_1])/\partial \hat{s}_1 = 0$. As Agent 1 finds neither private cost nor private benefit in the non-productive attribute of the candidate (i.e., $B_1 = 0$),

---

[8]This is easy to see in the symmetric case (i.e., $V_b = -V_g$, $V_0 = 0$, and $\alpha = 0.5$), as Agent 2's first-order condition collapses to $f_g(\hat{s}_2) = f_b(\hat{s}_2)$.

τ_1 drops from the agent's problem, and Agent 1's first-order condition in (4) simplifies to

$$\frac{\alpha f_g(\hat{s}_1)(1 - F_g(R_2(\hat{s}_1|_{B_2=0})))}{(1-\alpha)f_b(\hat{s}_1)(1 - F_b(R_2(\hat{s}_1|_{B_2=0})))} = \frac{V_0 - V_b}{V_g - V_0}. \tag{6}$$

which will vary with B_2 through its effect on $R_2(\cdot)$.

In Figure 1 we illustrating the tradeoffs in the sequential screening of candidates by plotting the optimally chosen $\hat{s}_1^\star$ and $\hat{s}_2^\star$ across a range of B_2 between $\tau_2 V_b$ (where the private cost to Agent 2 of hiring someone with this attribute completely offsets the benefit of hiring a "good" worker) and $\tau_2 V_g$ (where the private benefit to Agent 2 of hiring someone with this attribute completely offsets the cost of hiring a "bad" worker). For illustrative purposes, we impose *ex ante* symmetry.[9] Initially, we also abstract away from the role of incentive pay in agent behavior by setting $\tau_1 = \tau_2 = 0.5$. As changes in τ_1 and τ_2 determine the relative weights the private values play in agent decisions (i.e., where τ_i is large, Agent i's incentives are better aligned with the firm's) we will return to consider these margins below.

As illustrated in Figure 1, where B_2 decreases from zero and hiring the candidate imposes greater private costs on Agent 2, Agent 2 chooses a higher reservation signal, which is consistent with the agent's incentive to make it less likely that such a candidate successfully clears the required standard. While this exposes the firm to higher odds of making a type-I error (i.e., rejecting a good candidate) the perspective of Agent 2 is that the private costs of hiring an individual with the non-productive attribute are offset by the higher probability that the candidate is a good hire. That Agent 2 is motivated by this private value is clearly costly to the firm. Of course, any increase in B_2 from zero is also costly to the firm, as Agent 2 chooses a lower reservation signal in an attempt to increase the probability that the candidate is hired, where B_2 would be realized. This exposes the firm to higher odds of making a type-II error (i.e., hiring a bad candidate).

---

[9]Symmetry is defined as $V_b = -V_g$, $V_0 = 0$, and $\alpha = 0.5$. Collectively, the first-order condition for the choice of $\hat{s}_2$ is clear, as $f_g = f_b$ in equilibrium. In characterizing agent behavior, we adopt that $V_b = -4$, $V_g = 4$, $\mu_g = 1$, $\mu_b = -1$, and $\sigma_g = \sigma_b = 1$.

Figure 1 also reveals two interesting limiting cases in $B_2 = \tau_2 V_b$ and $B_2 = \tau_2 V_g$, where Agent 2's decision rule collapses on either "never hire" or "always hire." Again, this is in keeping with expectations. Where $B_2 = \tau_2 V_b$, the private cost associated with the non-productive attribute is sufficiently high that there is no possible outcome available (i.e., even $\tau_2 V_g$ is not sufficiently high) that would dominate the status quo of $\tau_2 V_0$ net of B_2. Likewise, where $B_2 = \tau_2 V_g$, the private benefit to the non-productive attribute is sufficiently high that there is no possible outcome available (i.e., even $\tau_2 V_b$ is not sufficiently low) that would dominate the potential that a "bad" hire is made and the firm realizes a value of $\tau_2 V_b$.

The shape of Agent 1's choice of $\hat{s}_1^\star$ across B_2 is where we first observe the behavior of consequence. First, as Agent 1 anticipates how $\hat{s}_2^\star$ varies with B_2, Agent 1's first-order condition in (6) implies that he adopts a *higher* reservation signal when B_2 is higher, requiring less uncertainty regarding the candidate's type before forwarding the candidate to Agent 2 where Agent 2 will be excessively favourable toward the candidate.

Proposition 1. *With top-down preferences, for any $|B_2| > 0$ Agent 1's choice of reservation signal acts as a weakly corrective force. That is, Agent 1's mitigating influence on firm value is non-negative as long as $|B_2| > 0$.*

Moreover, as B_2 approaches $\tau_2 V_g$ and Agent 2's decision rule collapses to the unproductive act of "always accepting" a candidate who provides the privately valued attribute, Agent 1's decision rule collapses to that which would be chosen by a single decision maker facing the same uncertainty (i.e., $\hat{s}_1^\star = 0$). In effect, while Agent 1's best response to Agent 2 favouring the candidate is corrective and valuable to the firm in expectation (i.e., it limits the potential losses that would otherwise result), Agent 2's private interest completely consumes the gains provided to the firm from having the second signal of the candidate's productivity.[10]

---

[10]Note that with symmetry assumed, a single decision maker would solve the first-order condition at $\hat{s} = 0$. In Figure 1, that $\hat{s}_1^\star < 0$ when $B_2 = 0$ is a reflection of the value to the firm of having a second agent. Agent 1 can adopt a lower reservation signal anticipating that Agent 2's independent draw and evaluation is pending. (While particularly evident at $B_2 = 0$, this is also driving the general result that $\hat{s}_1^* \leq 0$.)

However, this "corrective" ability of Agent 1 is not symmetric around $B_2 = 0$. As the private *costs* to Agent 2 increase and B_2 approaches $\tau_2 V_b$, Agent 2 never hires the candidate and Agent 1's choice is of no consequence to outcomes. The sequential nature of the hiring decision essentially limits the influence Agent 1 can have in offsetting $B_2 < 0$ and, in the limit, the firm suffers an unmitigated cost from Agent 2's bias. Again, this private cost results in the value to the firm created by the multiple signals of productivity being completely dissipated.

3.2 Implications for employment and firm value

In Panel A of Figure 2 we plot, across B_2, the employment rates associated with Agent 2 acting alone. While any observable attribute would work, we plot the relative treatments of gendered candidates, with the private value (B_2 in this case) capturing the private value associated with a female candidate. Clearly, without any offsetting influence of Agent 1, as B_2 increases from zero the probability a low-productivity female is hired clearly increases at a faster rate than does the probability a high-productivity female is hired. While optimal for Agent 2, this is destructive to firm value as this implies that the average productivity of female workers is falling. Likewise, as B_2 decreases from zero (and female hires are privately costly) the probability a low-productivity female is hired decreases at a slower rate than does the probability a high-productivity female is hired. This again decreases the value of the firm.

In Panel B of Figure 2, we plot employment rates across B_2, but with Agent 1 now actively participating in the hiring game. Relative to Agent 2 acting alone, the offsetting and corrective influence (from the firm's perspective) of Agent 1 is immediately evident. In fact, for both high- and low-productivity candidates, there is now significantly less separation in employment probabilities by gender, across all B_2 other than in the limiting case of $B_2 = \tau_2 V_b$. For high B_2 (i.e., those in the vicinity of $\tau_2 V_g$), high-productivity candidates can be strictly worse off than they would be without preference.

Proposition 2. *With top-down preferences, employment rates among low-ability candidates are strictly increasing in B_2. That is, low-ability candidates are always better off when they can offer employers a privately valued attribute. Alternatively, employment rates among high-productivity candidates are not monotonic in B_2. That is, there exists some $B_2 < 0$ for which the high-productivity candidate is strictly better off than he would be under a regime in which B_2 is large and positive. In a sequential hiring game, the early decision maker has enough influence on the candidate's prospect that the high-productivity candidate would prefer even mild discrimination in later rounds to having agents in later rounds offer strong favour.*

In Figure 3 we plot the expected value to the firm of a candidate with and without the influence of Agent 1 across B_2.[11] Not surprisingly, the firm values Agent 1's screen, which is evident in the higher firm values across B_2—Agent 1's screen better enables the hiring of "good" candidates. However, what is more interesting about the role of Agent 1 in the hiring game is the asymmetry introduced into the expected outcomes across B_2. In the absence of Agent 1, the expected costs to the firm associated with Agent 2 following his private interest are symmetric around $B_2 = 0$. However, when taking an active role in the hiring, Agent 1 is less effective at offsetting Agent 2's inclination to reject candidates (when $B_2 < 0$) than to hire candidates (when $B_2 > 0$), which introduces an asymmetry in firm value. Thus, given the ability of Agent 1 to unilaterally reject, the expected costs to the firm are higher with top-down discrimination (i.e., for $B_2 < 0$) than with top-down favouritism (i.e., for $B_2 > 0$).[12]

3.3 Extensions

Having modeled the direct outcomes of the hiring game, we consider two simple extensions.

---

[11]We normalize to one the expected value to the firm when Agent 2 is naive and there are no private values, $B_1 = B_2 = 0$.

[12]In the limit, as Agent 2's private values decrease, Agent 2 rejects all candidates with the private attribute, regardless of whether Agent 1 is present. In such cases, the expected value to the firm collapses to $V_0 = 0$.

3.3.1 Subsequent promotion games

As $B_2 \neq 0$ induces patterns of hiring that are specific to productivity-by-gender pools of candidates, in any subsequent period, average (within-firm) productivity levels will vary by gender. Even in the absence of private values playing a direct role in promotion decisions, promotion outcomes can be shown to depend on B_2.[13] For example, if $B_2 > 0$ at the hiring decision, the average female in the firm will be of lower productivity than the average male. Assuming that subsequent decision makers will perceive this difference in productivity, this disparity implies that females will suffer lower promotion probabilities within firms. While the implication of heterogeneous productivity in promotion games has been considered in the literature (Bjerk, 2008), we offer a source of heterogeneity—one driven, somewhat surprisingly, by favouritsm.

3.3.2 Performance pay

We next allow for $\tau_1 \leq \tau_2$ in order to consider the firm having taken steps to align the incentives differently across the internal hierarchy. In Figure 4 we show the optimal threshold levels for each agent across B_2 for a range of $\tau_2 \in [.5, 1)$, adjusting τ_1 accordingly, such that $\tau_1 = 1 - \tau_2$. For comparison with the baseline model, the solid lines indicate the $\hat{s}_1^\star$ and $\hat{s}_2^\star$ chosen when $\tau_1 = \tau_2 = 0.5$. Clearly, as τ_2 becomes increasingly large, any bias introduced in $\hat{s}_2^\star$ through $B_2 \neq 0$ (either discrimination or favouritism) is mitigated as Agent 2 cares more about the firm's value relative to his own private value as τ_2 increases. This is seen in the flattening of $\hat{s}_2^\star$ in B_2 in Figure 4. Importantly, the corresponding flattening of Agent 1's optimal $\hat{s}_1^\star$ in B_2 is entirely in response to B_2's influence on $\hat{s}_2^\star$. That is to say, because we have assumed $B_1 = 0$, any $\tau_1 > 0$ achieves unbiased decisions from Agent 1.[14]

---

[13]Of course, if the potential promotion of those with the privately valued attribute continue to be subject to the bias that occurred in the hiring process, outcomes will be affected. In fact, in such a setting, our "hiring" game can itself be recast as a promotion game of sorts.

[14]While we do not devote space to $\tau_1 \geq \tau_2$, these scenarios behave as expected. In the limit, where $\tau_2 = 0$, Agent 2 collapses to never hiring members of the non-preferred group for any $B_2 < 0$ and always hiring members of the preferred group for $B_2 > 0$.

In Figures 5 and 6 we plot the employment rates for good and bad workers respectively. As expected, increasing τ_2 works to offset biases arising from either $B_2 < 0$ or $B_2 > 0$, and allows for a larger range of these private values over which $\hat{s}_2$ does not collapse to either "always hire" or "never hire" rules.

3.4 The role of Agent 1's private value

As one last consideration before generalizing to both agents valuing the candidate's non-productive attribute, note the asymmetry in Agent 1's ability to mitigate Agent 2's biases—when Agent 1 foresees Agent 2's bias, Agent 1 plays a corrective role. Yet, a naive Agent 2 plays no such role when Agent 1 exercises favouritism or discrimination. In this way, our model reverts to the Becker (1957) intuition—Agent 2 simply facilitates a second signal of productivity and acts unbiasedly.

Proposition 3. *For a given private value, $W < 0$, the candidate would prefer to be subjected to a regime where $\{B_1, B_2\} = \{0, W\}$ than to a regime where $\{B_1, B_2\} = \{W, 0\}$. That is, if the candidate is to be discriminated against somewhere, she prefers discrimination to fall late in the sequence. Alternatively, for a given private value, $W > 0$, the candidate would prefer to be subjected to a regime where $\{B_1, B_2\} = \{W, 0\}$ than to a regime where $\{B_1, B_2\} = \{0, W\}$. That is, favouritism is more beneficial if experienced early in the sequence.*

In Figure 7, we allow for $B_1 \neq 0$ and $B_2 \neq 0$, capturing that both agents may value the candidate's non-productive attribute. As before, we plot Agent 2's choice of $\hat{s}_2$, but now with a menu of $\hat{s}_1$ corresponding to values of $B_1 \in (\tau_1 V_b, \tau_1 V_g)$. (As Agent 2 is naive, note that B_1 has no influence on $\hat{s}_2$.) Within the series of plots, Agent 1's decision rule in the strictly "top-down" case (i.e., that corresponding to $B_1 = 0$) can be seen in the solid line.

Figure 7 illustrates two results. First, as we have assumed that Agent 2 is not best responding to $\hat{s}_1$ at the margin, we document the expected pattern of behavior, that, for any $B_2 \in (\tau_2 V_b, \tau_2 V_g)$, $\hat{s}_1$ is strictly decreasing in B_1. As Agent 1's private value increases, holding constant Agent 2's private value, Agent 1 is less likely to reject those candidates who

14

have the attribute. The less-obvious takeaway from Figure 7, and one we wish to stress, we state as a proposition.

Proposition 4. *For all B_1, $\hat{s}_1^\star$ is strictly increasing in B_2. That is, Agent 1 raises the bar on candidates as Agent 2 is inclined to show less discrimination or more favour.*

In Figure 8 we plot the *ex post* rates of employment for "good" and "bad" female candidates, assuming that female is the private attribute around which the agents are potentially optimizing. As in Panel B of Figure 2, Figure 8 again captures that employment outcomes are sensitive to B_2, not only as a direct result of Agent 2's private value, but also indirectly through Agent 1's best response to $B_2 \neq 0$. Namely, employment rates among "good" female candidates eventually decline in B_2, reflecting Agent 1's ability to force the rejection of a particular candidate in response to a high B_2. As Agent 1 is less able to force the hiring of a candidate, employment rates among "bad" female candidates again monotonically increase in B_2. In panels A and B of Figure 8, then, we demonstrate that this strong tradeoff remains, across all B_1.

Proposition 5. *Both high- and low-productivity candidates prefer higher B_1 to lower B_1. That is, in a sequential hiring game when the late decision maker is naive, candidates weakly benefit from early preference as late decision makers provide no offsetting or corrective role.*

4   When Agent 2 is savvy

4.1   Agent behavior

In this section we relax the earlier assumption that Agent 2 is naive (i.e., does not recognize how Agent 1 best responds to $B_2 \neq 0$) and, instead, allow both agents to choose reservation signals while fully anticipating the effect that choice will have on the other agent's choice. While we are granting much more forethought and consideration to Agent 2 than may be evidenced in the field, this case fully bounds the possible scenarios relevant to policy and provides a richer understanding of the potential implications of private values in hiring games.

In Figure 9, we return to consider "top down" preferences (i.e., $B_1 = 0$) across a range of $B_2 \in (\tau_2 V_b, \tau_2 V_g)$, but allow Agent 2 to recognize that Agent 1 will adjust $\hat{s}_1$ in response to B_2. First, note that when $B_2 = 0$, both $\hat{s}_1^\star$ and $\hat{s}_2^\star$ are as they were in the case with a naive Agent 2. (This is expected, as one model nests the other when private values are absent.) Likewise, when $B_2 > 0$, the general patterns of behavior are similar to that in the naive-owner case. Yet, where $B_2 < 0$ and Agent 2 correctly anticipates $\hat{s}_1^\star$, both $\hat{s}_1^\star$ and $\hat{s}_2^\star$ behave differently in B_2 (than was the case with naiveté, in Figure 1). In particular, Agent 1's reservation signal is no longer monotonically increasing through $B_2 \in (\tau_2 V_b, \tau_2 V_g)$. To contrast, $\hat{s}_1^\star$ is now U-shaped, decreasing in B_2 for all $B_2 < 0$ in this range.

Proposition 6. *With top-down preferences, when Agent 2 is savvy in setting expectations of Agent 1's reservation signal, $\hat{s}_1^\star$ is monotonically decreasing in $B_2 \in (\tau_2 V_b, 0)$. (As when Agent 2 is naive, when Agent 2 is savvy $\hat{s}_1^\star$ is monotonically increasing in $B_2 \in (0, \tau_2 V_g)$.)*

The intuition for this result is again found in Agent 1's inability to fully offset prejudicial bias that arises late in the hiring sequence—while Agent 1 can secure a candidate's rejection, he cannot secure a candidate's hire. When Agent 2 anticipates a higher $\hat{s}_1$, he best responds by increasing $\hat{s}_2^\star$ all the more, which ultimately decreases employment rates among those presenting the privately costly attribute. By increasing $\hat{s}_1^\star$ as Agent 2 is more inclined to discriminate (i.e., as B_2 decreases from zero), Agent 1 is able to induce a lower $\hat{s}_2^\star$ than in the naive case. In essence, where Agent 2 is naive and Agent 1 then has no ability to influence Agent 2's decision, his decision rule was motivated solely by the potential to offset Agent 2's bias at the margin. Now, where Agent 2 is aware that $\hat{s}_1$ responds to B_2, Agent 1's choice of $\hat{s}_1$ influences $\hat{s}_2^\star$ at the margin. By raising his standard on candidates in the first period, Agent 1 lowers the marginal benefit to Agent 2 increasing $\hat{s}_2^\star$ in the second period, thereby allowing the firm to better exploit the gains available through the second signal of productivity. We learn by this that prejudicial bias introduced late in a sequential hiring game can motivate what looks like a prejudicial bias in earlier rounds; a preemptive bias-correction, of a sort. In this way taste-based discrimination introduced late in a sequence

can yield a sort of statistical discrimination earlier in the sequence. However, in this setting, Agent 1 is not responding to a perceived difference in the average productivity of female candidates—as would be the case in standard models of statistical discrimination—but in recognizing that subsequent decision makers will lean away from an unbiased assessment of productivity, treats female candidates differently as a corrective action.

4.2   Implications for employment and firm value

In Panel A of Figure 10 we again plot employment rates—the patterns are remarkably similar to those in the naive case. With Agent 2 now savvy, both high and low-productivity females are more likely to be hired for $B_2 > 0$ and less likely to be hired for $B_2 < 0$.

In Panel B of Figure 10 we plot the expected value to the firm of considering a candidate for the savvy and naive cases. While the firm's expected value is invariant to the assumption of naiveté when $B_2 = 0$, slight differences emerge at other values of B_2. In general, the firm suffers more from Agent 2's privately motivated decisions when Agent 2 is savvy; Agent 1 offers less of a corrective influence in such cases. The exception to this rule is for extreme discrimination (i.e., B_2 approaching V_b), where Agent 1's higher standard enables the firm to escape Agent 2's "always reject" regime.

4.3   The role of Agent 1's private value

In Panel A of Figure 11 for various values of B_1 we plot the rates at which high-productivity female candidates are hired across B_2. (Recall that we use the hiring of female candidates as a placeholder of sorts in the figures, which more-broadly apply to any observable non-productive attribute for which there may be private consideration.) The bold line captures that already represented in Figure 10. Around this line, however, we see the interesting asymmetry of employment rates. For example, where B_2 is large and negative and Agent 2 is increasingly inclined toward adopting a "never hire" position, Agent 1 has no ability to influence employment regardless of his inclination to do so (i.e., for any B_1). Thus, for all

B_1, employment rates converge to zero as B_2 decreases to $\tau_2 V_b$. As B_2 increases from $\tau_2 V_b$, employment rates fan out across B_1, with rates increasing faster in B_2 for higher values of B_1. This, again, reflects Agent 1's ability to "force" rejections (e.g., when B_1 is low), while being quite unable to force hires—even in the limit (as B_1 increases to $\tau_1 V_g$), employment is still very much dependent on Agent 2's private value (B_2).

In Panel B of Figure 11 we plot the expected value to the firm of a female candidate. That the expected value is highest when $B_1 = B_2 = 0$ again reflects that any privately motivated interest, in either agent, is costly to the firm. Moreover, it is interesting to note that for all B_2, firm value is maximized when $B_1 = 0$. That is, in the sequential hiring game, the full value to having multiple signals drawn and evaluated is only exploited when the first agent is free from bias. Any departure from this not only costs the firm directly (through Agent 1 choosing a standard that depends on B_1), but indirectly costs the firm through Agent 1's influence on Agent 2's decision (even when $B_2 = 0$).

The timing of preference—whether introduced with Agent 1 or Agent 2—yields striking differences in agents' optimal thresholds. In Figure 12, we impose bottom-up preferences (i.e, $B_2 = 0$) and plot agents' optimal thresholds (Panel A) and associated employment probabilities (Panel B) across B_1. Most notable, with bottom-up preferences, Agent 2's optimal threshold is monotonically increasing in B_1. This is different from the patterns evident with "top-down" preferences (recall Figure 9), where the agent without private preference appears to "buy" more-lenient treatment from the agent who finds the candidate's non-productive attribute privately costly.

The importance of the timing of bias is also seen in Panel B of Figure 12, where we plot associated employment probabilities by productivity. With discrimination, the timing of the introduction of private values is of little consequence to employment; either agent can unilaterally dismiss candidates. As no single agent can unilaterally hire a candidate, preference for a candidate's non-productive attribute yields different patterns of behaviour. With bottom-up preferences, both good and bad female candidates are more likely to be

18

hired than male candidates, for all B_1. This contrasts with top-down preferences (see Panel A of Figure 10) where strong preference on the part of Agent 2 ultimately leaves good female candidates less likely to be hired.

4.4 Extensions

Given the similarities in employment outcomes across the naive and savvy cases, we forgo additional discussion of subsequent hiring and promotion games and the implications of performance pay. Yet, unique to the environment in which Agent 2 fully anticipates Agent 1's best response to $B_2 \neq 0$ (which, loosely speaking, is to take corrective action and mitigate Agent 2's ability to act on his private valuations), it is interesting to consider the potential for a transfer, from Agent 2 to Agent 1, to incentivize Agent 1's cooperation.[15]

4.4.1 Can Agent 2 purchase Agent 1's cooperation?

Here we consider one important extension to the model—a potential transfer, from the firm (although at Agent 2's choosing) to Agent 1, attached to the hiring of a candidate presenting a particular non-productive attribute. We ask, then, whether there are private values $\{B_1, B_2\}$ for which Agent 2 will choose to reward Agent 1 for hiring such a candidate.[16]

Such practice appears in academic markets, for example, where payments would typically be made—by college-level administrators to departments—conditional on hiring a candidate

---

[15]We do not discuss such a payment in the "naive" case, as Agent 2 recognizing the potential for Agent 1's corrective action seems a prerequisite to explaining the use and effect of such payments.

[16]US labor law forbids deductions from employee pay without serious violations of workplace rules. As such, we do not consider whether there are values for which Agent 2 would tax Agent 1 for hiring a candidate with a particular non-productive attribute. Regardless, the sequential nature of the hiring process limits Agent 2's ability to require payment from Agent 1 for hiring a candidate, as Agent 1 can always avoid such penalties by raising the required standard for hire. Agent 1 still solves the first-order condition for $\hat{s}_1$, of course, so while Agent 1 will not collapse to an "always reject" position immediately, in the limit, $\hat{s}_1^\star$ approaches "always reject."

who presents with a non-productive attribute such as a minority race of gender. We parameterize this payment with ρ, through which we allow Agent 2 to transfer $\rho > 0$ from the firm to Agent 1, conditional on hiring a candidate with a particular (non-productive but verifiable) attribute. Agent 2's objective can therefore be written as,

$$
\begin{aligned}
\operatorname*{Max}_{\hat{s}_2, \rho} V_2(\hat{s}_2) \;=\;& \alpha[F_g(\mathbb{E}_2[\hat{s}_1]) + (1 - F_g(\mathbb{E}_2[\hat{s}_1]))F_g(\hat{s}_2)]\tau_2 V_0 \\
&+\alpha(1 - F_g(\mathbb{E}_2[\hat{s}_1]))(1 - F_g(\hat{s}_2))(\tau_2(V_g - \rho) + B_2) \\
&+(1 - \alpha)[F_b(\mathbb{E}_2[\hat{s}_1]) + (1 - F_b(\mathbb{E}_2[\hat{s}_1])F_b(\hat{s}_2)]\tau_2 V_0 \\
&+(1 - \alpha)(1 - F_b(\mathbb{E}_2[\hat{s}_1]))(1 - F_b(\hat{s}_2))(\tau_2(V_b - \rho) + B_2),
\end{aligned}
\tag{7}
$$

where the payment reflects a reduction in firm value by the amount ρ upon hiring. Similarly, as Agent 1 receives ρ, his objective equation becomes,

$$
\begin{aligned}
\operatorname*{Max}_{\hat{s}_1} V_1(\hat{s}_1) \;=\;& \alpha[F_g(\hat{s}_1) + (1 - F_g(\hat{s}_1))F_g(R_2)]\tau_1 V_0 \\
&+\alpha(1 - F_g(\hat{s}_1))(1 - F_g(R_2))(\tau_1(V_g - \rho) + B_1 + \rho) \\
&+(1 - \alpha)[F_b(\hat{s}_1) + (1 - F_b(\hat{s}_1))F_b(R_2)]\tau_1 V_0 \\
&+(1 - \alpha)(1 - F_b(\hat{s}_1))(1 - F_b(R_2))(\tau_1(V_b - \rho) + B_1 + \rho).
\end{aligned}
\tag{8}
$$

In giving away part of the firm, the private cost to Agent 2 is merely his share of the direct reduction in firm value, $\tau_2\rho$. On this margin, $\rho > 0$ is more desirable to Agent 2 for small τ_2. Regardless, however, Agent 2 benefits by any such payment only to the extent that it moves Agent 1 in his preferred direction. As Agent 1's share of the cost of $\rho > 0$ (in terms of firm value) increases in τ_1, $\rho > 0$ only has the desired effect for small τ_1. Thus, only for small τ_1 and τ_2 can Agent 2 benefit from a non-zero transfer of $\rho > 0$ from the firm to Agent 1.

In many cases, however, Agent 2 finds $\rho^\star = 0$ to be optimal. This implies that the additional dollar that would be used to influence $\hat{s}_1^\star$ generates less than a dollar's worth of return in noise reduction and increased probability a candidate will be hired. Intuitively,

Agent 2 is most likely to choose a non-zero ρ in cases where B_2 is large. In the extreme case, where $B_2 \to \tau_2 V_g$, we have shown (in Figure 9) that Agent 1 acts as though he were the only screen ($\hat{s}_1^\star = 0$) while Agent 2 collapses to always hiring candidates that make it through the first screen. This leads to a significant increase in the number of low-productivity workers hired relative to the number of high-productivity workers hired and limits the payoffs to all parties. By choosing $\rho > 0 > B_2$, Agent 2 incentivizes Agent 1 to lower his chosen threshold, bringing $\hat{s}_1^\star$ more in line with $\hat{s}_2^\star$ and increasing the average productivity of workers hired.

5 Conclusion

In this paper we consider a firm's hiring process, with agents of the firm each drawing a signal of a candidate's productivity and choosing to either reject or forward the candidate based on that signal (or make the hire, if last in the sequence). Into this setting we introduce that agents may also have private costs or benefits associated with some non-productive personal attribute of the candidate. The implications are interesting and non-trivial.

We show that private values introduced in one stage of such a game are evident not only in the actions of the agent harboring those private motivations, but also among agents in other stages of the game, even if they neither benefit nor suffer privately with the outcome of the game. In particular, where preference *for* a personal attribute is introduced late in the sequence, earlier decision makers partially offset this preference by raising the standard they impose on a candidate with that personal attribute. From the firm's perspective, this moves toward first best and we therefore characterize this potential as partially corrective. In the typical "up-or-out" hiring environment, where earlier decision makers have much more sway in rejecting candidates than in hiring candidates, the potential response among earlier decision makers who anticipate subsequent favourable treatment still has the potential to subject candidates who are "preferred," on average, to lower odds of employment than they would have experienced had their private attribute not been valued or observable.

In closing, we note four interesting implications, each of which may motivate additional exploration. First, where a single decision maker discriminates on taste, the average productivity among the "preferred" group decreases. However, as early movers in a sequential decision can take positions offsetting top-down preferences, average *ex post* productivity falls off more slowly among those who are "preferred" *a priori*. For example, with top-down preferences, early decision makers who anticipate excessively favourable treatment of female candidates in subsequent evaluations best respond by increasing the standards they impose on female candidates, which implies that later decision makers will be considering female candidates who are, on average, of higher quality (i.e., able to have cleared the higher standards imposed in early rounds). Therefore, while fewer female candidates advance in the sequence, the average productivity of those who do advance for final consideration is higher. As such, this may leave later decision makers increasingly misinformed of underlying female productivity, thereby reinforcing or strengthening prior beliefs among those in leadership positions. Overall, the influence of late-arriving preference for female candidates will change the mix of low- and high-productivity female employees such that average productivity falls among female employees. This, we presume, also introduces a source of downward pressure on female wages and thereby contributes to the persistence of male-female wage gaps.

Second, in a setting where late decision makers are savvy enough to anticipate the best responses of early decision makers, early-moving agents, who themselves may be uninclined to discriminate, will *raise* the bar on candidates against whom leadership is inclined to discriminate. Average productivity of female candidates is therefore higher coming out of early stages, thereby moving subsequent priors away from "reject" and toward "accept." Interestingly, where standard models of taste-based discrimination yield heterogeneity in *ex poste* productivity by gender and standard models of statistical discrimination yield homogeneity in *ex poste* productivity, the sequence of decision making in our setting allows for taste-based discrimination to exist, yet, due to the corrective action of an earlier agent of the firm, *not* be evidenced in *ex poste* heterogeneity in productivity by gender.

Third, the model offers interesting implications in light of existing evidence that resumes with African-American-sounding names receive fewer call backs (Bertrand and Mullainathan, 2004). While such an empirical regularity is consistent with either a single decision maker statistically discriminating or a single decision maker exercising a kind of taste-based discrimination, it is also consistent with the actions of the first of multiple decision makers in a regime where subsequent decision makers are expected to show preference *for* African-American candidates. (We assume that call-back decisions are made by initial screeners and not by those who will ultimately make the hire.) Of course, policy prescriptions across these potential mechanisms will differ significantly.

Finally, note that the model we present implies that if preferences for the private attribute are of the top-down variety we describe, we should be concerned that even in regimes where women and racial minorities are valued by leadership, such candidates can be harmed by revealing their identities early if initial screeners merely value those attributes less than leadership. Candidates will also experience tension, insofar as they do benefit from eventually revealing their identities. (In the model, they would choose to identify strictly between Agent 1 and Agent 2.) "Blind" assessments should arguably be considered in this context, as outcomes are certainly not neutral with respect to the information provided to reviewers. For example, in regimes where preferences for female recruitment are not uniformly held across the firm's hierarchy, pro-minority leadership meets with more success by incorporating blind-recruitment tools in early assessments of job candidates.

References

Aigner, Dennis J and Glen G Cain, "Statistical Theories of Discrimination in Labor Markets," *Industrial and Labor Relations Review*, 1977, pp. 175–187.

Altonji, Joseph G and Charles R Pierret, "Employer Learning and Statistical Discrimination," *The Quarterly Journal of Economics*, 2001, *116* (1), 313–350.

Arrow, Kenneth, "Some Models of Racial Discrimination in the Labor Market," 1971.

_ , "The Theory of Discrimination," *Discrimination in Labor Markets*, 1973, *3* (10), 3–33.

Becker, Gary S, *The Economics of Discrimination*, University of Chicago press, 1957.

Bertrand, Marianne and Sendhil Mullainathan, "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," *American Economic Review*, September 2004, *94* (4), 991–1013.

Bjerk, David James, "Glass Ceilings or Sticky Floors? Statistical Discrimination in a Dynamic Model of Hiring and Promotion," *The Economic Journal*, 2008, *118* (530), 961–982.

Breit, William and John B. Horowitz, "Discrimination and Diversity: Market and Non-Market Settings," *Public Choice*, 1995, *84*, 63–75.

Carlsson, Magnus and Dan-Olof Rooth, "Revealing Taste Based Discrimination in Hiring: A Correspondence Testing Experiment with Geographic Variation," *IZA Discussion Paper*, 2011, (6153).

Castillo, Marco, Ragan Petie, Maximo Torero, and Lise Vesterlund, "Gender Differences in Bargaining Outcomes: A Field Experiment on Discrimination," *Journal of Public Economics*, 2013, *99*, pp. 35–48.

Eriksson, Stefan and Jonas Lagerström, "Detecting Discrimination in the Hiring Process: Evidence From an Internet-Based Search Channel," *Empirical Economics*, 2012, *43* (2), 537–563.

Ewens, Michael, Bryan Tomlin, and Liang Choon Wang, "Statistical Discrimination or Prejudice? A Large Sample Field Experiment," *Review of Economics and Statistics*, 2012, (0).

Farber, Henry S. and Robert Gibbons, "Learning and Wage Dynamics," *Quarterly Journal of Economics*, 1996, *111* (4), 1007–47.

Green, Jerry R. and Jean-Jacques Laffont, "Posterior Implementability in a Two-Person Decision Problem," *Econometrica*, 1987, *55* (1), pp. 69–94.

Guryan, Jonathan and Kerwin Kofi Charles, "Taste-Based or Statistical Discrimination: The Economics of Discrimination Returns to its Roots," *The Economic Journal*, 2013.

Jacquemet, Nicolas and Constantine Yannelis, "Indiscriminate Discrimination: A Correspondence Test for Ethnic Homophily in the Chicago Labor Market," *Labour Economics*, 2012.

Kuhn, Peter and Kailing Shen, "Gender Discrimination in Job Ads: Evidence from China," *The Quarterly Journal of Economics*, 2013, *128* (1), 287–336.

Lang, Kevin and Michael Manove, "Education and Labor Market Discrimination," *The American Economic Review*, 2011, *101* (4), 1467–1496.

Lewis, Amy C. and Steven J Sherman, "Hiring You Makes Me Look Bad: Social-Identity Based Reversals of the Ingroup Favoritism Effect," *Organizational Behavior and Human Decision Processes*, 2003, pp. 262–276.

Luo, Guo Ying, "Collective Decision-Making and Heterogeneity in Tastes," *Journal of Business & Economic Statistics*, 2002, *20* (2), pp. 213–226.

McCall, John J, "The Simple Mathematics of Information, Job Search, and Prejudice," *Racial Discrimination in Economic Life, Lexington Books*, 1972, pp. 205–224.

Murphy, Kevin J., "Executive Compensation: Where We Are, and How We Got There," in G. Constantinides, M. Harris, and R. Stulz, eds., *Handbook of the Economics of Finance*, Elsevier Science North Holland, Elsevier, 2013.

Phelps, Edmund S, "The Statistical Theory of Racism and Sexism," *The American Economic Review*, 1972, pp. 659–661.

Pinkston, Joshua C, "A Test of Screening Discrimination with Employer Learning," *Industrial & Labor Relations Review*, 2005, *59*, 267.

Spence, Michael, "Job Market Signaling," *The Quarterly Journal of Economics*, 1973, pp. 355–374.

6 Figures

Figure 1: Optimal reservation signals with top-down preferences ($B_1 = 0$) and a naive Agent 2
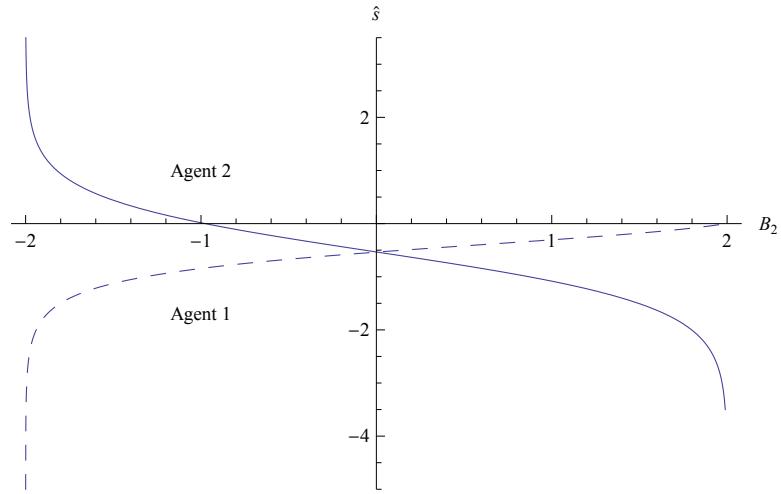
Figure 2: Employment probabilities with top-down preferences ($B_1 = 0$) and a naive Agent 2

Panel A: No screening provided by Agent 1

Employment Rate

High−productivity male

High−productivity female   Low−productivity female

Low−productivity male

B_2
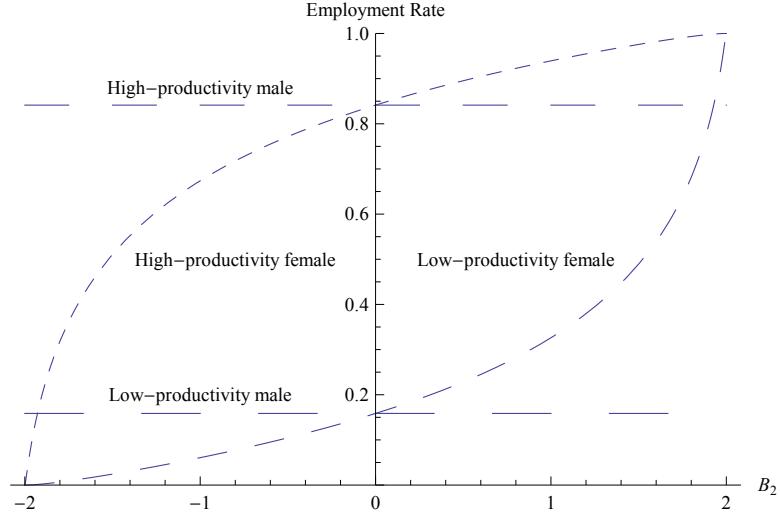
Panel B: Agent 1 screens candidates prior to Agent 2

Employment Rate

High−productivity male

High−productivity female

Low−productivity male   Low−productivity female

B_2
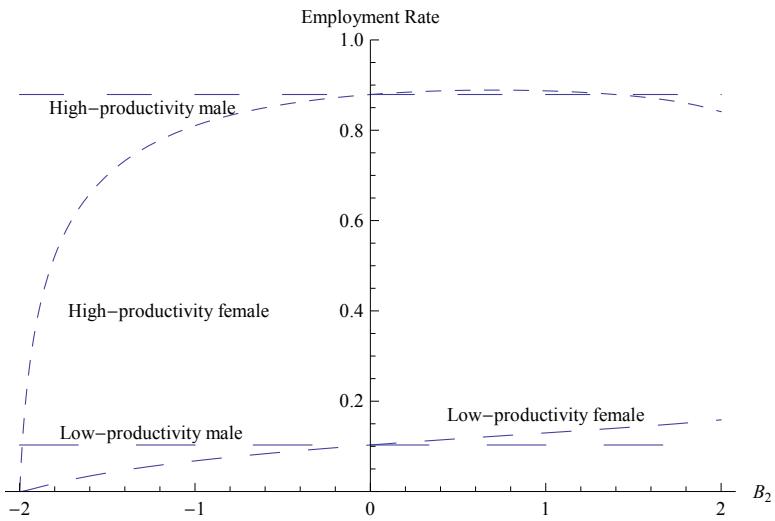
Figure 3: Firm value with top-down preferences ($B_1 = 0$) and a naive Agent 2
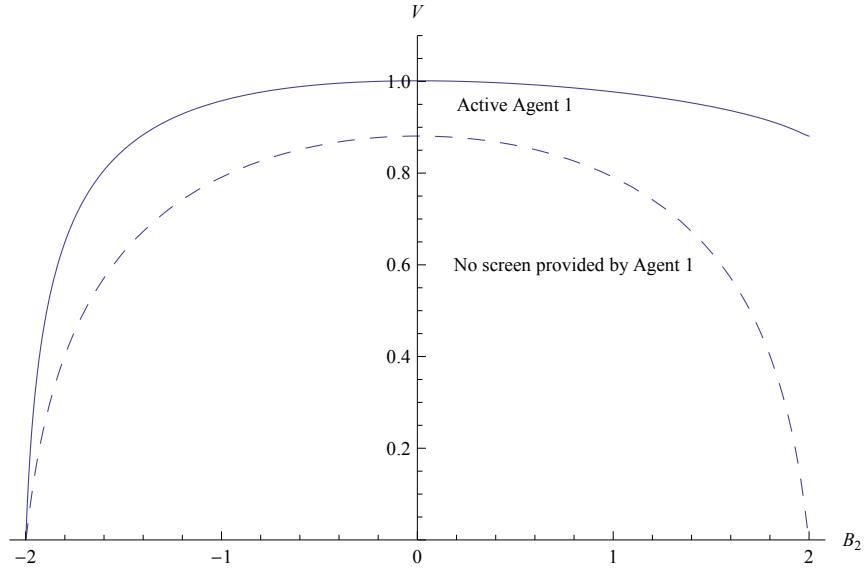
Figure 4: Reservation signals across τ_2 with top-down preferences ($B_1 = 0$) and a naive Agent 2
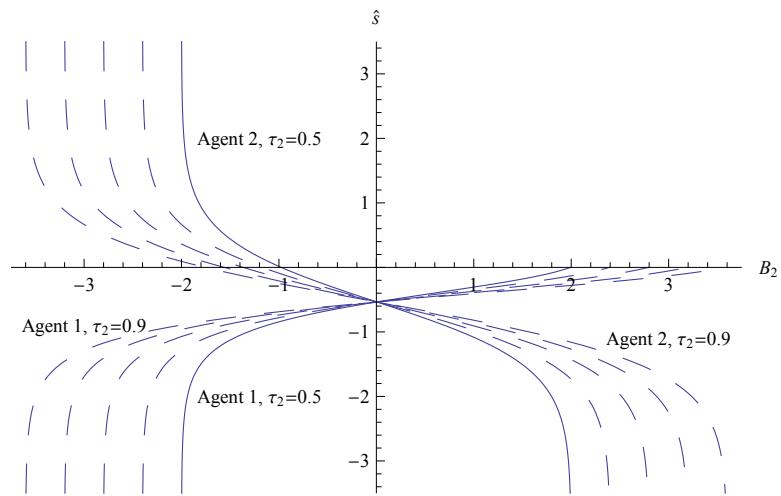
Figure 5: Employment rates for "good" workers across τ_2 with top-down preferences $(B_1 = 0)$ and a naive Agent 2
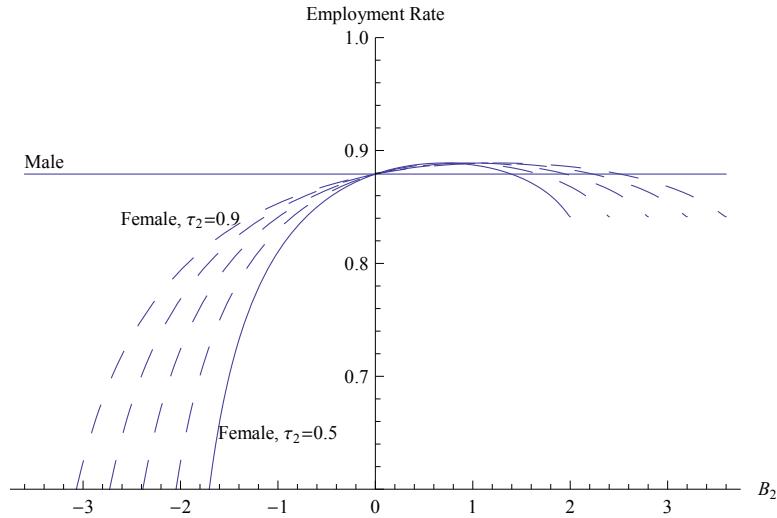
Figure 6: Employment rates for "bad" workers across τ_2 with top-down preferences $(B_1 = 0)$ and a naive Agent 2
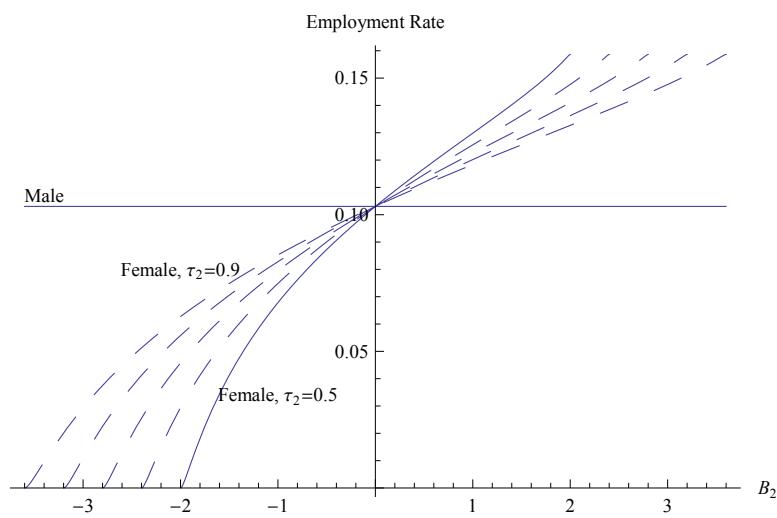
Figure 7: Reservation signals across B_1 when Agent 2 is naive
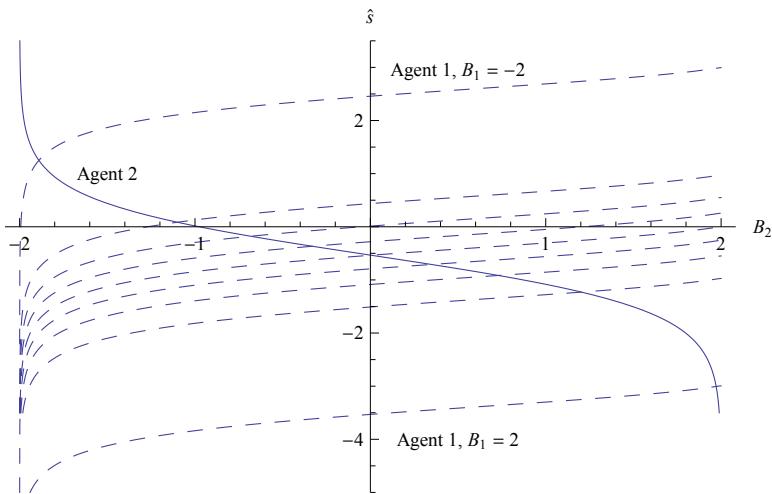
Figure 8: Rates of employment among female (i.e., the "preferred") candidates when Agent 2 is naive

Panel A: High-productivity female candidates
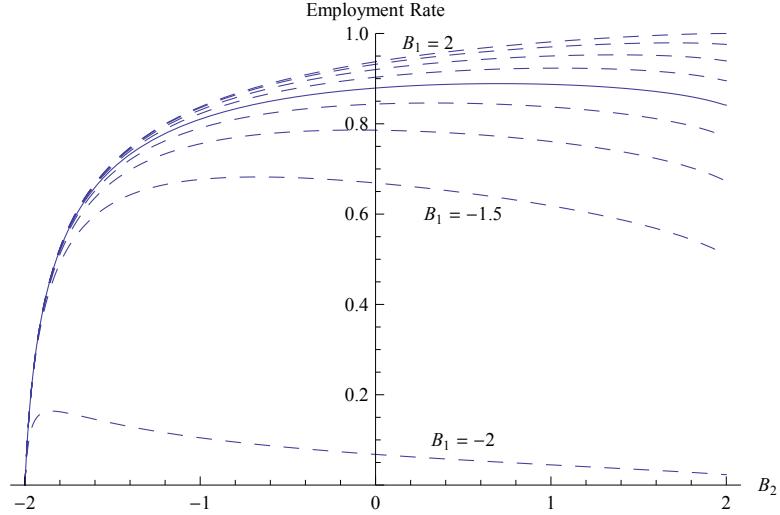
Panel B: Low-productivity female candidates
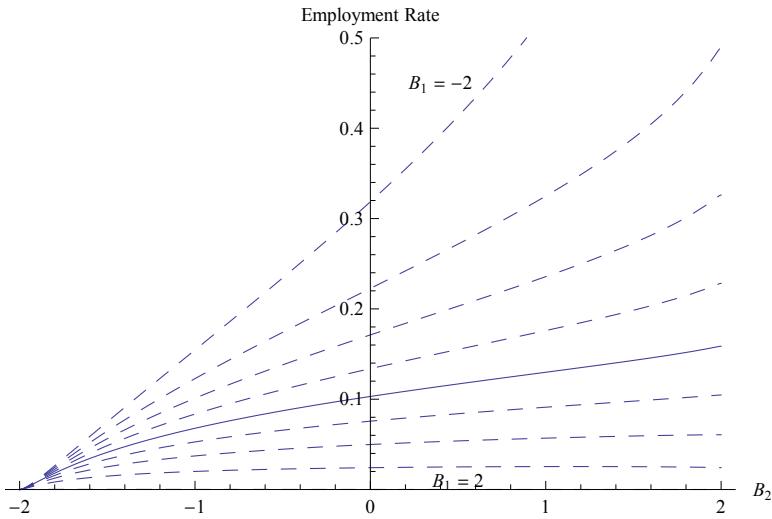
Figure 9: Optimal reservation signals with top-down preferences ($B_1 = 0$) and a savvy Agent 2
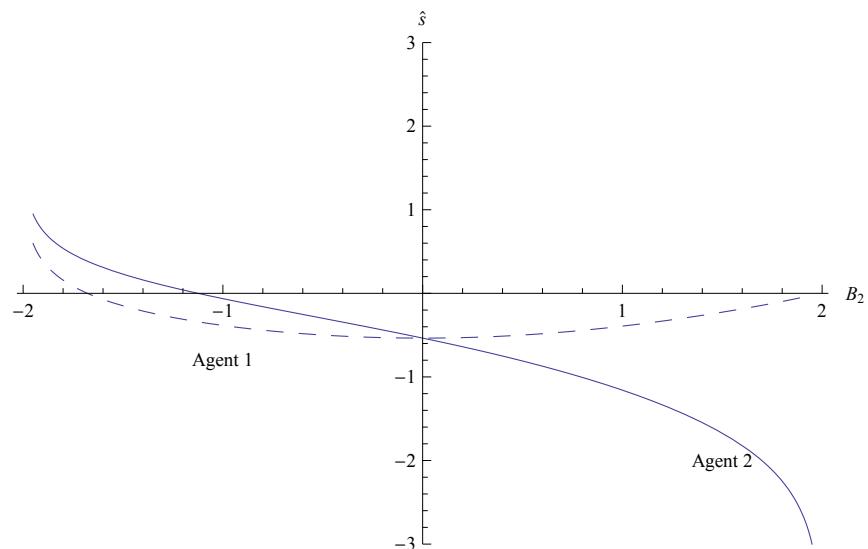
Figure 10: Employment probabilities and firm value with top-down preferences ($B_1 = 0$) and a savvy Agent 2

Panel A: Employment Probabilities
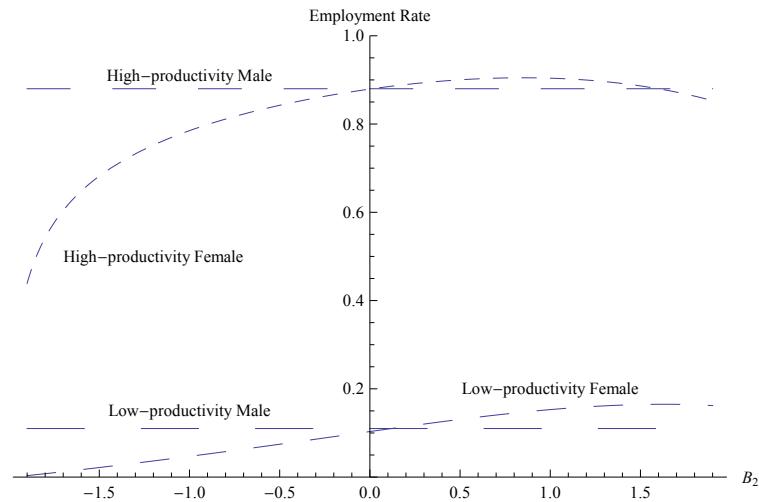
Panel B: Firm Value
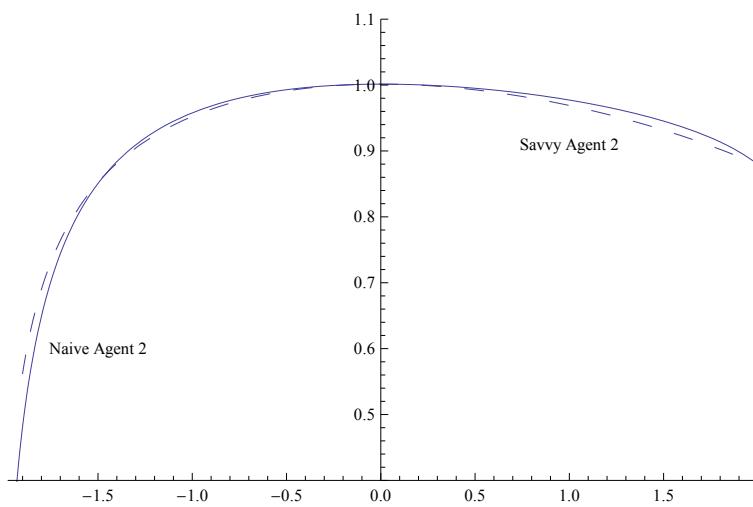
Figure 11: Employment probabilities and firm value when Agent 2 is savvy

Panel A: Employment probabilities among "good" female candidates
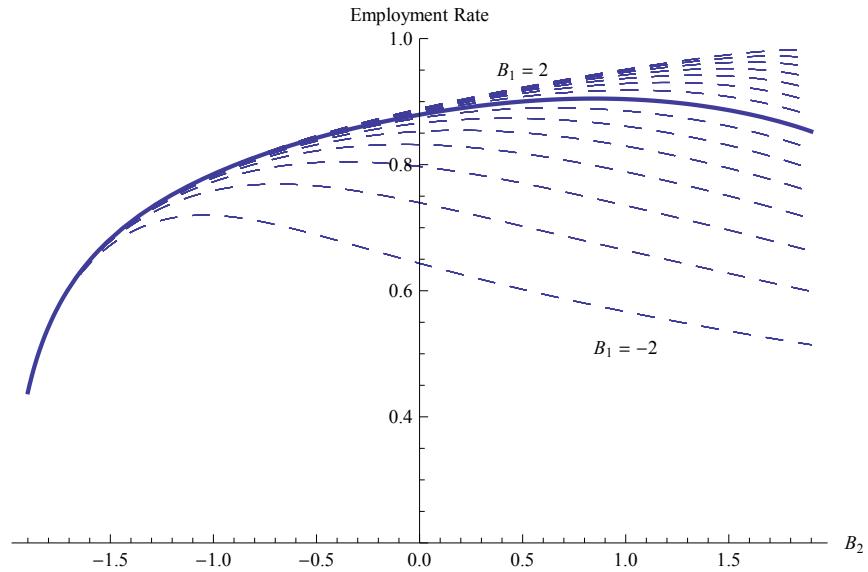
Panel B: Expected firm value in assessing a privately valued candidate
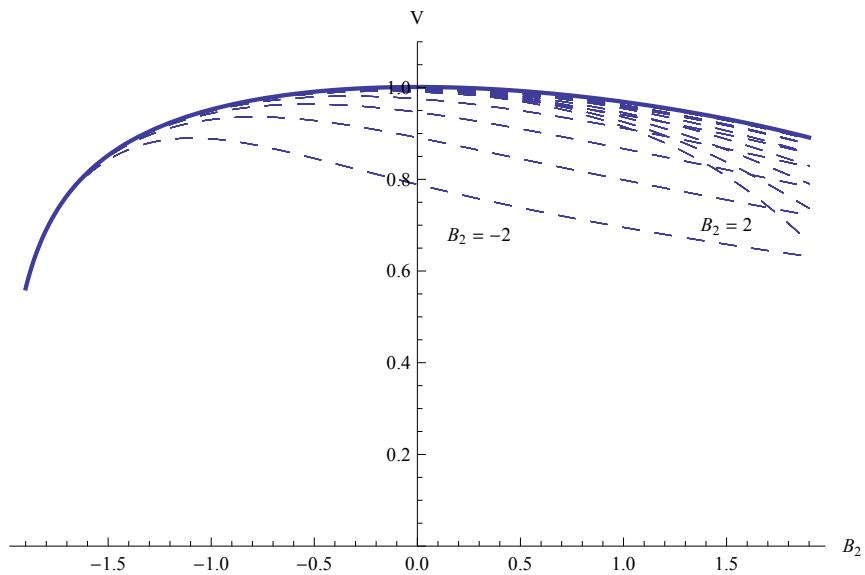
Figure 12: Optimal reservation signals and employment probabilities with bottom-up preferences ($B_2 = 0$) and a savvy Agent 2

Panel A: Optimal Reservation Signals
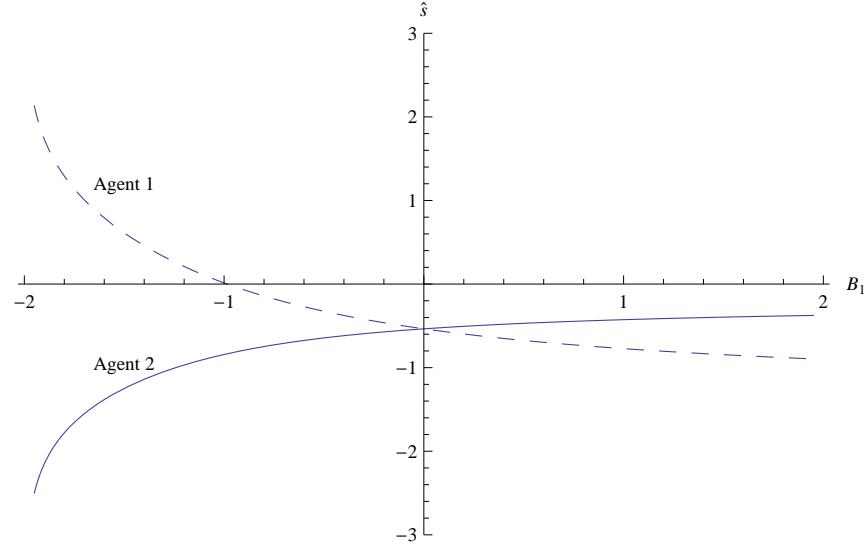
Panel B: Employment Probabilities