Rationally Inattentive Statistical Discrimination: Arrow Meets Phelps

SUBMISSION #

When information acquisition is costly but flexible, a principal may rationally acquire information that favors "majorities" over "minorities" unless the latter are strictly more productive than the former (the relative size of the groups plays no actual role). Majorities therefore face incentives to invest in becoming productive, whereas minorities are discouraged from such investments. The principal, in turn, focuses scarce attentional resources on majorities precisely because they are likely to invest. We give conditions under which a discriminatory equilibrium arises and is most preferred by the principal, despite that all groups are ex-ante identical. Our results have policy and welfare implications, as they add to the discussion of affirmative action, as well as the empirical literature on implicit bias, discrimination in subjective performance evaluation, and occupational segregation and stereotypes.

Manuscript submitted for review to the 24nd ACM Conference on Economics & Computation (EC'23).

1 INTRODUCTION

We propose a new model of statistical discrimination. A demographic group is discriminated against in the labor market because its members rationally choose to underinvest in the skills needed to succeed. Their choice is reinforced by the endogenous allocation of an employer's limited attention across groups, based on which beliefs are formed, and labor market decisions are made. In equilibrium, discriminatory (biased) attention allocation and beliefs can be optimal given the differing investments in skills between groups, who are ex-ante identical. Under some conditions, discriminatory equilibria are the most profitable to the employer.

The theory of statistical discrimination posits that groups of individuals with certain demographic traits are discriminated against in the labor market, because employers correctly infer that these groups should be treated differently. As an explanation for discrimination, it does not rely on bias or adversarial feelings towards discriminated groups, although both bias and rational beliefs may play a role in any given real-world instance of discrimination.

Economists have put forward two canonical models of statistical discrimination: the Arrovian model of coordination failure, and the Phelpsian model of information heterogeneity. Arrow [1971, 1998] argues that discrimination may arise as the result of coordination failure. One demographic collective, call it Group 1, expects to be discriminated against, and therefore does not undertake the costly investments that are needed to succeed in the labor market. Group 2 expects to be favored, and therefore finds it worthwhile to invest. Employers, in turn, rationally discriminate against Group 1 in favor of Group 2 because the latter is expected to invest and the former is not.

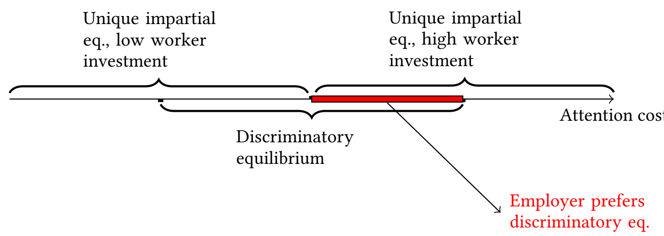
The second canonical model follows Phelps [1972] (see also Aigner and Cain, 1977) to argue that statistical discrimination emerges from differing qualities of information. Groups 1 and 2 have the same, exogenous, skill distribution, but employers have access to better-quality information about members of Group 2 than of Group 1. As a result, members of Group 2 enjoy, on average, a favorable treatment in the labor market.

The current paper combines ideas from the canonical Arrovian and Phelpsian models, with the chief aim of endogenizing employer's acquisition of information about workers' skills. In our story, workers choose whether to undertake a costly investment that results in an increased likelihood of being productive. An employer chooses a labor market outcome (a promotion decision, in our model), based on his endogenously-gleaned information about workers' productivity. Specifically, we borrow from the recent literature on rational inattention [Sims, 2003], to model how an employer chooses a costly signal structure that will inform him about workers' productivity. In equilibrium, workers' incentives to invest are affected by how they expect to be rewarded by the employer, a decision that is filtered through the endogenously-chosen information structure. In turn, the employer chooses an optimal information structure and labor market outcome, given his belief about workers' investment decisions.

We first demonstrate that there always exists an impartial equilibrium, one that is analogous to the equilibria without coordination failure in Arrow's model, but with the new feature that the information structure endogenously chosen by the employer is also impartial about groups. In an impartial equilibrium, there is neither Arrovian coordination failure nor Phelpsian information heterogeneity.

Our main results describe the emergence of a discriminatory equilibrium, one that is not impartial. In a discriminatory equilibrium, members of different groups face different incentives to undertake costly investments. Again, as in Arrow, some groups choose not to invest because they are not expected to, while others do invest, and correctly expect to be rewarded. In our model, however, workers' differing investment decisions are mirrored in the employer's choice of a discriminatory information structure—one that favors the group who chooses to invest, unless the underinvested group is strictly more productive than the former. In this way, the employer can efficiently deploy his limited attentional resources according to workers' investment decisions, focusing mainly on whether the underinvested group surprises him with a genuinely outstanding outcome. The resulting belief favors the invested group most of the time, and thus reinforces workers' expectations that they will be treated differently. The vicious circle is closed when *Arrow meets Phelps.*

The following diagram illustrates the model's behavior, as a function of an attention cost parameter that captures how costly it is for the employer to acquire information:



Our model exhibits two regimes: one in which the only equilibrium is impartial, and one where an impartial equilibrium and a discriminatory equilibrium coexist. The impartial equilibrium features high worker investments when the attention cost parameter is low, and low investments when the attention cost is high. A discriminatory equilibrium emerges when the attention cost parameter is intermediate, and it constitutes the most profitable equilibrium to the employer when it coexists with an impartial equilibrium that induces low worker investments.

The takeaways from this diagram are twofold. First, a discriminatory equilibrium can emerge in a model where a discriminatory information structure is chosen endogenously, and provides different groups of workers with different incentives to undertake costly investments. The workers are *ex-ante* identical, but are nonetheless treated differently in equilibrium.

Second, and most importantly, the discriminatory equilibrium may be strictly preferred by the employer to the impartial equilibrium. The reason is that, when the attention cost is high, the only way to maintain impartiality is to acquire noisy information that provides uniformly low incentives to all workers. Moreover, ranking these equally poorly motivated workers requires considerable time and energy from the employer, who thus prefers to live in a world in which only some workers are meticulously screened and properly incentivized, whereas others are rationally ignored and therefore underinvest. Such an outcome allows the employer to be rationally inattentive and therefore saves on attention cost, in addition to boosting the expected revenue. To the extent that employers can affect the selection of equilibrium in their interactions with workers, they may steer the system towards discrimination. This equilibrium selection feature of our model is absent from Arrow's explanation of statistical discrimination based on pure coordination failure.

Finally, we want to stress that rational inattention is essential to our story, because our game has no discriminatory equilibrium in the absence of costly information acquisition, or when the attention cost parameter is sufficiently close to zero.

Our model not only presents a novel explanation for statistical discrimination; it also provides a tractable framework to discuss various policy issues, as well as phenomena associated with labor market discrimination.

Attention and discrimination. Our model speaks to the connection between attention and discrimination, which has received a lot of attention in the empirical documentation of discriminatory outcomes. Much research in economics and psychology uses the Implicit Association Test (IAT) to detect and measure automatic, unconscious, biases, based on the premise that the latter are triggered by deficits in the decision-maker's attentional capacity [Greenwald et al., 1998]. For example, Chugh

[2004] argues that managers operate under time pressure, and that this leads to decisions that are tainted by automatic, unconscious, biases. Bertrand et al. [2005] interpret the well-known study of discrimination through African-American names of Bertrand and Mullainathan [2004], as evidence that time-constrained recruiters may allow implicit biases to guide their decisions.

Our model formally captures the attentional channel through which implicit biases could rise and fall. It predicts a non-monotonic relation between the attention cost parameter and the equilibrium degree of discrimination, as illustrated by the above diagram. In Section 3.3, we use this prediction to study the de-biasing techniques used by real-world organizations to address discrimination. The discussion therein speaks to the varying effectiveness of these techniques [Eberhardt, 2020, Greenwald and Lai, 2020], suggesting that such an ambiguity—which has annoyed and puzzled researchers and practitioners—should be the norm rather than the exception.

Affirmative action. Our model facilitates analysis of affirmative action policies, in particular the effectiveness of quotas and subsidies in eliminating a discriminatory situation. We show that mandating a quota that requires members of different groups be promoted with equal probability gives rise to a model that only has impartial equilibria. In contrast, a subsidy for promoting minorities will never achieve equity between different groups. The quota may thus seem like the best policy, except that our results regarding the most profitable equilibrium can call into question the long-term effects of affirmative action quotas, as well as the desirability of equity from the perspective of social welfare. Details are in Section 5.

Occupational discrimination. Our model can be used to capture occupational discrimination. There is clear evidence that men and women work on very different jobs, even within narrowly defined industries or firms [Blau and Kahn, 2017]; their performance evaluations are based on stereotypical traits, and overlook their achievements in counter-stereotypical tasks [Bohnet et al., 2016, Correll et al., 2020]. In Section 6, we extend our model to allow for multidimensional tasks and skills, and show that a similar mechanism to the one generating discriminatory outcomes in our main model, can also explain why different categories of workers (who are ex-ante identical) invest in different skills and are assigned different tasks. The channel is, again, that a rationally inattentive employer may save on attention cost by selectively paying attention to different workers in the assignment of different tasks. The employer's selectivity is then mirrored in workers' differential investments in task-specific skills, which, in equilibrium, gives rise to occupational segregation and stereotypes. Our results, as well as a detailed discussion, can be found in Section 6.

1.1 Related Literature

The current paper adds to four strands of the literature: statistical discrimination, rational inattention, and fairness (in AI).

Statistical discrimination. The literature on statistical discrimination is too vast to be exhausted here—we refer the reader to the surveys by Fang and Moro [2011] and Onuchic [2022]. Instead, we focus on the direct precedents to our work, and discuss the papers most closely related to ours.

The most important precedent to our work is Coate and Loury [1993]. These authors develop an Arrovian model of statistical discrimination, whereby each worker is assigned to either a standard task or a challenging task based on the realization of an exogenous, imperfect, signal of his skill. Discrimination emerges again as the consequence of coordination failure. One can think of our model of a variant of Coate and Loury's model, with two major differences: first, the signal structure is endogenously chosen by the principal à la rational inattention; second, agents compete in a tournament rather than being assigned to jobs on an individual basis. As will be discussed shortly, both differences are crucial for our result concerning discrimination as the most profitable equilibrium.[1] The model of Coate and Loury has been extended by several authors: for example by Fang [2001] to endogenous group identities, and by Chaudhuri and Sethi [2008] to allow for peer effects. The issue of endogenous information structure has, however, not been analyzed until recently (more on this later).

Our work provides a new foundation for the discriminatory information structure assumed by Phelpsian models of statistical discrimination. Recently, Chambers and Echenique [2021] examine Phelpsian statistical discrimination from the angle of information design, but they do not propose a strategic model of labor market outcomes and instead connect the presence of Phelpsisan statistical discrimination to the problem of identifying a skill distribution. Escudé et al. [2022] further the connection to Blackwell's theorem, and provide a more nuanced relation between discrimination and informativeness than allowed for in Chambers and Echenique.

Rational inattention. The literature on rational inattention (RI) pioneered by Sims [2003] has grown rapidly in recent years; see Maćkowiak et al. [2021] for a survey. We use the ideas and techniques developed in this literature to study statistical discrimination. Conceptually, our results concerning the discriminatory equilibrium exploit the flexibility associated with RI information acquisition, which has proven crucial for shaping the outcomes of financial contracting, political competition, and ultimatum bargaining [Hu et al., ming, Ravid, 2020, Yang, 2020]. Empirical relevance of attentional flexibility has been established by the lab experiments conducted by Ambuehl [2017] and Dean and Neligh [2019], and by Eberhardt [2020] in the context of discrimination. Technically, Matějka and McKay [2015] and Yang [2020] provide a complete characterization of the optimal signal structure for binary RI decision problems; our analysis builds on their results.

Recently, Bartoš et al. [2016] and Fosgerau et al. [ming] propose models of job market outcomes in which employers choose costly information structures, but the mechanisms they seek to explore differ from ours.[2] In the model of Bartoš et al. [2016], employers screen job applicants, taking into account the exogenous difference between groups. Stylized, costly, information acquisition (in the form of variance reduction of a normal random variable) is shown to magnify this difference, as employers acquire information about the disadvantageous group when the market is slack, and ignore them when the market is tight.

Fosgerau et al. [ming] study an Arrovian model where a screener acquires costly information about job candidates on an individual basis, in a model with a continuum of agents, and allowing a general posterior-separable attention cost. The focus of their comparative statics results is not on when the screener can sustain discrimination among ex-ante identical groups as the most profitable equilibrium, but on how rational inattention interacts with several natural intrinsic differences between groups: differential screening costs, prejudice, and asymmetric access to social capital. A key difference with our paper, is that the screener considers each agent individually, while in our model agents compete for a limited opportunity. Under rational inattention, this becomes a competition for the principal's limited attention. Using a discriminatory signal structure to screen and select, the principal saves on attention cost, and can sometimes sustain discrimination as the most profitable equilibrium among ex-ante identical agents. The channel uncovered in our paper has not been explored by the existing literature on rational inattention and statistical discrimination.

[1]We are not the first to study Arrovian discrimination with tournament being the incentive scheme. de Haan et al. [2017] examine, theoretically and experimentally, the stability of equilibria in a variant of Coate and Loury's model, whereby agents can undertake continuous investments to improve their chances of winning a tournament, and the principal's decision is made based on an exogenous signal structure. The focus here is on how rational inattention could bias the equilibrium signal structure and effort choices.

[2]Recently, Bartoš et al. [2016], Glover et al. [2017], and Huang et al. [2022] provide evidence for attentional discrimination using field experiments and administrative data.

The work of Matveenko and Mikhalishchev [2021] studies how imposing quotas on average decision probabilities affects the solution to the rational inattention problem studied by Matějka and McKay [2015]. Our analysis on affirmative action quotas builds on their analysis, but enriches it with the endogenous investments by agents.

Incentive contracting. We join Alchian and Demsetz [1972] in formalizing the role of monitoring cost in shaping the organization of multi-agent production. Most existing contract theory papers restrict the principal to drawing signals from exogenously given distributions, with or without an information acquisition cost. Li and Yang [2020] consider optimal incentive contracting between a rationally inattentive principal and a single agent, whereby both the incentive scheme and monitoring structure are part of the principal's strategic planning. Here, the incentive scheme is taken as given, and the focus is on the optimal information structure that guides the competition between multiple agents.

We study a contracting problem with subjective monitoring, whereby the principal's choice of the signal structure is his private information. The existing literature on incentive contracting with subjective performance evaluation focuses on three kinds of contractual frictions: (1) the non-verifiability of subjective performance evaluation, (2) the unobservability thereof, and (3) the favoritism practiced by supervisors (see Prendergast, 1999 for a survey). We instead formalize the role of costly and yet flexible monitoring in engendering biased subjective performance evaluation, a result that is, to the best of our knowledge, new to the literature.

The theory of contests has been widely applied to the study of affirmative action policies that level the playing field for heterogeneous participants. Most existing studies attribute the biases of optimal contests to the asymmetry between contestants or the favoritism practiced by the principal; see Chowdhury et al. [2020] for a survey. Recently, a few authors start to realize that the optimal contest between ex-ante identical agents can be biased, provided that the principal's objective is sufficiently general [Drugov and Ryvkin, 2017], or there are sufficiently many agents [Fu and Wu, 2020]. We consider a stylized tournament between two ex-ante identical agents, and use the rational inattention of the principal to bias the optimal tournament.

Fairness. There is a large and active literature on fairness in machine learning and algorithm design, whose chief aim is to formulate algorithmic criteria and solutions that will undo the biases that algorithms may inherit from the sources they are based on [Barocas et al., 2019, Chouldechova and Roth, 2018]. The effort of the literature is therefore mainly normative—see, for example, the arguments articulated by Dwork et al. [2012] and Kearns and Roth [2019], which seek to develop the normative basis for fairness in the design of classification algorithms. When it comes to uncovering the reasons behind discriminatory outcomes, the literature focuses on the properties of training data and other issues that are relevant for algorithmic decision-making. Our model is a complementary effort to understand a novel channel of discrimination that stems from rational yet imperfect human decision-making.

2 MODEL

We study a game between three players: a principal and two agents, who are called Michael (m) and Wendy (w). The principal employs Michael and Wendy, and has to choose one of them to promote. The promotion decision serves to induce the agents to exert effort so as to be more productive. It delivers a unit benefit to the promoted agent, as well as the agent's productivity to the principal.[3]

[3]One can broadly interpret the promotion opportunity as a reward (e.g., a bonus payment) that motivates agents to exert effort. For concreteness' sake, we shall stick to the interpretation of promotion throughout.

Specifically, each agent $i \in \{m, w\}$ chooses a level of *effort* $\mu_i \in \{\underline{\mu}, \overline{\mu}\}$, with $0 < \underline{\mu} < \overline{\mu} < 1$, at a cost $C(\mu_i)$. Suppose that $C(\underline{\mu}) = 0$ and $C(\overline{\mu}) = C > 0$. The effort μ_i determines a random *productivity* $\theta_i \in \{0, 1\}$ for agent i, with μ_i being the probability that $\theta_i = 1$ and $1 - \mu_i$ the probability that $\theta_i = 0$. Given the profile $\boldsymbol{\mu} = (\mu_m, \mu_w)$, each θ_i is drawn independently.

The principal does not known the values of θ_m and θ_w, but can choose to acquire information about them. Information, however, is costly. Given the information that the principal gleans about $\boldsymbol{\theta} = (\theta_m, \theta_w)$, he chooses $a \in \{0, 1\}$, where $a = 0$ means that Wendy is promoted, and $a = 1$ means that Michael is promoted.

The principal's information acquisition is modeled as the choice of a signal structure $\pi : \{0, 1\}^2 \to \Delta(S)$, which maps each profile of productivity values to a random signal taking values in a set S. We assume that S is finite and that $|S| \geq 2$; later we shall demonstrate that these assumptions about S are without loss of generality. Otherwise we impose no restriction on the signal structure π. A promotion rule is a function $a : S \to \Delta(\{0, 1\})$, which maps each signal realization to a (random) decision on whether to promote Michael or Wendy.

Given a profile $\boldsymbol{\mu}$ of effort choices by the agents, the principal's expected payoff is

$$\mathbb{E}\left[\tilde{a}\tilde{\theta}_m + (1 - \tilde{a})\tilde{\theta}_w | \boldsymbol{\mu}, \pi, a(\cdot)\right] - \lambda I(\pi | \boldsymbol{\mu}),$$

where $\lambda > 0$ parameterizes the cost of information acquisition, and is henceforth referred to as the *attention cost parameter*; and I is the mutual information between the random productivity profile $\tilde{\boldsymbol{\theta}}$ and the random signal generated by π. In words, the principal's payoff equals the productivity of the promoted agent, which is estimated according to the information generated by the signal structure of his choice. As the latter becomes more informative of agents' productivities, the cost of information acquisition increases.

The game begins with the principal and agents moving simultaneously: The former specifies a signal structure π and a promotion rule $a(\cdot)$, whereas the latter make effort choices μ_is. After that, productivities and signals are realized, and the principal makes a promotion decision. When choosing, the principal observes neither agent's effort or productivity, thus facing a moral hazard problem. Agents do not observe the principal's choice of the signal structure or promotion rule—an assumption that reflects the subjective nature of employee evaluation and promotion in practice.

Our solution concept is *pure strategy Bayes Nash equilibrium* (hereinafter, equilibrium for short). When multiple equilibria coexist, we characterize them all, with a particular focus on the *most profitable equilibrium to the principal.*

3 RESULTS

3.1 Preliminaries

First, it is helpful to simplify the principal's strategy. Define $\Delta\theta = \theta_m - \theta_w$ as the differential productivity between m and w, and note that $\Delta\theta \in \{-1, 0, 1\}$. For any given effort profile $\boldsymbol{\mu}$, rewrite the principal's expected payoff as

$$\underbrace{\mathbb{E}\left[\tilde{a}\Delta\tilde{\theta} \mid \boldsymbol{\mu}, \pi, a(\cdot)\right] + \mu_w}_{\text{Expected revenue}} - \lambda I(\pi \mid \boldsymbol{\mu}),$$

and note that the expected revenue depends on his strategy $(\pi, a(\cdot))$ and $\boldsymbol{\mu}$ jointly only through $\Delta\theta$. By Matějka and McKay [2015], we may restrict attention to signal structures that prescribe a (random) *promotion recommendation* to the principal based on the differential productivity between m and w, i.e., $\pi : \{-1, 0, 1\} \to \Delta(\{0, 1\})$. Intuitively, any information beyond the aforementioned is redundant and therefore shouldn't be acquired. Moreover, promotion recommendations must be *strictly obeyed* by the principal, i.e., $a(1) = 1$ and $a(0) = 0$, because otherwise the principal has a (weakly) preferred candidate regardless of the promotion recommendations, and can therefore always promote that agent without acquiring information. Hereinafter, we shall represent the principal's strategy by $\pi : \{-1, 0, 1\} \to [0, 1]$, where each $\pi(\Delta\theta)$, $\Delta\theta \in \{-1, 0, 1\}$, specifies the probability that m is recommended for promotion when the differential productivity between m and w equals $\Delta\theta$.

Next are the key concepts that embody the notion of discrimination.

Definition 3.1. A signal structure π is *impartial* if the probability of promoting an agent depends *only* on his or her productivity difference with the other agent, and *not* on agents' identities. That is, $\pi(\Delta\theta) = 1 - \pi(-\Delta\theta)\ \forall \Delta\theta \in \{-1, 0, 1\}$. Otherwise π is *discriminatory.*

Definition 3.2. An equilibrium is *impartial* (resp. *discriminatory*) if the equilibrium signal structure is impartial (resp. discriminatory).

As will later be demonstrated, an impartial equilibrium must induce the same level of effort from both agents, whereas a discriminatory equilibrium must induce different levels of effort from the two agents. By symmetry, it is without loss of generality (w.l.o.g.) to focus on discriminatory equilibria that induce high effort from m and low effort from w—a convention we will follow in the remainder of the paper.

Lastly we introduce a regularity condition that will be maintained throughout the paper. For ease of notation, we write $\Delta\mu$ for $\overline{\mu} - \underline{\mu}$, c for $C/\Delta\mu$, A for $\overline{\mu}(1 - \underline{\mu})$, and B for $\underline{\mu}(1 - \overline{\mu})$:

Assumption 3.1. *$\overline{\mu} + \underline{\mu} > 1$ and $c < \overline{\mu}(1 - \overline{\mu})/(A + B)$.*

The role of Assumption 3.1 will be discussed in Appendix A.

3.2 Main Results

We present our two most important results. The first concerns the existence and uniqueness of impartial and discriminatory equilibria. The second pinpoints the most profitable equilibrium to the principal.

Theorem 3.3. *Fix any C, $\overline{\mu}$, and $\underline{\mu}$ that satisfy Assumption 3.1. These determine values $\underline{\lambda}$, $\overline{\lambda}$, and λ^* of the attention cost parameter for which $0 < \underline{\lambda} < \overline{\lambda} < +\infty$ and $\lambda^* > 0$. The following statements are true:*

1. *An impartial equilibrium always exists, and it is unique if and only if $\lambda \neq \lambda^*$. When unique, the impartial equilibrium sustains the high-effort profile $(\overline{\mu}, \overline{\mu})$ if the attention cost parameter is low, namely $\lambda < \lambda^*$; and it sustains the low-effort profile $(\underline{\mu}, \underline{\mu})$ if the attention cost parameter is high, i.e., $\lambda > \lambda^*$.*
2. *A discriminatory equilibrium exists if and only if the attention cost parameter is intermediate, i.e., $\lambda \in [\underline{\lambda}, \overline{\lambda}]$. Whenever a discriminatory equilibrium exists, it is unique.*
3. *$\underline{\lambda} < \lambda^*$ always holds. $\lambda^* < \overline{\lambda}$ holds if and only if $\underline{\mu} > 1/2$ and Condition (4) in Appendix B holds.*

Theorem 3.4. *Let everything be as in Theorem 3.3, and suppose that $\lambda^* < \overline{\lambda}$. Then the most profitable equilibrium to the principal is discriminatory if and only if $\lambda \in (\lambda^*, \overline{\lambda}]$.*

To better understand the intuitions behind these results, we first restrict the principal to using impartial signal structures. Under this restriction, the signal acquired by the principal becomes less informative about agents' productivities (in the sense of Blackwell) as the attention cost parameter increases. Agents best respond by exerting high effort when the attention cost parameter is low, and low effort when the attention cost parameter is high. The symmetry in agents' effort choices, in turn, justifies the use of an impartial signal structure to begin with. The two regimes are separated by the threshold value $\lambda^* > 0$, at which the game has two impartial equilibria. For all $\lambda \neq \lambda^*$, the impartial equilibrium is unique.

We next allow the principal to use discriminatory signal structures. When the attention cost parameter is intermediate, i.e., $\lambda \in [\underline{\lambda}, \overline{\lambda}]$, the principal can sustain a discriminatory effort profile through conducting discriminatory performance evaluations.

As an illustration, consider the numerical example in Table 1, which takes a discriminatory effort profile as given and solves for the optimal signal structure, i.e., one that maximizes the principal's expected profit:

Table 1. Optimal signal structure for $\boldsymbol{\mu} = (\overline{\mu}, \underline{\mu}) = (.8, .6)$, $\lambda = .3$.

$\Delta\theta$	1	0	-1
$\mathbb{P}(\Delta\theta \mid \boldsymbol{\mu})$	.32	.56	.12
$\pi(\Delta\theta)$	.98	.74	.09

Since m is known to work harder than w, promoting m over w is the safer choice for the principal. In consequence, a rationally inattentive principal will favor m, unless w is strictly more productive. As depicted in Table 1, doing so does not require a careful distinction between whether m is more productive than, or equally productive as w (indeed $\pi(1) = .98$ is not very different from $\pi(0) = .74$), and therefore saves on information acquisition cost. At the same time, the signal structure still does a decent job in selecting the most productive agent, as it generates an expected revenue of .90, compared to the expected revenue .92 in the benchmark case where information acquisition is costless. While w is strongly favored by the principal when she is strictly more productive than m (i.e., $\pi(-1) = .09$), that event occurs with a small probability because m works harder than w. w is treated unfavorably otherwise. In particular, and importantly, this occurs when she is as productive as m (i.e., $\pi(0) = .74$).

Turning to the agents' incentives to exert effort, under the above numerical assumptions, w can increase her winning probability by

$$\Delta\mu[\overline{\mu}(\pi(1) - \pi(0)) + (1 - \overline{\mu})(\pi(0) - \pi(-1))] = .081$$

if she exerts high effort rather than low effort. The analogous increase for m is

$$\Delta\mu[(1 - \underline{\mu})(\pi(1) - \pi(0)) + \underline{\mu}(\pi(0) - \pi(-1))] = .098.$$

If $C \in (.081, .098)$, then it is indeed optimal for m to exert high effort and w low effort. In turn, this justifies the principal's use of the discriminatory signal structure that favors m.

Taken together, our main results present an important lesson: Discrimination in labor market outcomes could stem from the discrimination in information acquisition. Conducting discriminatory performance evaluations allows the principal to be rationally inattentive and to sustain a discriminatory effort profile in equilibrium. Compared to the impartial equilibrium that sustains the low effort profile, the discriminatory equilibrium enjoys both a *revenue advantage* and a *cost advantage*. That is, it generates higher expected revenue to the principal and incurs a lower attention cost, thus constituting the most profitable equilibrium to the principal when both kinds of equilibria coexist (i.e., when $\lambda \in (\lambda^*, \overline{\lambda}]$). The comparison between the discriminatory equilibrium and the impartial equilibrium that sustains the high effort profile is more delicate, because the former has, roughly speaking, a revenue disadvantage and a cost advantage over the latter. As it turns out, the cost-saving concern is always of a secondary importance. Hence the discriminatory equilibrium is the least profitable to the principal when these kinds of equilibria coexist (i.e., when $\lambda \in [\underline{\lambda}, \lambda^*]$).

3.3 Implications

Our results have direct implications for the literatures on implicit bias and discrimination in subjective performance evaluation. We proceed to discuss these connections in detail, as well as the welfare consequences of our results.

Implicit bias, stereotype, and the effectiveness of de-biasing programs. Perhaps the most obvious implication of our results is the connection between attention and discrimination. Many scholars, across multiple disciplines, have advanced the notion that limited attention triggers implicit biases and stereotypes. The idea is that in attempting to make sense of other people, we regularly construct and use categorical representations to simplify our process of perception. This mode of thought, formally known as social categorization, offers tangible cognitive benefits such as rapid inferences, and the efficient deployment of limited processing resources.[4] A popular idea among social psychologists is that the activation of social categories is modulated by the availability of attentional resources [Greenwald and Banaji, 1995, Macrae and Bodenhausen, 2000]. Deficits in the attentional capacity increase the likelihood that decision-makers will apply stereotypes when dealing with others—an idea that lays the foundation for the Implicit Association Test (IAT), developed by Greenwald et al. [1998] to detect and measure automatic, unconscious, biases.

Evidence on the connection between attention and discrimination abounds. In human resource management, Chugh [2004] argues that managers operate under time pressure, and that this leads to decisions that are tainted by automatic, unconscious, biases. Bertrand et al. [2005] interpret the well-known study of discrimination through African-American names of Bertrand and Mullainathan [2004], as evidence that time-constrained recruiters may allow implicit biases to guide their decisions. Similar arguments have been used to explain the discriminatory practices observed in other contexts, such as criminal justice, education, healthcare, and sport [Chapman et al., 2013, Eberhardt, 2020, Price and Wolfers, 2010, Warikoo et al., 2016]. Yet despite the richness of evidence, a theory that establishes the causal link between limited attention and implicit bias is lacking. Our results fill this intellectual gap by establishing a formal, explicit, mechanism through which costly and yet flexible attention allocation gives rise to discrimination.

Our model predicts a nonmonotonic relation between the attention cost parameter and the equilibrium degree of discrimination. Recall the statement of Theorem 3.3, or the diagram in the introduction. The non-monotone nature of the comparative-statics speaks to the varying effectiveness of the de-biasing techniques used by real-world organizations to address discrimination. On the positive side, the Oakland Policy Department recently adjusted its foot pursuit policy so that officers could no longer follow suspects as they run into backyards or blind alleys. Instead, officers were instructed to "step back, slow down, call for backup, and think it through." According to Eberhardt [2020], this simple adjustment, together with use of body cameras to monitor the languages used by police officers, has not only has led to fewer civilians being shot but also has made cops safer: Injuries among officers dropped by 70 percent, and the number of officer-involved shootings fell dramatically, from an average of eight every year to about the same number in

[4]Fryer and Jackson [2008] propose a model of social categorization, based on the idea that the same rule of simplification must be applied across multiple social contexts (e.g., how one should interact with people with different races during and after work is governed by the same rule). In contrast, rationally inattentive information aggregation is context-dependent.

the past five years, even while the arrest rate held steady and crime levels have fallen.[5] On the negative side, Greenwald and Lai [2020] recently concluded, based on their meta analysis of the implicit bias training programs offered in corporations, nonprofit organizations, hospitals, public welfare organizations, schools, universities, court systems, and police departments, that "The popular media often suggests relying on one's own resources to intercept implicit biases—perhaps by pausing to think deliberately or by meditating before making decisions that might adversely affect others. Convincing evidence for the effectiveness of these strategies is not yet available in peer-reviewed publications." Our results put these seemingly conflicting findings into perspective, and suggest that they may share the same root. Rather than to abandoning the premise that limited attention triggers implicit biases (as suggested by Greenwald and Lai, 2020), an alternative way to reconcile the aforementioned findings is to recognize that the exact relation between attention and discrimination is more nuanced than previously thought.

Gender and racial gap in subjective performance evaluation. Gender and racial stereotypes continue to disadvantage women and minorities through biased performance appraisals. This unfortunate reality does not surprise many managers: In one recent survey conducted by Mackenzie et al. [2019], only 15% of women and 24% of men managers had confidence in the performance evaluation process, while most viewed it as subjective and highly ambiguous. Our model formalizes a causal link between limited managerial attention and biased subjective performance evaluation. To the extent that subjective performance evaluation affects labor market outcomes, such as termination, bonus pay, and career trajectories [Baker et al., 1988, Prendergast, 1999], our model sheds light on the on the role of limited managerial attention in shaping the various labor market outcomes.

Years of sociological research into the gender and racial gap in subjective performance evaluations has established several salient patterns. For example, women tend to get shorter, more vague, and less constructive critical feedback during performance reviews, which inhibits their ability to learn what they need to do to grow and advance [Correll and Simard, 2016, Jampol and Zayas, 2021]. They are also held to higher performance standards, and face increased scrutiny and shifting criteria when being evaluated [Wynn and Correll, 2018]. In a related study, Upton and Arrington [2012] find a negative relation between balanced scorecard performance evaluations and evaluators' racial biases measured by IAT.

Our model speaks to these stylized facts. It predicts that minorities are rated more harshly than majorities in the discriminatory equilibrium. Indeed, in our model, minority workers are recognized by the employer only when they truly are strictly more productive than their majority counterparts. Otherwise their chances of getting a promotion is slim, which is the case even if they are equally productive as their majority counterparts. Such a hurdle discourages minorities from undertaking costly investments. In equilibrium, minorities are promoted less frequently and earn less on average than majorities.

Welfare. An important implication of our results—which sets them apart from the standard Arrovian mechanism such as Coate and Loury [1993]—is that one cannot Pareto rank the various kinds of equilibria generated by the model. While coordination failure is clearly part of our story, the welfare implications of our results are more subtle. This is illustrated by Figure 1, which plots the varying welfare regimes against the attention cost parameter.

From the principal's perspective, he most prefers the impartial equilibrium that sustains the high effort profile, followed by the discriminatory equilibrium, and then the impartial equilibrium that

[5]Relatedly, Nextdoor Neighbor, a social network (developed by Meta, formerly Facebook) for residential neighbors to communicate through, recently started asking its users to provide detailed descriptions about the suspicious activities they wish to report to the system, because "adding frictions allows users to act based on information rather than instinct."

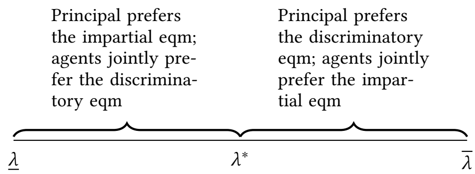


Fig. 1. Welfare regimes.

sustains the low effort profile. Meanwhile, agents together most prefer the impartial equilibrium that sustains the low effort profile, followed by the discriminatory equilibrium, and finally the impartial equilibrium that sustains the high effort profile (as the sum of their expected utilities in these equilibria are 1, $1-C$, and $1-2C$, respectively). Thus depending on the welfare weights of the principal and agents, reduced discrimination may either enhance or undermine utilitarian social welfare. This finding further complicates the picture painted by our results, suggesting that the aforementioned de-biasing programs might not only send the equilibrium degree of discrimination in the wrong direction, but might also have unintended welfare consequences.

4 ANALYSIS

This section provides a detailed analysis of Theorem 3.3. The proof of Theorem 3.4 is more technical and is relegated to Appendix B.

For ease of notation, we shall hereinafter write $\overline{\pi}$ for the average probability that an arbitrary signal structure π recommends m for promotion, as well as X for $\pi(1)-\pi(0)$ and Y for $\pi(0)-\pi(-1)$. Then π is impartial if and only if $\pi(0)=1/2$ and $X=Y$. We will also write γ for $\exp(1/\lambda)$ and note that γ is strictly decreasing in λ, $\gamma \to +\infty$ as $\lambda \to 0$, and $\gamma \to 1$ as $\lambda \to +\infty$.

4.1 Best Response Functions

Consider first the problem faced by the principal, holding agents' effort profile $\boldsymbol{\mu}$ fixed. Call the solution to this problem the *optimal signal structure* for $\boldsymbol{\mu}$. By Matějka and McKay [2015], this signal structure is either degenerate, satisfying $\pi(\Delta\theta) \equiv 0$ or 1, or it is nondegenerate and satisfies $\pi(\Delta\theta) \in (0,1)\ \forall \Delta\theta$. The next lemma solves for the optimal signal structure for every effort profile.

Lemma 4.1. (1) *The optimal signal structure for $(\overline{\mu}, \overline{\mu})$ or $(\underline{\mu}, \underline{\mu})$ is nondegenerate and impartial. It satisfies $\overline{\pi} = \pi(0) = 1/2$ and $X = Y = g(\gamma)$, where*

$$g(\gamma) = \frac{\gamma - 1}{2(\gamma+1)} \text{ satisfies } g > 0 \text{ and } \frac{dg(\gamma)}{d\lambda} < 0\ \forall \lambda > 0.$$

(2) *The optimal signal structure for $(\overline{\mu}, \underline{\mu})$ is degenerate if $\lambda \geq \check{\lambda} = (\ln(A/B))^{-1} > 0$, and it is nondegenerate otherwise. In the second case, the signal structure is discriminatory and satisfies $\overline{\pi} = \pi(0) = (\gamma A - B)[(\gamma-1)(A+B)]^{-1} \in (1/2, 1)$ and $X = f(\gamma) < Y = Af(\gamma)/B$, where*

$$f(\gamma) = \frac{(\gamma A - B)(\gamma B - A)}{(\gamma^2-1)(A+B)A} \text{ satisfies } f > 0 \text{ and } \frac{df(\gamma)}{d\lambda} < 0\ \forall \lambda \in (0, \check{\lambda}).$$

Lemma 4.1 conveys three important messages. First, in the case where an optimal signal structure is nondegenerate, the conditional probability that it recommends m for promotion is strictly increasing in the differential productivity between m and w, i.e., $X, Y > 0$. When both agents attain the same level of productivity, the conditional probability that m is recommended for promotion equals the average probability, i.e., $\pi(0) = \overline{\pi}$. In light of these findings, we shall hereinafter interpret

X as the extent to which outperforming w increases m's promotion probability above the average, and Y as the extent to which underperforming w reduces m's promotion probability below the average.

Second, the optimal signal structure is impartial when both agents exert the same level of effort, and it is discriminatory otherwise. The first result is easy to understand. To gain insights into the second result, notice that when m is more hard-working than w, promoting m is a safe option. The optimal signal structure favors m unless w is strictly more productive, as doing so doesn't require a careful distinction between whether m is strictly more productive than, or equally productive as w (i.e., X is small), and therefore saves on information acquisition cost. At the same time, it still does a decent job in selecting the most productive agent, since m works harder than w after all. While w is strongly favored by the principal when she is strictly more productive than m (i.e., Y is large), that event occurs with a small probability because m works hard. w is treated unfavorably otherwise and, in particular, when she is equally productive as m (i.e., $\pi(0), \pi(1) > 1/2$). Since $\pi(0) = \overline{\pi}$, w is also treated less favorably on average.

Finally, as the attention cost parameter λ increases, any optimal signal structure becomes "noisier," in that the conditional probabilities that it recommends the most productive agent for promotion become more concentrated around the average probability, i.e., X and Y are both decreasing in λ.

We next turn to agents' best response functions. The next lemma solves for an agent's best response to a given signal structure and the other agent's effort choice.

Lemma 4.2. *Fix any signal structure π. For any $\mu_w \in \{\overline{\mu}, \underline{\mu}\}$, m prefers to exert high effort rather than to exert low effort if and only if*

$$(1 - \mu_w)X + \mu_w Y \geq c. \tag{IC$_m$}$$

For any $\mu_m \in \{\overline{\mu}, \underline{\mu}\}$, w prefers to exert high effort rather than to exert low effort if and only if

$$\mu_m X + (1 - \mu_m)Y \geq c. \tag{IC$_w$}$$

From m's perspective, X is a carrot that is effective when w has a low productivity (hence m can outperform w and raise his chance of getting promoted), whereas $-Y$ is a stick that is effective when w has a high productivity. The overall incentive power that a signal structure provides to him is thus $(1 - \mu_w)X + \mu_m Y$. By exerting high effort rather than low effort, m can increase his chance of getting promoted by $\Delta\mu[(1 - \mu_w)X + \mu_m Y]$. In the case where $(1 - \mu_w)X + \mu_m Y$ exceeds the effective cost $c = C/\Delta\mu$ of exerting high effort, exerting high effort is optimal for m.

The problem faced by w can be solved analogously. In case π is an optimal signal structure, Lemma 4.1 implies that sustaining high effort becomes harder as λ increases.

4.2 Equilibria

The analysis differs, depending on whether the equilibrium is impartial or discriminatory.

Consider first the impartial case, in which the optimal signal structure satisfies $X = Y = g(\gamma)$. It induces both agents to exert high effort if $g(\gamma) \geq c$, and low effort if $g(\gamma) \leq c$. The two regimes are separate by a single threshold:

$$\lambda^* = (\ln g^{-1}(c))^{-1},$$

at which the game has two impartial equilibria. For all $\lambda \neq \lambda^*$, the impartial equilibrium is unique.

The discriminatory case is illustrated by Figure 2. On the one hand, (X, Y) must lie in the grey area in order to satisfy both agents' incentive compatibility constraints. On the other hand, it must lie on the red ray $Y = AX/B$ in order for the signal structure to be optimal. Since $\overline{\mu} + \underline{\mu} > 1$, the grey area must lie above the 45-degree line. Then from $A > B$, it follows that the red ray must cross the

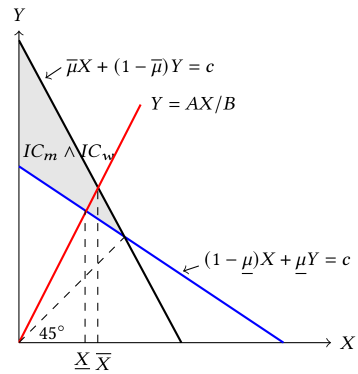


Fig. 2. Analysis of the discriminatory case.

grey area twice, at $(\underline{X}, A\underline{X}/B)$ and $(\overline{X}, A\overline{X}/B)$, respectively. Thus for any $X = f(\gamma) \in [\underline{X}, \overline{X}]$, the profile $(X, AX/B)$ can arise in an equilibrium. The last condition is equivalent to $\lambda \in [\underline{\lambda}, \overline{\lambda}]$, where

$$\underline{\lambda} = (\ln f^{-1}(\overline{X}))^{-1} \text{ and } \overline{\lambda} = (\ln f^{-1}(\underline{X}))^{-1}.$$

It remains to sign and rank $\underline{\lambda}$, λ^*, and $\overline{\lambda}$. This step is technical and is relegated to Appendix B. The regularities of these thresholds are ensured by Assumption 3.1, whose role is discussed in Appendix A.

5 AFFIRMATIVE ACTION POLICIES

This section uses the model to evaluate two affirmative-action policies that have been used to deal with discrimination in practice: quotas and subsidies.

5.1 Quotas

Suppose that the principal faces a hard quota mandating that the two agents must be promoted with equal probability on average:

$$\overline{\pi} = 1/2. \tag{Q}$$

The next theorem pinpoints the channel through which the promotion quota operates in our model.

THEOREM 5.1. *Under the assumption that $\overline{\mu} + \underline{\mu} > 1$, a strategy profile constitutes an equilibrium of the game with (Q) if and only if it is an impartial equilibrium of the baseline model.*

Time has not quelled controversy over affirmative action policies since their introductions in the 1960s and 1970s. Recent studies seek to understand the channels through which these policies operate, as well as the duration of their effects (see Holzer and Neumark, 2000, Fang and Moro, 2011, and Doleac, 2021 for surveys). Among others, Miller [2017] argues that affirmative action policies operate through inducing firms to undertake long-term investments in their employee screening procedures. This finding is reinforced by Dianat et al. [2022], who find, in the absence of the screening channel, that affirmative action can have only temporary effects.

Theorem 5.1 adds to this debate. It shows that in the current context, the promotion quota operates only through eliminating the discriminatory equilibrium. No additional effect should be expected, since the policy intervention does not impact on any impartial equilibrium, or generate any new equilibria as a byproduct. While the first two findings are somewhat anticipated, the last one is more subtle and sets our analysis apart from alternative models of Arrovian discrimination.[6]

As for the duration of quota's effect, our model offers a bleak possibility: In the case where the discriminatory equilibrium is the most profitable to the principal, lifting the quota will probably reverse its effect, as the principal's ultimate goal is best achieved by the discriminatory equilibrium. Such a reversal may not be welfare detrimental though, since we cannot Pareto rank the discriminatory equilibrium against the impartial equilibria in general. In the case where the discriminatory equilibrium attains the greatest level of social welfare, affirmative action may achieve its goal, but at a cost.

The remainder of this section provides a more detailed analysis of Theorem 5.1. Our starting observation is that any impartial equilibrium of the baseline model satisfies (Q) and therefore remains an equilibrium with and without the quota. It remains to show that the discriminatory effort profile $(\overline{\mu}, \underline{\mu})$ cannot be sustained in any equilibrium in the presence of the quota. To prove this claim, we first characterize the optimal signal structure for $(\overline{\mu}, \underline{\mu})$ that satisfies (Q) in the next lemma.

Lemma 5.2. *The optimal signal structure for $(\overline{\mu}, \underline{\mu})$ subject to (Q) is unique and satisfies $\pi(0) < 1/2$ and $X > Y > 0$.*

Comparing and contrasting Lemmas 4.1 and 5.2 reveals how the introduction of the quota reverses the situations face by m and w. Without the quota, m is favored by the principal unless w is strictly more productive. In the presence of the quota, it is w who is favored by the principal (i.e., $\pi(0) < 1/2$ and $Y > 0$), unless m is strictly more productive (i.e., $X > 0$). While m is strongly favored by the principal in the last event (i.e., X is large), he is discriminated against otherwise, and, in particular, when he is equally productive as w. In this way, the principal can meet the quota constraint, despite that m works harder than w.

We next argue that the optimal signal structure satisfying (Q) must violate some agent's incentive compatibility constraint. This can be seen from Figure 3, which depicts the signal structures that satisfy both agents' incentive compatibility constraints in the grey area:

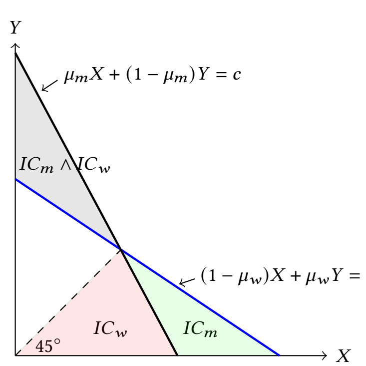


Fig. 3. When $\mu_m + \mu_w > 1$, a signal structure with $X > Y > 0$ cannot be incentive compatible for both agents.

Under the assumption that $\overline{\mu} + \underline{\mu} > 1$, the grey area lies above the 45-degree line. Therefore, it mustn't contain the optimal signal structure satisfying (Q), which lies below the 45-degree line by

[6]For example, Coate and Loury [1993] predict that affirmative action quotas could operate through generating new, "patronizing," equilibria, whereby the minority group works even less harder than before.

Lemma 5.2. In that area, satisfying one agent's incentive compatibility constraint would necessarily violate the incentive compatibility constraint of the other agent.

5.2 Subsidies

Now suppose that the principal receives a subsidy $s \geq 0$ for promoting w (the case $s < 0$ can be solved analogously). Subsidy is an important kind of affirmative action policies. In practice, it may take an explicit form such as employment subsidy, or an implicit form such as promises of more resources, supports, or future hires.

The case $s = 0$ coincides with the baseline model. For any positive level of subsidy $s > 0$, we ask whether its very use helps achieve *equity*, in the following sense.

Definition 5.3. An equilibrium is *equitable* if agents exert the same level of effort and get promoted with equal probability on average.

What distinguishes equity from impartiality is that, while the former only requires that agents be promoted with equal probability on average, the latter mandates that the promotion decision must be independent of agents' identities under all circumstances. An impartial equilibrium must be equitable, but the converse is in general false.

The next theorem provides a negative answer to the aforementioned question, showing that it is difficult, if not impossible, to achieve equity through the use of subsidies.

Theorem 5.4. *Under the assumption that $\overline{\mu}+\underline{\mu} > 1$, there exists no $s > 0$ such that the corresponding game sustains $(\overline{\mu}, \overline{\mu})$ in an equilibrium. If, in addition, $\underline{\mu} > 1/2$, then there exists no $s > 0$ such that the corresponding game has an equitable equilibrium.*

Thus when choosing which level of subsidy to provide to the principal, we face the following dilemma: On the one hand, if we do not subsidize the principal for hiring w, then we cannot rule out the discriminatory equilibrium, especially when it is most preferred by the principal. But once we start to subsidize the principal for hiring w, we will lose equity, sometimes completely. Such an dilemma sets subsidy apart from quota, as the latter always achieves equity by Theorem 5.1.

A sizable economic literature dating back to Weitzman [1974] examines the differences between price versus quantities regulations. In our model, quota and subsidy operate through related, but distinct, channels. On the one hand, they both change the principal's expected payoff to:

$$\mathbb{E}\left[\tilde{a}(\Delta\tilde{\theta} - \nu) \mid \boldsymbol{\mu}, \pi, a(\cdot)\right] + \mu_w - \lambda I(\pi \mid \boldsymbol{\mu}) + \text{other terms, for some } \nu \geq 0.$$

On the other hand, ν differs, depending on which policy is being used. In the case of quota, $\nu \geq 0$ is the Lagrange multiplier associated with the constraint (Q); it equals zero if (Q) holds automatically in the baseline model, and it is strictly positive otherwise. Such a flexibility explains why quota could eliminate the discriminatory equilibrium without impacting on any impartial equilibrium.

In the case of subsidy, however, $\nu = s \geq 0$, and it is set exogenously and rigidly by the authority. The next lemma establishes the counterpart of Lemma 5.2 for any positive level of subsidy.

Lemma 5.5. *Fix any effort profile $\boldsymbol{\mu}$ and any positive level $s > 0$ of subsidy. If the optimal signal structure for $\boldsymbol{\mu}$, given the subsidy, satisfies (Q), then it must also satisfy $\pi(0) < 1/2$ and $X > Y$.*

Combining Lemma 5.5 with the argument articulated in the previous section shows why equity cannot be achieved by a positive level of subsidy when $\underline{\mu} > 1/2$. In that case, satisfying both agents' incentive compatibility constraints at either $(\overline{\mu}, \overline{\mu})$ or $(\underline{\mu}, \underline{\mu})$ requires that we stay in the grey area depicted in Figure 3. Since the grey area lies above the 45 degree line, it mustn't contain the optimal signal structure that promotes agents with equal probability, which is shown to lie below the 45 degree line by Lemma 5.5. As a result, equity can be achieved through the use of quota but not any subsidy.

6 MULTIPLE TASKS AND AND OCCUPATIONAL DISCRIMINATION

This section examines a variant of the baseline model featuring multiple tasks that require distinct skills to fulfill. Agents may undertake multidimensional investments that affect their productivity in each task, and they are screened and selected by the principal to perform the various tasks. Our main finding is that rational inattention may give rise to occupational segregation and stereotypes, whereby the principal favors different agents in the screening and assignment of different tasks, and agents invest in the skills that they are treated favorably for. This occurs in spite of agents having a priori symmetrical aptitudes towards the differing tasks, and can indeed constitute the most profitable equilibrium to the principal.

Setup. There are two tasks that need to be performed: $t = 1, 2$, each arriving randomly with probability $\alpha^t \in (0, 1/2]$. The two tasks never arrive simultaneously, hence it is always the case that exactly one of the tasks needs to be performed.

Agents can undertake multidimensional, costly, investments to improve their task-specific skills. Agent i's investment in skill t is $\mu_i^t \in \{\underline{\mu}, \overline{\mu}\}$. Investment yields a high skill, $\theta_i^t = 1$, with probability μ_i^t, and a low skill, $\theta_i^t = 0$, with the complementary probability $1 - \mu_i^t$. Investing incurs a cost $C^t(\mu_i^t)$ to the agent, where $C^t(\underline{\mu}) = 0$ and $C^t(\overline{\mu}) = C^t > 0$. If agent i is chosen to perform task t, then he earns a reward $\beta^t > 0$. He delivers a benefit θ_i^t to the principal, who values the skill of the agent that is assigned to perform the task.

The principal does not directly observe θ_i^ts, but can acquire costly information about them. The signal that he uses to screen agents for task t is $\pi^t : \{-1, 0, 1\} \to [0, 1]$. For each level of the differential productivity $\Delta\theta^t = \theta_m^t - \theta_w^t$ between m and w, the signal specifies the probability $\pi^t(\Delta\theta^t)$ that m is assigned to perform task t.

The game begins with all players moving simultaneously: The principal specifies the signal structures π^t, $t = 1, 2$; agents decide whether to invest in each skill. After players have made their choices, the task that needs to be performed arrives, and agents are screened according to the pre-specified signal structure. If t is the relevant task, then $\pi^t(\Delta\theta^t)$ is the probability that m is assigned to perform the task. We examine the pure strategy Bayes Nash equilibria of this game.

Preliminaries. First, it is useful to develop some notational conventions. For each $t \in \{1, 2\}$, define $c^t = C^t/(\alpha^t\beta^t\Delta\mu)$, and assume w.l.o.g. that $c^1 \leq c^2$. Intuitively, c^t captures the effective cost that agents must incur in order to win the assignment of task t. $c^1 \leq c^2$ implies that skill 1 is more valuable than skill 2.

In the baseline model, we defined three cutpoints in the attention cost parameter: λ^*, $\overline{\lambda}$, and $\underline{\lambda}$. As we increase c—the effective cost of exerting high effort—these cutpoints must decrease, because more information (and, hence, a reduced information acquisition cost) is needed to motivate agents to work hard. In what follows, we shall write the cutpoints as $\lambda^*(c)$, $\overline{\lambda}(c)$, and $\underline{\lambda}(c)$ in order to signify their dependence on c. The assumption $c^1 \leq c^2$ then implies that the cutpoints are higher for task 1 than for task 2.

Next is our notation of specialization.[7]

[7]To keep the exposition simple, we omit, from the main text, hybrid equilibria in which agents adopt the same investment strategy for one task but different investment strategies for the other task. However, nothing prevents us from conceptualizing these equilibria and comparing them with specialized and non-specialized equilibria. The proof presented in the appendix covers all equilibria.

Definition 6.1. Call an equilibrium *non-specialized* if both agents adopt the same investment strategy. Call an equilibrium *specialized* if one agent invests in skill 1 and the other agent invests in skill 2.

One may think of a non-specialized equilibrium as the multidimensional analog of an impartial equilibrium. In a non-specialized equilibrium, agents invest in the same skill and are screened indiscriminately by the principal. In a specialized equilibrium, however, agents invest in different skills and are screened differently. In the case where m invests in skill 1 and w in skill 2 (which will be our focus), the principal labels task 1 as "traditionally male" and task 2 as "traditionally female," and screens m and w favorably for their respective tasks. Anticipating the discriminatory behavior on the part of the principal, agents invest in the skills that they are screened favorably for and, in turn, reinforce the use of specialized screening. In equilibrium, occupational segregation and stereotypes emerge, whereby m and w are believed to possess the needed skills for succeeding in different tasks, and they do so indeed despite being identical ex ante.

Results. We present two results that are analogous to Theorems 3.3 and 3.4. The first result establishes the existence and uniqueness of specialized and non-specialized equilibria.

THEOREM 6.2. *Suppose that the regularity conditions stated in Theorem 3.3 hold for each* $t \in \{1, 2\}$, *and hence that* $0 < \underline{\lambda}(c^t) < \lambda^*(c^t) < \overline{\lambda}(c^t)$ *for each* $t \in \{1, 2\}$. *Then the following statements are true:*

(1) *A non-specialized equilibrium always exists and is generically unique. When unique, the equilibrium induces both agents to invest in both skills when* $\lambda < \lambda^*(c^2)$, *no agent to invest in any skill when* $\lambda > \lambda^*(c^1)$, *and both agents to invest in skill 1 but not skill 2 when* $\lambda \in (\lambda^*(c^2), \lambda^*(c^1))$.
(2) *A specialized equilibrium exists if and only if*

$$\frac{c^1}{c^2} \geq \frac{\overline{\mu}(1-\overline{\mu})}{\underline{\mu}(1-\underline{\mu})} \text{ and } \lambda \in [\underline{\lambda}(c^1), \overline{\lambda}(c^2)].$$

Whenever a specialized equilibrium exists, it is unique.

Theorem 6.2 extends Theorem 3.3 to multidimensional tasks and skills. In the non-specialized case, the signal structures used to screen agents become less Blackwell informative as the attention cost parameter increases. When the attention cost parameter is below $\lambda^*(c^2)$, screening is meticulous for both tasks, and agents best respond by investing in both skills. When the attention cost parameter is above $\lambda^*(c^1)$, screening is too sloppy to incentivize high levels of investment. For the in-between case $\lambda \in (\lambda^*(c^2), \lambda^*(c^1))$, screening provides agents with just enough incentives to invest in the most valuable skill, but not enough incentives to invest in the other skill.

The specialized case arises when the attention cost parameter is intermediate. To induce one and only one agent to invest in skill t, $t \in \{1, 2\}$, we need $\lambda \in [\underline{\lambda}(c^t), \overline{\lambda}(c^t)]$. Taking intersections between skills, and simplifying using $\overline{\lambda}(c^2) \leq \overline{\lambda}(c^1)$ and $\underline{\lambda}(c^2) \leq \underline{\lambda}(c^1)$, we obtain $[\underline{\lambda}(c^1), \overline{\lambda}(c^2)]$ as the parameter region that sustains specialization in an equilibrium. To ensure that $\underline{\lambda}(c^1) \leq \overline{\lambda}(c^2)$, the two tasks must be sufficiently similar in terms of their costs and benefits to the agents, i.e., $c^1/c^2 \geq \overline{\mu}(1-\overline{\mu})/\underline{\mu}(1-\underline{\mu})$. If the last condition fails, then both agents prefer to invest in the more valuable skill, hence the force behind specialization will unravel.

The second result concerns which of the specialized and non-specialized equilibria is the most profitable to the principal. The comparison is the most straightforward when the two tasks are equally profitable to the principal, i.e., $\alpha^1 = \alpha^2$.

THEOREM 6.3. *Let everything be as in Theorem 6.2, and suppose that* $\alpha^1 = \alpha^2$. *Then,*

(1) *When the game has a specialized equilibrium and a non-specialized equilibrium in which both agents invest in both skills, i.e.,* $\lambda \in [\underline{\lambda}(c^1), \overline{\lambda}(c^2)] \cap [0, \lambda^*(c^2)]$, *the non-specialized equilibrium is the most profitable.*
(2) *When the game has a specialized equilibrium and a non-specialized equilibrium in which no agent invests in any skill, i.e.,* $\lambda \in [\underline{\lambda}(c^1), \overline{\lambda}(c^2)] \cap (\lambda^*(c^1), +\infty)$, *the specialized equilibrium is the most profitable.*
(3) *When the game has a specialized equilibrium and a non-specialized equilibrium in which both agents invest in skill 1 but not skill 2, i.e.,* $\lambda \in [\underline{\lambda}(c^1), \overline{\lambda}(c^2)] \cap (\lambda^*(c^2), \lambda^*(c^1)]$, *the specialized equilibrium is the most profitable.*

Parts (1) and (2) of Theorem 6.3 are immediate from Theorem 3.4. Part (3) of the theorem is new. To understand the intuition behind it, notice that when the attention cost parameter is intermediate, each agent has just enough incentives to invest in one skill, but no more. Now, who should invest in which skill? In the non-specialized case, both agents invest in the same skill. As a result, the principal has to compare and contrast them carefully every time a task needs to be assigned, which incurs a significant attention cost. In the specialized case, agents are expected to opt into separate career trajectories, one labeled as "traditionally male" and the other labeled as "traditionally female." This is achieved by giving stereotypical performance evaluations that favor m in the assignment of the traditionally male task, and w in the assignment of the traditionally female task. Anticipating this, m and w invest in different skills and specialize in different tasks. In turn, this allows the principal to be rationally inattentive, favoring m unless w is strictly more productive in the assignment of the male task, and doing the opposite for the female task.

Implications. There is ample evidence that men and women work on very different jobs even within narrowly defined firms or industries (see Blau and Kahn, 2017 for a survey). Recent sociological and experimental research stresses the role of gender-stereotypical performance evaluations in sustaining and perpetuating this pattern. For example, after coding and analyzing managers' written reviews of employees at a Fortune 500 tech company, Correll et al. [2020] find that women are evaluated based on their personalities and likeabilities, and they are under-rewarded for traits associated with men such as taking charges and being visionary. In a related lab experiment, Bohnet et al. [2016] find that both genders are overlooked for counter-stereotypical tasks, although the problem can be alleviated if employees are evaluated jointly as a team.

Stereotypical performance evaluation is also cited as a culprit for women's underrepresentation in STEM fields. Among others, Lavy and Sand [2018] compare the scores between school exams graded by teachers and national exams graded blindly by external examiners. On subjects such as math and sciences, a gender gap exists and is positively related to the teacher's bias in favor of boys. Female evaluators are not exempt from stereotypes: In a double-blinded study, Moss-Racusin et al. [2012] find that both male and female faculties give lower ratings to female applicants for a lab manager position, despite that the latter are equally capable as their male counterpart.

Our results throw new light on these empirical findings, by telling a story of endogenous stereotype formation and occupational segregation based on limited attention only. While our model abstracts away from many important, practical, considerations—such as the differing attitudes of men and women towards risks and competition, gender social roles, as well as factors inside and outside families that affect women's supply of labor, demand for flexibility, and cost of investing in human capital (see Niederle and Vesterlund, 2011, Blau and Kahn, 2017, and Bertrand, 2018 for surveys of these topics)—it singles out a new channel through which occupational segregation and stereotypes could arise and perpetuate, and raises the possibility of curtailing these phenomena through modulating the availability of attentional resources.

REFERENCES

Dennis J Aigner and Glen G Cain. 1977. Statistical theories of discrimination in labor markets. *Industrial and Labor Relations Review* 30, 2 (1977), 175–187.

Armen A Alchian and Harold Demsetz. 1972. Production, information costs, and economic organization. *American Economic Review* 62, 5 (1972), 777–795.

Sandro Ambuehl. 2017. An offer you can't refuse? Incentives change what we believe. *CESifo Working Paper* (2017).

Kenneth J Arrow. 1971. *Some models of racial discrimination in the labor market.* Technical Report RM-6253-RC. RAND.

Kenneth J Arrow. 1998. What has economics to say about racial discrimination? *Journal of Economic Perspectives* 12, 2 (1998), 91–100.

George P Baker, Michael C Jensen, and Kevin J Murphy. 1988. Compensation and incentives: Practice vs. theory. *Journal of Finance* 43, 3 (1988), 593–616.

Solon Barocas, Moritz Hardt, and Arvind Narayanan. 2019. *Fairness and Machine Learning: Limitations and Opportunities.* fairmlbook.org. http://www.fairmlbook.org.

Vojtěch Bartoš, Michal Bauer, Julie Chytilová, and Filip Matějka. 2016. Attention discrimination: Theory and field experiments with monitoring information acquisition. *American Economic Review* 106, 6 (2016), 1437–1475.

Marianne Bertrand. 2018. Coase lecture–the glass ceiling. *Economica* 85, 338 (2018), 205–231.

Marianne Bertrand, Dolly Chugh, and Sendhil Mullainathan. 2005. Implicit discrimination. *American Economic Review* 95, 2 (2005), 94–98.

Marianne Bertrand and Sendhil Mullainathan. 2004. Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review* 94, 4 (2004), 991–1013.

Francine D Blau and Lawrence M Kahn. 2017. The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature* 55, 3 (2017), 789–865.

Iris Bohnet, Alexandra Van Geen, and Max Bazerman. 2016. When performance trumps gender bias: Joint vs. separate evaluation. *Management Science* 62, 5 (2016), 1225–1234.

Christopher P Chambers and Federico Echenique. 2021. A characterisation of "Phelpsian" statistical discrimination. *The Economic Journal* 131, 637 (2021), 2018–2032.

Elizabeth N Chapman, Anna Kaatz, and Molly Carnes. 2013. Physicians and implicit bias: How doctors may unwittingly perpetuate health care disparities. *Journal of General Internal Medicine* 28, 11 (2013), 1504–1510.

Shubham Chaudhuri and Rajiv Sethi. 2008. Statistical discrimination with peer effects: Can integration eliminate negative stereotypes? *Review of Economic Studies* 75, 2 (2008), 579–596.

Alexandra Chouldechova and Aaron Roth. 2018. The frontiers of fairness in machine learning. *arXiv preprint arXiv:1810.08810* (2018).

Subhasish M Chowdhury, Patricia Esteve-González, and Anwesha Mukherjee. 2020. Heterogeneity, leveling the playing field, and affirmative action in contests. *Southern Economic Journal* (2020).

Dolly Chugh. 2004. Societal and managerial implications of implicit social cognition: Why milliseconds matter. *Social Justice Research* 17, 2 (2004), 203–222.

Stephen Coate and Glenn C Loury. 1993. Will affirmative-action policies eliminate negative stereotypes? *American Economic Review* 83, 5 (1993), 1220–1240.

Shelley Correll and Caroline Simard. 2016. Research: Vague feedback is holding women back. *Harvard Business Review* 29 (2016).

Shelley J Correll, Katherine R Weisshaar, Alison T Wynn, and JoAnne Delfino Wehner. 2020. Inside the black box of organizational life: The gendered language of performance assessment. *American Sociological Review* 85, 6 (2020), 1022–1050.

Thomas de Haan, Theo Offerman, and Randolph Sloof. 2017. Discrimination in the labour market: The curse of competition between workers. *The Economic Journal* 127, 603 (2017), 1433–1466.

Mark Dean and Nate Leigh Neligh. 2019. Experimental tests of rational inattention. *Working Paper* (2019).

Ahrash Dianat, Federico Echenique, and Leeat Yariv. 2022. Statistical discrimination and affirmative action in the lab. *Games and Economic Behavior* 132 (2022), 41–58.

Jennifer L Doleac. 2021. A Review of Thomas Sowell's Discrimination and Disparities. *Journal of Economic Literature* 59, 2 (2021), 574–89.

Mikhail Drugov and Dmitry Ryvkin. 2017. Biased contests for symmetric players. *Games and Economic Behavior* 103 (2017), 116–144.

Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. 2012. Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*. 214–226.

Jennifer L Eberhardt. 2020. *Biased: Uncovering the Hidden Prejudice that Shapes What We See, Think, and Do.* Penguin.

Matteo Escudé, Paula Onuchic, Ludvig Sinander, and Quitzé Valenzuela-Stookey. 2022. Statistical discrimination and statistical informativeness. *arXiv preprint arXiv:2205.07128* (2022).

Hanming Fang. 2001. Social culture and economic performance. *American Economic Review* 91, 4 (2001), 924–937.

Hanming Fang and Andrea Moro. 2011. Theories of statistical discrimination and affirmative action: A survey. *Handbook of Social Economics* 1 (2011), 133–200.

Mogens Fosgerau, Rajiv Sethi, and Jorgen W Weibull. forthcoming. Equilibrium screening and categorical inequality. *American Economic Journal: Microeconomics* (forthcoming).

Roland Fryer and Matthew O Jackson. 2008. A categorical model of cognition and biased decision-making. *BE Journal of Theoretical Economics* 8, 1 (2008), 1–42.

Qiang Fu and Zenan Wu. 2020. On the optimal design of biased contests. *Theoretical Economics* 15, 4 (2020), 1435–1470.

Dylan Glover, Amanda Pallais, and William Pariente. 2017. Discrimination as a self-fulfilling prophecy: Evidence from French grocery stores. *Quarterly Journal of Economics* 132, 3 (2017), 1219–1260.

Anthony G Greenwald and Mahzarin R Banaji. 1995. Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review* 102, 1 (1995), 4–27.

Anthony G Greenwald and Calvin K Lai. 2020. Implicit social cognition. *Annual Review of Psychology* 71, 1 (2020), 419–445.

Anthony G Greenwald, Debbie E McGhee, and Jordan LK Schwartz. 1998. Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology* 74, 6 (1998), 1464–1480.

Harry Holzer and David Neumark. 2000. Assessing affirmative action. *Journal of Economic literature* 38, 3 (2000), 483–568.

Lin Hu, Anqi Li, and Ilya Segal. forthcoming. The politics of personalized news aggregation. *Journal of Political Economy Microeconomics* (forthcoming).

Bo Huang, Jiacui Li, Tse-Chun Lin, Mingzhu Tai, and Yiyuan Zhou. 2022. Attention constraints and financial inclusion. *Working Paper* (2022).

Lily Jampol and Vivian Zayas. 2021. Gendered white lies: Women are given inflated performance feedback compared with men. *Personality and Social Psychology Bulletin* 47, 1 (2021), 57–69.

Michael Kearns and Aaron Roth. 2019. *The Ethical Algorithm: The Science of Socially Aware Algorithm Design*. Oxford University Press.

Victor Lavy and Edith Sand. 2018. On the origins of gender gaps in human capital: Short-and long-term consequences of teachers' biases. *Journal of Public Economics* 167 (2018), 263–279.

Anqi Li and Ming Yang. 2020. Optimal incentive contract with endogenous monitoring technology. *Theoretical Economics* 15, 3 (2020), 1135–1173.

Lori Mackenzie, JoAnne Wehner, and Shelley J Correll. 2019. Why most performance evaluations are biased, and how to fix them. *Harvard Business Review* (2019).

Bartosz Maćkowiak, Filip Matějka, and Mirko Wiederholt. 2021. Rational inattention: A review. *ECB Working Paper* (2021).

C Neil Macrae and Galen V Bodenhausen. 2000. Social cognition: Thinking categorically. *Annual Review of Psychology* 51 (2000), 93–120.

Filip Matějka and Alisdair McKay. 2015. Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review* 105, 1 (2015), 272–98.

Andrei Matveenko and Sergei Mikhalishchev. 2021. Attentional role of quota implementation. *Journal of Economic Theory* 198 (2021), 105356.

Conrad Miller. 2017. The persistent effect of temporary affirmative action. *American Economic Journal: Applied Economics* 9, 3 (2017), 152–190.

Corinne A Moss-Racusin, John F Dovidio, Victoria L Brescoll, Mark J Graham, and Jo Handelsman. 2012. Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences* 109, 41 (2012), 16474–16479.

Muriel Niederle and Lise Vesterlund. 2011. Gender and competition. *Annual Review of Economics* 3, 1 (2011), 601–630.

Paula Onuchic. 2022. Recent contributions to theories of discrimination. *arXiv preprint arXiv:2205.05994* (2022).

Edmund S Phelps. 1972. The statistical theory of racism and sexism. *American Economic Review* 62, 4 (1972), 659–661.

Canice Prendergast. 1999. The provision of incentives in firms. *Journal of economic literature* 37, 1 (1999), 7–63.

Joseph Price and Justin Wolfers. 2010. Racial discrimination among NBA referees. *Quarterly Journal of Economics* 125, 4 (2010), 1859–1887.

Doron Ravid. 2020. Ultimatum bargaining with rational inattention. *American Economic Review* 110, 9 (2020), 2948–2963.

Christopher A Sims. 2003. Implications of rational inattention. *Journal of Monetary Economics* 50, 3 (2003), 665–690.

David R Upton and C Edward Arrington. 2012. Implicit racial prejudice against African-Americans in balanced scorecard performance evaluations. *Critical Perspectives on Accounting* 23, 4-5 (2012), 281–297.

Natasha Warikoo, Stacey Sinclair, Jessica Fei, and Drew Jacoby-Senghor. 2016. Examining racial bias in education: A new approach. *Educational Researcher* 45, 9 (2016), 508–514.

Martin L Weitzman. 1974. Prices vs. quantities. *Review of Economic Studies* 41, 4 (1974), 477–491.

Alison T Wynn and Shelley J Correll. 2018. Combating gender bias in modern workplaces. In *Handbook of the Sociology of Gender*. Springer, 509–521.

Ming Yang. 2020. Optimality of debt under flexible information acquisition. *Review of Economic Studies* 87, 1 (2020), 487–536.

A MODEL DISCUSSION

This appendix clarifies the roles played by the various assumptions and model ingredients.

Consider first Assumption 3.1, which has two parts. The first part: $\overline{\mu} + \underline{\mu} > 1$, is necessary for a discriminatory equilibrium to exist. If, instead, $\overline{\mu} + \underline{\mu} \leq 1$, then (X, Y) must lie below the 45-degree line in order to satisfy both agents' incentive compatibility constraints (as depicted in Figure 4). Since this area does not intersect with the red ray on which the optimal signal structure lies, no discriminatory equilibrium exists when $\overline{\mu} + \underline{\mu} \leq 1$.

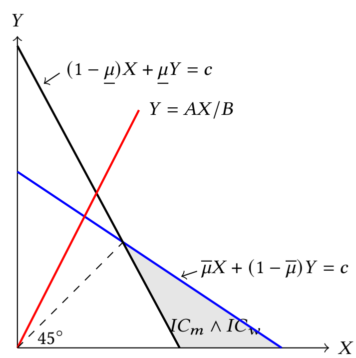


Fig. 4. No discriminatory equilibrium exists when $\overline{\mu} + \underline{\mu} \leq 1$.

The second part of Assumption 3.1: $c < \overline{\mu}(1 - \overline{\mu})/(A + B)$, ensures that $\lambda^*, \underline{\lambda} > 0$. We postpone the proof of this claim to Appendix B, and focus here on its implication, namely in the benchmark case where information acquisition is (almost) costless, i.e., $\lambda \approx 0$, our game has a unique, impartial, equilibrium that sustains the high effort profile. Given this, one can safely attribute all our findings—especially those regarding the discriminatory equilibrium—to rational inattention.

Our story is incomplete without the competition between agents for the limited promotion opportunity. Rational inattention turns this competition into a competition for the principal's limited attention, which in turn justifies the use of a discriminatory signal structure to screen and select agents in the most profitable equilibrium. If, instead, the principal enters a separate contractual relationship with each individual agent (as in, e.g., Coate and Loury, 1993 and Fosgerau et al., ming), then the most profitable equilibrium signal structure between a principal-agent pair is generically unique. This implies that discrimination cannot generically arise as the most profitable equilibrium among ex-ante identical agents—a prediction that stands in sharp contrast to ours.

B PROOFS

Throughout this appendix, we follow the notational conventions developed in the main text. Specifically, we use $\boldsymbol{\mu}$ denote the profile of effort choices by the agents, and $\Delta\theta$ to denote the differential productivity between m and w. For any signal structure π, we use $\overline{\pi}$ to denote the average probability that m is recommended for promotion, and write X and Y for $\pi(1) - \pi(0)$ and $\pi(0) - \pi(-1)$, respectively. Finally, recall the following definitions: $\Delta\mu := \overline{\mu} - \underline{\mu}$, $c := C/\Delta\mu$, $A := \overline{\mu}(1 - \underline{\mu})$, $B := \underline{\mu}(1 - \overline{\mu})$, and $\gamma := \exp(1/\lambda)$. Note that $A > B$, and that γ is decreasing in λ and satisfies $\gamma \longrightarrow +\infty$ as $\lambda \longrightarrow 0$ and $\gamma \longrightarrow 1$ as $\lambda \longrightarrow +\infty$.

B.1 Useful Lemmas and Their Proofs

Proof of Lemma 4.1. Fix any effort profile $\boldsymbol{\mu} \in \{\underline{\mu}, \overline{\mu}\}^2$. By Yang [2020], the optimal signal for $\boldsymbol{\mu}$, denoted simply by π, uniquely exists, and it satisfies $\pi(\Delta\theta) \equiv 1$ if $\mathbb{E}[\exp(-\Delta\theta/\lambda) \mid \boldsymbol{\mu}] \leq 1$, $\pi(\Delta\theta) \equiv 0$ if $\mathbb{E}[\exp(\Delta\theta/\lambda) \mid \boldsymbol{\mu}] \leq 1$, and $\pi(\Delta\theta) \in (0,1)$ $\forall \Delta\theta$ otherwise. Let $p(\Delta\theta)$ denote the probability that $\Delta\theta$ occurs under $\boldsymbol{\mu}$. Simplifying the last condition yields $\forall \Delta\theta \in \{-1, 0, 1\}$:

$$\pi(\Delta\theta) \begin{cases} = 1 & \text{if } p(1)/p(-1) \geq \gamma, \\ = 0 & \text{if } p(1)/p(-1) \leq 1/\gamma, \\ \in (0,1) & \text{else.} \end{cases} \tag{1}$$

In what follows, we say that π is degenerate in the first two case, and that it is nondegenerate in the last case. When nondegenerate, π satisfies the multinomial logit formula prescribed by Matějka and McKay [2015]:

$$\pi(\Delta\theta) = \frac{\overline{\pi} \exp(\gamma \Delta\theta)}{\overline{\pi} \exp(\gamma\Delta\theta) + 1 - \overline{\pi}} \quad \forall \Delta\theta, \tag{2}$$

where $\overline{\pi}$ denotes the average probability that π recommends m for promotion. Bayes plausibility mandates that

$$\sum_{\Delta\theta \in \{-1,0,1\}} p(\Delta\theta)\pi(\Delta\theta) = \overline{\pi}, \tag{3}$$

which, together with (2), pins down π.

Part (1): When $\boldsymbol{\mu} = (\overline{\mu}, \overline{\mu})$, we have $p(1) = p(-1) = \overline{\mu}(1-\overline{\mu})$ and so $p(1)/p(-1) = 1 \in (1/\gamma, \gamma)$. Thus π is always nondegenerate, and it is fully pinned down by (2) and (3). Solving π explicitly yields:

$$\overline{\pi} = \pi(0) = \frac{1}{2} \text{ and } X = Y = g(\gamma) := \frac{\gamma - 1}{2(\gamma + 1)},$$

where $g > 0$ and $g' > 0$ $\forall \gamma > 1$. Since $\gamma := \exp(1/\lambda)$ is decreasing in λ, the last result can be rewritten as $dg(\gamma)/d\lambda < 0$ $\forall \lambda > 0$. The proof for the case $\boldsymbol{\mu} = (\underline{\mu}, \underline{\mu})$ is analogous and hence is omitted for brevity.

Part (2): When $\boldsymbol{\mu} = (\overline{\mu}, \underline{\mu})$, we have $p(1) = A$, $p(-1) = B$, and so $p(1)/p(-1) = A/B > 1$. Thus $p(1)/p(-1) < 1/\gamma$ can never happen, whereas $p(1)/p(-1) \geq \gamma$ holds if and only if

$$\gamma \leq \check{\gamma} := \frac{A}{B}, \text{ or equivalently } \lambda \geq \check{\lambda} := (\ln \check{\gamma})^{-1}.$$

For all $\lambda < \check{\lambda}$, π is nondegenerate and is fully pinned down by (2) and (3). Solving π explicitly for this case yields:

$$\overline{\pi} = \pi(0) = \frac{\gamma A - B}{(\gamma - 1)(A + B)}, \; X = f(\gamma) := \frac{(\gamma A - B)(\gamma B - A)}{(\gamma^2 - 1)(A + B)A}, \text{ and } Y = \frac{A}{B} f(\gamma),$$

where $f > 0$ and $f' > 0$ $\forall \gamma > \check{\gamma}$ (or equivalently $df(\gamma)/d\lambda < 0$ $\forall \lambda < \check{\lambda}$). The proof for the case $\boldsymbol{\mu} = (\underline{\mu}, \overline{\mu})$ is analogous and is thus omitted. □

Proof of Lemma 4.2. For any given μ_w and π, m prefers to exert high effort rather than low effort if and only if

$$\overline{\mu}(1-\mu_w)\pi(1)+(1-\overline{\mu})\mu_w\pi(-1)+[1-\overline{\mu}(1-\mu_w)-(1-\overline{\mu})\mu]\pi(0)-C \\ \geq \underline{\mu}(1-\mu_w)\pi(1)+(1-\underline{\mu})\mu_w\pi(-1)+[1-\underline{\mu}(1-\mu_w)-(1-\underline{\mu})\mu]\pi(0),$$

or equivalently

$$(1-\mu_w)X+\mu_w Y \geq c := \frac{C}{\Delta\mu}.$$

Likewise, w prefers to exert high effort rather than low effort if and only if

$$(1-\overline{\mu})\mu_m\pi(1)+\overline{\mu}(1-\mu_m)\pi(-1)+[1-(1-\overline{\mu})\mu_m-\overline{\mu}(1-\mu_m)]\pi(0)-C \\ \geq (1-\underline{\mu})\mu_m\pi(1)+\underline{\mu}(1-\mu_m)\pi(-1)+[1-(1-\underline{\mu})\mu_m-\underline{\mu}(1-\mu_m)]\pi(0),$$

or equivalently

$$\mu_m X+(1-\mu_m)Y \geq c. \quad \square$$

Proof of Lemma 5.2. Fix the effort profile $\boldsymbol{\mu}$ to be $(\overline{\mu},\underline{\mu})$. Write the principal's problem as:

$$\mathbb{E}\left[\tilde{a}\Delta\tilde{\theta} \mid \boldsymbol{\mu},\pi,a(\cdot)\right]+\mu_w-\lambda I(\pi \mid \boldsymbol{\mu}) \text{ s.t. } \frac{1}{2} \geq \mathbb{E}\left[\tilde{a} \mid \boldsymbol{\mu},\pi,a(\cdot)\right].$$

Since $I(\pi \mid \boldsymbol{\mu})$ is convex in π, and there clearly exists $(\pi,a(\cdot))$ that strictly satisfies the quota constraint, the above problem satisfies Slater's condition; and, hence, strong duality holds. Let $\nu \geq 0$ denote the Lagrange multiplier associated with the quota constraint, and define the Lagrangian function as:

$$\mathcal{L}(\pi,a(\cdot),\nu)=\mathbb{E}\left[\tilde{a}(\Delta\tilde{\theta}-\nu) \mid \boldsymbol{\mu},\pi,a(\cdot)\right]-\lambda I(\pi \mid \boldsymbol{\mu})+\mu_w+\frac{\nu}{2}.$$

The primal and dual problems are:

$$\max_{\pi,a(\cdot)}\min_{\nu\geq 0}\mathcal{L}(\pi,a(\cdot),\nu) \text{ and } \min_{\nu\geq 0}\max_{\pi,a(\cdot)}\mathcal{L}(\pi,a(\cdot),\nu), \text{ respectively.}$$

Let $(\pi^*,a^*(\cdot),\nu^*)$ denote a solution to these problems. A careful inspection of the problem $\max_{\pi,a(\cdot)}\mathcal{L}(\pi,a(\cdot),\nu^*)$ reveals its equivalence to the baseline problem, had the differential productivity between m and w been $\Delta\theta-\nu$. As a result, any solution to this problem, including $(\pi^*,a^*(\cdot))$, must satisfy $\pi:\{-1,0,1\}\to\Delta(\{0,1\})$, $a(0)=0$, and $a(1)=1$. As before, write $\pi^*(\Delta\theta)$ for the probability that π^* recommends m for promotion in state $\Delta\theta$, and $\overline{\pi}^*$ for the average probability that π^* recommends m for promotion. Since the quota constraint must be binding at the optimum, $\nu^*>0$ and $\overline{\pi}^*=1/2$ must hold by complementary slackness. Then

$$\pi^*(\Delta\theta)=\frac{\overline{\pi}^*\exp(\gamma(\Delta\theta-\nu^*))}{\overline{\pi}^*\exp(\gamma(\Delta\theta-\nu^*))+1-\overline{\pi}^*}=\frac{\exp(\gamma(\Delta\theta-\nu^*))}{\exp(\gamma(\Delta\theta-\nu^*))+1} \quad \forall\Delta\theta$$

by Matějka and McKay [2015], and ν^* can be obtained by solving:

$$\frac{1}{2}=\sum_{\Delta\theta\in\{-1,0,1\}}p(\Delta\theta)\pi(\Delta\theta)=A\pi^*(1)+B\pi^*(-1)+(1-A-B)\pi^*(0),$$

where $p(\Delta\theta)$ denotes the probability that $\Delta\theta$ occurs under $\boldsymbol{\mu}$. Write the right-hand side of the last equation as RHS(ν). Straightforward algebra shows that

$$\text{RHS}(0)-\frac{1}{2}=\frac{A-B}{2}\frac{\exp(\gamma)-1}{\exp(\gamma)+1}>0,\ \lim_{\nu\to+\infty}\text{RHS}(\nu)-\frac{1}{2}=-\frac{1}{2}<0, \text{ and } \frac{d\text{RHS}}{d\nu}(\nu)<0,$$

hence RHS(ν) $=1/2$ has a unique, positive, root. This completes the proof that the optimal signal structure for $(\overline{\mu},\underline{\mu})$ subject to (Q) uniquely exists.

It remains to show that π^* satisfies $\pi^*(0) < 1/2$ and $X > Y$. The first claim is easy to prove: $\pi^*(0) = \exp(-\nu^*)/(\exp(-\nu^*) + 1) < 1/2$. Further algebra shows that

$$X = \frac{\exp(\gamma) - 1}{[\exp(\gamma(1 - \nu^*)) + 1][\exp(\gamma\nu^*) + 1]}, \quad Y = \frac{\exp(\gamma\nu^*)(\exp(\gamma) - 1)}{[\exp(\gamma\nu^*) + 1][\exp(\gamma(\nu^* + 1)) + 1]},$$

and hence that

$$\frac{X}{Y} = \frac{\exp(\gamma(\nu^* + 1)) + 1}{\exp(\gamma) + \exp(\gamma\nu^*)} > 1,$$

where the last inequality holds because $\exp(\cdot)$ is convex. □

Proof of Lemma 5.5. For any positive level $s > 0$ of subsidy, we can formalize the principal's problem as follows:

$$\max_{\pi, a(\cdot)} \mathbb{E}\left[\tilde{a}(\Delta\tilde{\theta} - s) \mid \boldsymbol{\mu}, \pi, a(\cdot)\right] + (\mu_w + s) - \lambda I(\pi \mid \boldsymbol{\mu}).$$

By Matějka and McKay [2015], the optimal signal structure must satisfy:

$$\pi(\Delta\theta) = \frac{\overline{\pi}\exp(\gamma(\Delta\theta - s))}{\overline{\pi}\exp(\gamma(\Delta\theta - s)) + 1 - \overline{\pi}} \quad \forall\Delta\theta$$

if it is nondegenerate. Letting $\overline{\pi} = 1/2$ in the above expression and repeating the argument for Lemma 5.2 shows that $X > Y$. □

Lemma B.1. *Let $V(\boldsymbol{\mu};\gamma)$ and $I(\boldsymbol{\mu};\gamma)$ denote the expected revenue and mutual information cost generated by the optimal signal structure for $\boldsymbol{\mu}$, respectively, when the attention cost parameter is $(\ln\gamma)^{-1}$. Then $V(\cdot;\gamma)$ satisfies:*

$$\begin{aligned}
V((\overline{\mu}, \overline{\mu}); \gamma) &= \overline{\mu} + \overline{\mu}(1 - \overline{\mu})\frac{\gamma - 1}{\gamma + 1}, \\
V((\overline{\mu}, \underline{\mu}); \gamma) &= \underline{\mu} + \frac{\gamma A - B}{\gamma + 1}, \\
V((\underline{\mu}, \underline{\mu}); \gamma) &= \underline{\mu} + \underline{\mu}(1 - \underline{\mu})\frac{\gamma - 1}{\gamma + 1}, \\
V((\overline{\mu}, \overline{\mu}); \gamma) - V((\overline{\mu}, \underline{\mu}); \gamma) &= \frac{\Delta\mu}{\gamma + 1}[\gamma - (\gamma - 1)\overline{\mu}], \\
V((\overline{\mu}, \underline{\mu}); \gamma) - V((\underline{\mu}, \underline{\mu}); \gamma) &= \frac{\Delta\mu}{\gamma + 1}[\gamma - (\gamma - 1)\underline{\mu}], \\
\frac{d}{d\gamma}V((\overline{\mu}, \overline{\mu}); \gamma) - V((\overline{\mu}, \underline{\mu}); \gamma) &= \frac{\Delta\mu(1 - 2\overline{\mu})}{(\gamma + 1)^2}, \\
\text{and } \frac{d}{d\gamma}V((\overline{\mu}, \underline{\mu}); \gamma) - V((\underline{\mu}, \underline{\mu}); \gamma) &= \frac{\Delta\mu(1 - 2\underline{\mu})}{(\gamma + 1)^2},
\end{aligned}$$

whereas $I(\cdot;\gamma)$ satisfies:

$$
\begin{aligned}
I((\overline{\mu},\overline{\mu});\gamma) &= 2\overline{\mu}(1-\overline{\mu})[h\left(\frac{\gamma}{\gamma+1}\right) - h\left(\frac{1}{2}\right)],\\
I((\overline{\mu},\underline{\mu});\gamma) &= Ah\left(\frac{\gamma(\gamma A-B)}{(\gamma^2-1)A}\right) + Bh\left(\frac{\gamma A-B}{(\gamma^2-1)B}\right) - (A+B)h\left(\frac{\gamma A-B}{(\gamma-1)(A+B)}\right),\\
I((\underline{\mu},\underline{\mu});\gamma) &= 2\underline{\mu}(1-\underline{\mu})[h\left(\frac{\gamma}{\gamma+1}\right) - h\left(\frac{1}{2}\right)],\\
\frac{d}{d\gamma}I((\overline{\mu},\overline{\mu});\gamma) - I((\overline{\mu},\underline{\mu});\gamma) &= \frac{\Delta\mu(1-2\overline{\mu})\ln\gamma}{(\gamma+1)^2},\\
\text{and } \frac{d}{d\gamma}I((\overline{\mu},\underline{\mu});\gamma) - I((\underline{\mu},\underline{\mu});\gamma) &= \frac{\Delta\mu(1-2\underline{\mu})\ln\gamma}{(\gamma+1)^2},
\end{aligned}
$$

where $h(x) := x\ln x + (1-x)\ln(1-x)\ \forall x \in [0,1]$.

Proof. In the proof of Lemma 4.1, we solved for the optimal signal structure for any given $\boldsymbol{\mu}$. Substituting these solutions into the expressions for $V(\cdot;\gamma)$ and $I(\cdot;\gamma)$ gives the desired result. We omit most algebra, but point out an intermediate result we used when calculating $I(\boldsymbol{\mu};\gamma) - I(\boldsymbol{\mu}';\gamma)$, $\boldsymbol{\mu} \neq \boldsymbol{\mu}'$:

$$
\frac{d}{d\gamma}I((\overline{\mu},\overline{\mu});\gamma) = \frac{2\overline{\mu}(1-\overline{\mu})\ln\gamma}{(\gamma+1)^2},\ \frac{d}{d\gamma}I((\overline{\mu},\underline{\mu});\gamma) = \frac{(A+B)\ln\gamma}{(\gamma+1)^2},
$$
$$
\text{and } \frac{d}{d\gamma}I((\underline{\mu},\underline{\mu});\gamma) = \frac{2\underline{\mu}(1-\underline{\mu})\ln\gamma}{(\gamma+1)^2}.
$$

This result follows from doing lengthy algebra, which is available upon request. □

B.2 Proofs of Theorems and Propositions

Proof of Theorem 3.3. The following observation will be useful for the proof: Under Assumption 3.1: $\overline{\mu}+\underline{\mu} > 1$ and $c < \overline{\mu}(1-\overline{\mu})/(A+B)$, $c < \min\{1/2, \underline{\mu}(1-\underline{\mu})/(A+B)\}$ must hold, because

$$
\frac{\overline{\mu}(1-\overline{\mu})}{A+B} - \frac{1}{2} = \frac{\Delta\mu(1-2\overline{\mu})}{2(A+B)} < 0 \text{ and } \frac{\overline{\mu}(1-\overline{\mu})}{A+B} - \frac{\underline{\mu}(1-\underline{\mu})}{A+B} = \frac{\Delta\mu(1-\overline{\mu}-\underline{\mu})}{A+B} < 0.
$$

Part (1): Combining Lemma 4.1(1) and Lemma 4.2 shows that $(\overline{\mu},\overline{\mu})$ can be sustained in an equilibrium if and only if $g(\gamma) \geq c$. Since $g(1) = 0$, $g' > 0$ on $(1,+\infty)$, and $\lim_{\gamma\to+\infty} g(\gamma) = 1/2 > c$, $g(\gamma) \geq c$ if and only if

$$
\gamma \geq \gamma^* := g^{-1}(c), \text{ or equivalently } \lambda \leq (\ln\gamma^*)^{-1} := \lambda^* > 0.
$$

When this condition fails, we have $g(\gamma) < c$ and so can sustain $(\underline{\mu},\underline{\mu})$ can in an equilibrium. At $\gamma = \gamma^*$ (or $\lambda = \lambda^*$), both $(\overline{\mu},\overline{\mu})$ and $(\underline{\mu},\underline{\mu})$ can be sustained in equilibrium.

Part (2): We can sustain $(\overline{\mu},\underline{\mu})$ in an equilibrium if and only the optimal signal structure for $(\overline{\mu},\underline{\mu})$ satisfies (i) $X = f(\gamma)$, (ii) $Y = AX/B$, and (iii) agents' incentive compatibility constraints, i.e., $(1-\underline{\mu})X + \underline{\mu}Y \geq c$ and $\overline{\mu}X + (1-\overline{\mu})Y \leq c$. Solving (ii) and (iii) simultaneously yields $X \in [\underline{X}, \overline{X}]$, where

$$
\underline{X} = \frac{c(1-\overline{\mu})}{1-\underline{\mu}} \text{ and } \overline{X} = \frac{c\underline{\mu}}{\overline{\mu}}.
$$

Note that $\underline{X}$ and $\overline{X}$ are both independent of γ. Moreover, $\underline{X} < B/(A+B)$ because

$$\underline{X} < B/(A+B) \Longleftrightarrow c < \underline{\mu}(1-\underline{\mu})/(A+B) \Longleftarrow \text{Assumption 3.1},$$

and $\overline{X} < B/(A+B)$ because

$$\overline{X} = \frac{c\underline{\mu}}{\overline{\mu}} < \frac{\overline{\mu}(1-\overline{\mu})}{A+B}\frac{\underline{\mu}}{\overline{\mu}} = \frac{B}{A+B}.$$

Together with $f' > 0\ \forall \gamma \in (\check{\gamma}, +\infty)$, $f(\check{\gamma}) = 0$, and $\lim_{\gamma\to+\infty} f(\gamma) = B/(A+B)$, these observations imply that (i) holds if and only if $\gamma \in [\underline{\gamma}, \overline{\gamma}]$, where

$$\underline{\gamma} := f^{-1}(\underline{X}) \text{ and } \overline{\gamma} := f^{-1}(\overline{X})$$

are both finite. Define

$$\underline{\lambda} := (\ln \overline{\gamma})^{-1} \text{ and } \overline{\lambda} := (\ln \underline{\gamma})^{-1},$$

and note that $0 < \underline{\lambda} < \overline{\lambda} < \check{\lambda} < +\infty$.

It remains to show that $\underline{\lambda} < \lambda^*$ (equivalently $\gamma^* < \overline{\gamma}$) always holds, and that $\lambda^* < \overline{\lambda}$ (equivalently $\underline{\gamma} < \gamma^*$) holds under additional conditions. To show that $\gamma^* < \overline{\gamma}$, rewrite $f(\gamma) = \overline{X}$ as

$$\varphi(\gamma) := \frac{(\gamma A - B)(\gamma B - A)}{(\gamma^2-1)\underline{\mu}(1-\underline{\mu})(A+B)} = c,$$

where $\varphi : [\check{\gamma}, +\infty) \longrightarrow \mathbb{R}$ satisfies $\varphi(\check{\gamma}) = 0$ and $\varphi' > 0\ \forall \gamma > \check{\gamma}$. Then $\overline{\gamma}$ is the unique root of $\varphi(\gamma) = c$, and γ^* is the unique root of $g(\gamma) = c$, where $g : [1, +\infty) \longrightarrow \mathbb{R}$ satisfies $g(1) = 0$ and $g' > 0\ \forall \gamma > 1$. Tedious algebra shows that

$$\frac{d}{d\gamma}\frac{\varphi(\gamma)}{g(\gamma)} = \frac{2(A-B)^2(\gamma+1)}{\underline{\mu}(1-\underline{\mu})(A+B)(\gamma-1)^3} > 0$$

and that

$$\lim_{\gamma\to+\infty} \varphi(\gamma) = \frac{\overline{\mu}(1-\overline{\mu})}{A+B} < \frac{1}{2} = \lim_{\gamma\to+\infty} g(\gamma).$$

Therefore, $\varphi(\gamma) < g(\gamma)\ \forall \gamma \in [\check{\gamma}, +\infty)$, and so $\gamma^* < \overline{\gamma}$.

To pin down the conditions for $\underline{\gamma} < \gamma^*$ to hold, rewrite $f(\gamma) = \underline{X}$ as

$$\psi(\gamma) := \frac{\underline{\mu}(1-\underline{\mu})}{\overline{\mu}(1-\overline{\mu})}\varphi(\gamma) = c,$$

and $\underline{\gamma}$ as the unique root of $\psi(\gamma) = c$. From the above derivation, we know that

$$\frac{d}{d\gamma}\frac{\psi(\gamma)}{g(\gamma)} > 0$$

and that

$$\lim_{\gamma\to+\infty} \psi(\gamma) - \lim_{\gamma\to+\infty} g(\gamma) = \frac{\underline{\mu}(1-\underline{\mu})}{A+B} - \frac{1}{2} = \frac{\Delta\mu(2\underline{\mu}-1)}{2(A+B)}.$$

Thus $\gamma^* > \underline{\gamma}$ if and only if $\underline{\mu} > 1/2$ and

$$c > g(\hat{\gamma}), \tag{4}$$

where $\hat{\gamma}$ is the unique root of $g(\gamma) = \psi(\gamma)$. Numerical analysis shows that (4) can hold simultaneously with Assumption 3.1. □

Proof of Theorem 3.4. We proceed in three steps.

Step 1. By Lemma B.1, the following must hold for all $\gamma > 1$:

$$V((\overline{\mu},\overline{\mu});\gamma) - V((\underline{\mu},\underline{\mu});\gamma) = \frac{\Delta\mu}{\gamma+1}[2\gamma - (\gamma-1)(\overline{\mu}+\underline{\mu})] > 0,$$

$$\text{and } I((\overline{\mu},\overline{\mu});\gamma) - I((\underline{\mu},\underline{\mu});\gamma) = -2\Delta\mu(\overline{\mu}+\underline{\mu}-1)[h\left(\frac{\gamma}{\gamma+1}\right) - h\left(\frac{1}{2}\right)] < 0,$$

where the last inequality follows from the assumption that $\overline{\mu}+\underline{\mu} > 1$ and the fact that $\mathrm{argmin}_{[0,1]} h = 1/2$. Thus at $\gamma = \gamma^*$, the impartial equilibrium sustaining $(\overline{\mu},\overline{\mu})$ is more profitable than the impartial equilibrium sustaining $(\underline{\mu},\underline{\mu})$. The remaining analysis divides $[\underline{\gamma},\overline{\gamma}]$ into two disjoint intervals $[\underline{\gamma},\gamma^*)$ and $[\gamma^*,\overline{\gamma}]$. The most profitable impartial equilibrium sustains $(\underline{\mu},\underline{\mu})$ on the first interval, and $(\overline{\mu},\overline{\mu})$ on the second interval.

Step 2. Show that the discriminatory equilibrium is the most profitable equilibrium on $[\underline{\gamma},\gamma^*)$. Write $\Delta V(\gamma)$ for $V((\overline{\mu},\underline{\mu});\gamma) - V((\underline{\mu},\underline{\mu});\gamma)$, $\Delta I(\gamma)$ for $I((\overline{\mu},\underline{\mu});\gamma) - I((\underline{\mu},\underline{\mu});\gamma)$, and $\Delta R(\gamma)$ for $\Delta V(\gamma) - \Delta I(\gamma)/\ln\gamma$. We wish to show that $\Delta V(\gamma) - \Delta I(\gamma)/\ln\gamma > 0$. First, recall from Lemma B.1 that $\forall\gamma \in [\check{\gamma},+\infty)$:

$$\Delta V(\gamma) > 0 \text{ and } \frac{d}{d\gamma}\Delta I(\gamma) = \frac{\Delta\mu(1-2\underline{\mu})\ln\gamma}{(\gamma+1)^2}.$$

Thus when $\underline{\mu} > 1/2$ (as required by the theorem), $\Delta I(\gamma)$ is decreasing in γ on $[\check{\gamma},+\infty)$. Then from

$$\Delta I(\check{\gamma}) = 0 - 2\underline{\mu}(1-\underline{\mu})[h\left(\frac{\check{\gamma}}{\check{\gamma}+1}\right) - h\left(\frac{1}{2}\right)] < 0, \qquad (\because \check{\gamma} > 1 \text{ and } \mathrm{argmin}_{[0,1]} h = 1/2)$$

it follows that $\Delta I(\gamma) < 0\ \forall\gamma \geq \check{\gamma}$, and hence that $\Delta R(\gamma) > 0\ \forall\gamma \geq \check{\gamma}$ as desired.

Step 3. Show that the discriminatory equilibrium is the least profitable equilibrium on $[\gamma^*,\overline{\gamma}]$. Write $\Delta V(\gamma)$ for $V((\overline{\mu},\overline{\mu});\gamma) - V((\overline{\mu},\underline{\mu});\gamma)$, $\Delta I(\gamma)$ for $I((\overline{\mu},\overline{\mu});\gamma) - I((\overline{\mu},\underline{\mu});\gamma)$, and $\Delta R(\gamma)$ for $\Delta V(\lambda) - \Delta I(\gamma)/\ln\gamma$. Since

$$\Delta I(\check{\gamma}) = 2\overline{\mu}(1-\overline{\mu})[h\left(\frac{\check{\gamma}}{\check{\gamma}+1}\right) - h\left(\frac{1}{2}\right)] - 0 > 0 \qquad (\because \check{\gamma} > 1 \text{ and } \mathrm{argmin}_{[0,1]} h = 1/2)$$

and

$$\frac{d}{d\gamma}\Delta I(\gamma) = \frac{\Delta\mu(1-2\overline{\mu})\ln\gamma}{(\gamma+1)^2} < 0, \qquad (\because \overline{\mu} > \tfrac{1}{2})$$

either $\Delta I(\gamma) > 0\ \forall\gamma \in [\check{\gamma},+\infty)$, or it single-crosses the horizontal line from above at some $\tilde{\gamma} > \check{\gamma}$. Then from

$$\begin{aligned}\frac{d}{d\gamma}\Delta R(\gamma) &= \frac{d}{d\gamma}[\Delta V(\gamma) - \frac{1}{\ln\gamma}\Delta I(\gamma)]\\ &= \frac{d\Delta V(\gamma)}{d\gamma} - \frac{1}{\ln\gamma}\frac{d\Delta I(\gamma)}{d\gamma} + \frac{\Delta I(\gamma)}{\gamma(\ln\gamma)^2}\\ &= \frac{\Delta\mu(1-2\overline{\mu})}{(\gamma+1)^2} - \frac{1}{\ln\gamma}\frac{\Delta\mu(1-2\overline{\mu})\ln\gamma}{(\gamma+1)^2} + \frac{\Delta I(\gamma)}{\gamma(\ln\gamma)^2} \qquad (\because \text{Lemma B.1})\\ &= \frac{\Delta I(\gamma)}{\gamma(\ln\gamma)^2},\end{aligned}$$

it follows that $\Delta R(\gamma)$ is either monotonically increasing on $[\check{\gamma},+\infty)$, or it first increases on $[\check{\gamma},\tilde{\gamma}]$ and then decreases on $(\tilde{\gamma},+\infty)$. In both situations, we have

$$\lim_{\gamma\to+\infty}\Delta R(\gamma) = \lim_{\gamma\to+\infty}\Delta V(\gamma) - 0\cdot\lim_{\gamma\to+\infty}\Delta I(\gamma) = \Delta\mu(1-\overline{\mu}) - 0 > 0.$$

Thus if $\Delta R(\check{\gamma}) > 0$, then $\Delta R(\gamma) > 0 \; \forall \gamma \in [\check{\gamma}, +\infty)$ as desired.

To show that $\Delta R(\check{\gamma}) > 0$, note that $V((\overline{\mu}, \underline{\mu}); \check{\gamma}) = \underline{\mu}$ by Lemma B.1, and that $I(\overline{\mu}, \underline{\mu}); \check{\gamma}) = 0$ by Lemma 4.1. Also note that $V((\overline{\mu}, \overline{\mu}); \check{\gamma}) - \frac{1}{\ln \check{\gamma}} I((\overline{\mu}, \overline{\mu}); \check{\gamma}) \geq \overline{\mu}$, where $\overline{\mu}$ is the expected profit generated by $(\overline{\mu}, \overline{\mu})$ if the principal uses a degenerate signal structure that recommends m for promotion for sure, and the inequality follows from optimality, i.e., the optimal signal structure for $(\overline{\mu}, \overline{\mu})$ generates a (weakly) higher expected profit to the principal than the degenerate signal structure. Taken together, we conclude that $\Delta R(\check{\gamma}) > \Delta \mu > 0$ as conjectured. □

Proof of Theorems 5.1 and 5.4. Combining Lemmas 5.2 and 5.5 with the graphical arguments offered in Section 5 gives the desired results. □

Proof of Theorem 6.2. For starters, notice that for each task $t \in \{1, 2\}$ and effort profile $\boldsymbol{\mu}^t = (\mu_m^t, \mu_w^t)$, the principal's problem is the same as before. Thus, what is left is to verify that the joint signal structure (π^1, π^2) satisfies agents' incentive compatibility constraints. Compared to the baseline model, agents can now commit two-step deviations that revise their effort choices for both tasks, in addition to one-step deviations that revise their effort choices for a single task. However, since their problems are additive separable across tasks, it suffices to deter one-step deviations only. Given this, we can solve the multidimensional problem as two separate single-dimensional problems—an approach we will follow in the remainder of the proof.

Part (1): The optimal signal structure for $((\overline{\mu}, \overline{\mu}), (\overline{\mu}, \overline{\mu}))$ is incentive compatible if and only if $\lambda \leq \min\{\lambda^*(c^1), \lambda^*(c^2)\}$. Since $c^1 \leq c^2$ and $\lambda^*(\cdot)$ is decreasing in its argument, the last condition is equivalent to $\lambda \leq \lambda^*(c^2)$. Likewise, the optimal signal structure for $((\underline{\mu}, \underline{\mu}), (\underline{\mu}, \underline{\mu}))$ is incentive compatible if and only if $\lambda \geq \max\{\lambda^*(c^1), \lambda^*(c^2)\} = \lambda^*(c^1)$, and the optimal signal structure for $((\overline{\mu}, \overline{\mu}), (\underline{\mu}, \underline{\mu}))$ is incentive compatible if and only if $\lambda \in [\lambda^*(c^2), \lambda^*(c^1)]$. The optimal signal structure for $((\underline{\mu}, \underline{\mu}), (\overline{\mu}, \overline{\mu}))$ isn't incentive compatible unless $c^1 = c^2$.

Part (2): The optimal signal structure for $((\overline{\mu}, \underline{\mu}), (\underline{\mu}, \overline{\mu}))$ is incentive compatible if and only if $\lambda \in \cap_{t=1}^2 [\underline{\lambda}(c^t), \overline{\lambda}(c^t)]$. Since $\underline{\lambda}(\cdot)$ and $\overline{\lambda}(\cdot)$ are decreasing in their arguments, the last expression is a nonempty set if and only if $\underline{\lambda}(c^1) \leq \overline{\lambda}(c^2)$. To reduce this condition to primitives, let $(X, AX/B)$ be the optimal signal structure for $(\overline{\mu}, \underline{\mu})$ when the attention cost parameter is given by λ. In the proof of Theorem 3.3, we established that $\lambda \geq \underline{\lambda}(c^1)$ if and only if

$$X \leq \overline{X}(c^1) = \frac{c^1 \underline{\mu}}{\overline{\mu}},$$

and that $\lambda \geq \overline{\lambda}(c^2)$ if and only if

$$X \geq \underline{X}(c^2) = \frac{c^2(1 - \overline{\mu})}{1 - \underline{\mu}}.$$

Thus $\underline{\lambda}(c^1) \leq \overline{\lambda}(c^2)$ if and only if $\overline{X}(c^1) \geq \underline{X}(c^2)$, which, after simplifying, becomes

$$\frac{c^1}{c^2} \geq \frac{\overline{\mu}(1 - \overline{\mu})}{\underline{\mu}(1 - \underline{\mu})}.$$

Before we conclude, notice that the method developed above also speaks to situations in which agents undertake the same level of investment in one task, but different levels of investment in the other task. For example, the optimal signal structure for $((\overline{\mu}, \overline{\mu}), (\overline{\mu}, \underline{\mu}))$ is incentive compatible if and only if $\lambda \leq \lambda^*(c^1)$ and $\lambda \in [\underline{\lambda}(c^2), \overline{\lambda}(c^2)]$. To save space, we choose not to exhaust all possibilities, but instead focus on the specialized and non-specialized equilibria only. □

Proof of Theorem 6.3. Parts (1) and (2) are immediate from Theorem 3.4. To prove Part (3), let $V(\boldsymbol{\mu};\gamma)$ and $I(\boldsymbol{\mu};\gamma)$ denote the expected revenue and mutual information cost generated by the optimal signal structure for $\boldsymbol{\mu}$, respectively, when the attention cost parameter is $(\ln\gamma)^{-1}$. Write $\Delta V^1(\gamma)$ for $V((\overline{\mu},\overline{\mu});\gamma) - V((\overline{\mu},\underline{\mu});\gamma)$, $\Delta V^2(\gamma)$ for $V((\overline{\mu},\underline{\mu});\gamma) - V((\underline{\mu},\underline{\mu});\gamma)$, $\Delta I^1(\gamma)$ for $I((\overline{\mu},\overline{\mu});\gamma) - I((\overline{\mu},\underline{\mu});\gamma)$, and $\Delta I^2(\gamma)$ for $I((\overline{\mu},\underline{\mu});\gamma) - I((\underline{\mu},\underline{\mu});\gamma)$. We wish to show that

$$\Delta V^1(\gamma) - \frac{1}{\ln\gamma}\Delta I^1(\gamma) - [\Delta V^2(\gamma) - \frac{1}{\ln\gamma}\Delta I^2(\gamma)] < 0 \quad \forall\gamma \in [\check{\gamma}, +\infty).$$

In what follows, we prove a stronger claim, namely $\Delta V^1(\gamma) < \Delta V^2(\gamma)$ and $\Delta I^1(\gamma) > \Delta V^2(\gamma)$ $\forall\gamma \in [\check{\gamma}, +\infty)$.

To show that $\Delta V^1(\gamma) < \Delta V^2(\gamma)$, recall from Lemma B.1 that

$$\Delta V^1(\gamma) = \frac{\Delta\mu}{\gamma+1}[\gamma - (\gamma-1)\overline{\mu}] \text{ and } \Delta V^2(\gamma) = \frac{\Delta\mu}{\gamma+1}[\gamma - (\gamma-1)\underline{\mu}].$$

Thus,

$$\Delta V^1(\gamma) - \Delta V^2(\gamma) = -\frac{(\gamma-1)(\Delta\mu)^2}{\gamma+1} < 0$$

as desired.

It remains to show that $\Delta I^1(\gamma) > \Delta I^2(\gamma)$ $\forall\gamma \in [\check{\gamma}, +\infty)$. The claim is clearly true at $\gamma = \check{\gamma}$, since $\Delta I^1(\check{\gamma}) > 0$ and $\Delta I^2(\check{\gamma}) < 0$. It is also true when γ is very large, since

$$\begin{aligned}
&\lim_{\gamma\to+\infty} \Delta I^1(\gamma) - \Delta I^2(\gamma) \\
&= 2[\overline{\mu}(1-\overline{\mu}) + \underline{\mu}(1-\underline{\mu})]\ln 2 - 2[A\ln\left(\frac{A+B}{A}\right) + B\ln\left(\frac{A+B}{B}\right)] \\
&> 0. \qquad \text{(Verify using Mathematica)}
\end{aligned}$$

Then from

$$\begin{aligned}
&\frac{d}{d\gamma}\Delta I^1(\gamma) - \Delta I^2(\gamma) \\
&= \frac{\Delta\mu(1-2\overline{\mu})\ln\gamma}{(\gamma+1)^2} - \frac{\Delta\mu(1-2\overline{\mu})\ln\gamma}{(\gamma+1)^2} \qquad (\because \text{Lemma B.1}) \\
&= -\frac{2(\Delta\mu)^2\ln\gamma}{(\gamma+1)^2} \\
&< 0,
\end{aligned}$$

it follows that $\Delta I^1(\gamma) - \Delta I^2(\gamma)$ is everywhere positive on $[\check{\gamma}, +\infty)$ as desired. □