Subtle Discrimination

Finance Working Paper N° 903/2023
August 2024

Daniel Ferreira
London School of Economics, CEPR and ECGI

Elena S. Pikulina
University of British Columbia

ECGI Working Paper Series in Finance

Subtle Discrimination

Working Paper N° 903/2023
August 2024

Daniel Ferreira
Elena S. Pikulina

We thank our reviewers, an associate editor, Thomas Philippon (editor), Murray Carlson, Hans Degryse, Alex Edmans, Jack Favilukis, Guido Friebel, Paolo Fulghieri, Lorenzo Garlappi, Will Gornall, Sebastian Gryglewicz, Isaac Hacamo, Alex Imas, Elena Pastorino, Luca Picariello, Alesio Piccolo, Maddalena Ronchi, and Edward Van Wesep for valuable feedba?

Abstract

We introduce the concept of subtle discrimination - biased acts that cannot be objectively ascertained as discriminatory - and study its implications in a model of promotions. We show that subtle (as opposed to overt) discrimination has unique implications. Discriminated candidates perform better in low-stakes careers, while favored candidates perform better in high-stakes careers. In equilibrium, firms are polarized: high-productivity firms become "progressive" and have diverse management teams, while low-productivity firms choose to be "conservative" and homogeneous at the top. Subtle discrimination also has unique empirical predictions in contexts such as equity analysis, lending, fund flows, banking careers, and entrepreneurial finance.

Keywords: Bias, human capital, promotions, firm performance

JEL Classification: M51, J71, J31

Daniel Ferreira

London School of Economics, CEPR and ECGI
Professor of Finance
London School of Economics
e-mail: d.ferreira@lse.ac.uk

Elena S. Pikulina

University of British Columbia
Assistant Professor
e-mail: elena.pikulina@sauder.ubc.ca

Subtle Discrimination*

Elena S. Pikulina Daniel Ferreira

August 2024
Forthcoming in the Journal of Finance

Abstract

We introduce the concept of *subtle discrimination*—biased acts that cannot be objectively ascertained as discriminatory—and study its implications in a model of promotions. We show that subtle (as opposed to overt) discrimination has unique implications. Discriminated candidates perform better in low-stakes careers, while favored candidates perform better in high-stakes careers. In equilibrium, firms are polarized: high-productivity firms become "progressive" and have diverse management teams, while low-productivity firms choose to be "conservative" and homogeneous at the top. Subtle discrimination also has unique empirical predictions in contexts such as equity analysis, lending, fund flows, banking careers, and entrepreneurial finance.

Keywords: Bias, human capital, promotions, firm performance

JEL Classification: M51, J71, J31

*Pikulina: elena.pikulina@sauder.ubc.ca, Finance Division, Sauder School of Business, University of British Columbia. Ferreira: d.ferreira@lse.ac.uk, Department of Finance, London School of Economics and Political Science, CEPR and ECGI. We thank our reviewers, an associate editor, Thomas Philippon (editor), Murray Carlson, Hans Degryse, Alex Edmans, Jack Favilukis, Guido Friebel, Paolo Fulghieri, Lorenzo Garlappi, Will Gornall, Sebastian Gryglewicz, Isaac Hacamo, Alex Imas, Elena Pastorino, Luca Picariello, Alesio Piccolo, Maddalena Ronchi, and Edward Van Wesep for valuable feedback and, on occasion, lively discussions. We also thank seminar and conference participants at Schulich School of Business (York University), Sauder Business School (University of British Columbia), London School of Economics and Political Science, New Economic School, the University of Naples, the University of Hong Kong, Erasmus University Rotterdam, Bayes Business School, Goethe University Frankfurt, Georgetown University, Amsterdam Business School, 2022 Annual Pacific Northwest Finance Conference, 2022 North-American Economic Science Association Conference, 2023 American Finance Association Meetings, 2023 FMA Napa/Sonoma Conference, ESCP ESG Worskshop, 2023 Corporate Finance NBER Meetings, 2023 Adam Smith Conference, CSEF Conference on Finance, Labor and Inequality, 2023 Personnel Economics NBER SI Meetings, and the Vienna Festival of Finance Theory for valuable comments.

1 Introduction

Economists traditionally classify discriminatory acts based on their causes. Some view such acts as a consequence of rational statistical discrimination (Phelps (1972); Arrow (1973)). A second view is that biases in tastes, beliefs, or incentives can cause discrimination (Becker (1957); Bordalo et al. (2016); Bohren et al. (2023); Dobbie et al. (2021)). While such a classification is undoubtedly useful, other perspectives are also possible. Social and organizational psychologists classify discriminatory acts based on their transparency, i.e., whether discrimination is overt or subtle. These scholars define *subtle discrimination* as acts that are ambiguous in intent to harm, ex-post rationalizable, difficult to verify, and often (but not always) unintentional.[1] In the workplace, examples include asking female employees to perform menial tasks, failing to praise the performance of minority employees, and disproportionately promoting men to managerial positions when choosing among equally qualified candidates.

It is hard to substantiate claims of subtle discrimination. In the United States, to prosecute a person or company for discrimination, the discriminated party must either provide direct evidence of intent to harm or deny rights or prove a clear pattern of adverse events that can be explained only by discrimination. Partially due to the threat of legal action, overt discrimination has become relatively rare. In contrast, subtle discrimination often hides behind non-discriminatory narratives, making it a common and insidious phenomenon.[2]

Despite its prevalence, the impact of subtle discrimination on workers and firms has received scant attention in the economics and finance literatures. Our paper is a first attempt at formalizing the concept of subtle discrimination. We define subtle discrimination as biased acts that cannot be objectively ascertained as discriminatory. The defining feature of subtle discrimination is *plausible*

[1]See, for example, Dovidio and Gaertner (1986), Essed (1991), Dipboye and Halverson (2004), Van Laer and Janssens (2011), Jones et al. (2017), Dhanani et al. (2018), and Hebl et al. (2020). While some of these studies often describe this type of discrimination using different terms, such as "modern," "aversive," "every day," "ambivalent" and "covert," they all contrast subtle discrimination with "old-time" overt discrimination and emphasize its ambiguous, hard-to-detect and yet pernicious nature.

[2]Several law scholars argue that the burden of proof is particularly high in discrimination cases. They demonstrate that discrimination cases proceed and terminate less favorably for plaintiffs than other kinds of civil cases, including *habeas corpus* cases (Selmi (2000); Clermont and Schwab (2009)).

deniability. For instance, when a biased decision-maker acts in a discriminatory way, he or she may offer an explanation for such actions, such as using subjective criteria like "potential" to choose between two candidates with similar objective qualifications, thereby concealing a bias against one of the candidates. When such a narrative is plausible, the decision-maker is *subtly biased*. For example, a manager who asks a female subordinate to perform a menial task could argue that he sometimes asks men to perform such duties and that the frequency of such requests is low and unrelated to gender. Likewise, if performance is subjectively assessed, a manager accused of not praising someone's performance could claim that they did not perform adequately.

We apply our notion of subtle discrimination to a model of promotions. In the model, two ex-ante identical agents with labels "blue" and "red" compete for promotion by investing in human capital. Labels are payoff-irrelevant; profit depends only on the acquired skill of the promoted agent. The decision-maker (the *principal*) has a slight bias toward the blue agent. Because the bias is small, the principal prefers to promote the red agent when her objective qualifications are considerably better than those of the blue agent. That is, the principal does not *overtly discriminate*. However, when both candidates are similarly qualified, the principal is likelier to promote the blue candidate. This form of discrimination is subtle because the principal's "tie-breaking" rule is not observable, implying that any promotion decision can be construed as unbiased.[3]

Our analysis draws from the reality that choosing among candidates with similar objective qualifications is often challenging. In such cases, the principal is likely to rely on subjective assessments. Such discretion allows biases to influence choices. In an employment setting, Hoffman et al. (2018) show that biases, not superior information, explain most cases in which managers use discretion to select candidates. Using internal data from a large bank, Bircan et al. (2023) found that supervisors' discretion in job assignments explains most of the gender promotion gap, implying the existence of "*subtle mechanisms that disadvantage women*."

Our paper makes three primary contributions. First, we propose a taxonomy of discriminatory

[3] Our definition does not imply that the principal always breaks ties in favor of the blue candidate. Dipboye and Halverson (2004) and Gaertner and Dovidio (2005) emphasize that subtle biases tend to be variable. Sometimes, individuals behave in discriminatory ways, and at other times, they demonstrate their egalitarian views.

acts with two categories: subtle and overt. This distinction is novel to the theoretical literature on discrimination in economics, which we review in Section 2. Second, in a model of promotions, we show that subtle and overt discrimination have different empirical predictions. Third, our model of subtle discrimination in promotions generates a rich set of empirical predictions relating firm characteristics to the performance of different groups of workers, the diversity of top management teams, and firms' choices of anti-discrimination policies.

We distinguish subtle discrimination from overt discrimination based on the ease of objectively ascertaining particular acts as discriminatory. When two candidates have similar qualifications, passing over one candidate for a promotion is not clear evidence of discrimination. In our model, when blue and red agents are similarly skilled, the principal can use his private signals to rationalize the act of promoting the blue agent, both to others and to himself. In contrast, promoting an unskilled blue agent over a skilled red agent is clear evidence of discrimination; we classify such acts as overt discrimination. While our particular formalization necessarily leaves out many nuances in real life, it is simple, intuitive, and tractable.

Although individual acts of subtle discrimination cannot be undeniably exposed, agents may form beliefs about the principal's subtle bias in several ways. First, agents may estimate the principal's bias in promotion decisions by observing his behavior in other settings, such as preferential treatment of workers from some groups in day-to-day interactions. Second, the principal may voluntarily take some actions that signal his bias. Third, agents may extrapolate from the past promotion decisions of a set of similar firms and assign an "average" bias to all firms in the set. In all such cases, agents' beliefs can be (approximately) correct even when individual acts of subtle discrimination are unverifiable.

Beliefs about subtle biases affect the agents' decisions to invest in human capital. Because promotions are competitive, subtle biases also affect how the agents react to each other's decisions. In equilibrium, agents differ in their investment choices, which creates an *achievement gap*, i.e., a difference in accumulated human capital and qualifications. Two opposing forces contribute to the achievement gap. On the one hand, unfavored agents are discouraged from investing in human capital because they anticipate a low probability of promotion. We call this force the *discouragement effect*. On the other hand, an unfavored agent may choose to overinvest in skills in an attempt to separate herself from the favored agent. We call this force the *overcompensation effect*.

We show that the sign and magnitude of the achievement gap depend on the stakes faced by the agents. When the net benefit from promotion is large – a *high-stakes career path* – the discouragement effect dominates, implying that favored (blue) agents invest <u>more</u> than unfavored (red) agents. In this case, the achievement gap is positive: favored agents have more visible achievements (e.g., better qualifications and performance records) than unfavored agents. In contrast, when the net benefit from promotion is small – a *low-stakes career path* – the overcompensation effect dominates and, thus, favored agents invest <u>less</u> than unfavored agents, leading to a negative achievement gap. We show that these results hinge crucially on discrimination being subtle instead of overt.

These results are helpful when interpreting the evidence on the professional advancement of women and minorities. Evidence that women have lower promotion rates in high-skilled occupations can be found in Hospido et al. (2019) for central bankers, Bosquet et al. (2019) for academic economists, Azmat et al. (2020) for lawyers, and Bircan et al. (2023), Huang et al. (2023), and Ceccarelli et al. (2024) for bankers.

Azmat et al. (2020) show that female associates in law firms invest less in the qualifications required for promotion (e.g., hours billed) than male associates. Hospido et al. (2019) and Bosquet et al. (2019) find that women are less likely to seek promotion in the first place. Similarly, Linos et al. (2023) find that black-white promotion gaps in a professional services firm can be explained partly by black employees receiving worse subjective evaluations when assigned to teams with more white coworkers. In contrast, Benson et al. (2024) find that women in management-track careers in retail have better (pre-promotion) performance than men. These facts are consistent with our prediction that discriminated groups are discouraged from investing in promotable tasks in high-stakes careers while being over-incentivized to undertake such investments in low-stakes careers.

Our model also predicts that, in high-stakes careers, differences in observable achievements (such as human capital, performance, experience, and effort) explain most of the *promotion gap* (i.e., the difference in promotion rates between groups). Because the promotion gap increases with the expected benefits of promotion, the model can also explain the evidence of growing promotion gaps at the top of hierarchies, a fact that is known as the "leaky pipe" phenomenon (Lundberg and Stearns (2019); Sherman and Tookes (2022)). Such evidence is at odds with Lazear and Rosen's (1990) prediction that promotion gaps should decrease as the skill required for a task increases.

In the model, firms can change their subtle biases through anti-discrimination (i.e., diversity, equity, and inclusion) policies. Firms can become more progressive (i.e., less biased) or conservative (i.e., more biased) at no direct cost. In equilibrium, firms become polarized. On one side, we have high-productivity firms offering high-stakes careers to their employees. Such firms choose to become progressive and, thus, have greater diversity in their top management teams. On the other side, we have low-productivity firms that offer low-stakes careers. Such firms choose to be conservative and, thus, have little diversity at the top. The model predicts that even small differences in firm productivity can account for large differences in corporate culture and top-level diversity. Furthermore, market forces cannot eliminate such differences. Thus, our model provides novel empirical predictions relating firm quality to observed diversity metrics. Consistent with the model's predictions, Edmans et al. (2023) find that employees' perception of diversity, equity, and inclusion is stronger in growing, high-valuation, and financially strong firms. Firm polarization in diversity preferences is also a potential explanation for the evidence that large and well-performing firms have more women on their boards (Adams and Ferreira (2009)).

Our model offers a novel perspective on the costs and benefits of corporate focus on social issues, especially regarding workforce diversity. While some argue that progressive values typically do not conflict with the pursuit of profits (e.g., Edmans (2020)), others claim that some businesses excessively focus on promoting progressive causes to the detriment of profits (see Edgecliffe-Johnson (2022)). Our model illustrates one mechanism through which progressive firms can increase profits: Employees who believe that a company does not discriminate in promotions are encouraged to invest in their human capital. Moreover, the model shows that such benefits accrue primarily to firms with high returns to human capital. In contrast, firms in which investment in human capital has low returns have less to gain from eliminating discrimination in promotions. For such firms, investing in a reputation for progressiveness does not pay off.

Finally, we show how the concept of subtle discrimination can help interpret evidence of discrimination in other contexts. Because equity and credit analysts are rewarded for both accuracy and optimism (see Hong and Kubik (2003) and Kumar (2010)), in the presence of subtle biases against women analysts, we expect women to make more accurate forecasts, which is in line with the evidence (Kumar (2010)). Subtle discrimination can also explain why minority loan applicants have lower approval rates (Butler et al. (2023); Frame et al. (2023)). It also helps understand the evidence linking mutual fund flows to fund managers' personal characteristics, such as name origins (Kumar et al. (2015)) and gender (Niessen-Ruenzi and Ruenzi (2019)).

In the next section, we describe the related theoretical literature. In Section 3, we define subtle discrimination in the context of a choice between two candidates for a position. We present our main analysis in Section 4. Section 5 briefly considers other applications. We discuss further empirical implications in Section 6 and conclude in Section 7. The Online Appendix (OA) presents welfare and policy analyses, as well as several variations of our main model.

2 Related Theoretical Literature

Our article contributes to the theoretical literature on discrimination (see Fang and Moro (2011), Lang and Lehmann (2012), and Onuchic (2023) for reviews). Our main contribution to this literature is the formalization of the concepts of "subtle discrimination" and "overt discrimination," and a demonstration of the relevance of this classification: subtle discrimination and overt discrimination typically have different empirical consequences.

In their seminal work on affirmative action, Coate and Loury (1993) show that negative stereotypes can be self-fulfilling because discriminated agents may underinvest in human capital. Similarly, in our model, discrimination may discourage some agents from investing. However, because workers compete for the same position, their investment decisions are interdependent. Such strategic considerations may further discourage investment or, instead, provide discriminated agents with stronger incentives to invest. Thus, differently from Coate and Loury (1993), in our model, the unfavored group may invest more than the favored group. This result is obtained only when discrimination is subtle. Building upon Coate and Loury (1993), Fryer Jr. (2007) presents a statistical discrimination model with two stages: hiring and promotion. Employer beliefs may "flip" between stages: in the second stage, the employer expects a disadvantaged worker to be more qualified than a favored worker. Our model has a similar flavor: promoted red agents are more qualified on average. However, because qualifications are observable, subtle biases do not flip.

Our model has peer effects: agents impose externalities on one another, as in Mailath et al. (2000), Moro and Norman (2004) and Chaudhuri and Sethi (2008). In these models, asymmetric equilibria exist in which agents with identical ex-ante qualifications receive different wages. In contrast, in our model, wages do not vary across groups, and discrimination cannot be verified ex-post. Unlike theories of discrimination based on differential screening abilities (Cornell and Welch (1996); Fershtman and Pavan (2021)), our model assumes that the principal knows each candidate's type. While we can still interpret subtle discrimination as a form of incorrect or exaggerated beliefs, as in Bordalo et al. (2016), it can also be seen as a limiting case of taste-based discrimination when the taste parameter is arbitrarily small.

Our paper also relates to a strand of the discrimination literature focusing on bias amplification. Lang et al. (2005) show that weak discriminatory preferences can cause large wage differentials in markets where firms post wages. Bartoš et al. (2016) show how "attention discrimination" can amplify animus and prior beliefs about group quality. Davies et al. (2021) demonstrate that an arbitrary small bias towards one candidate can have large consequences when the principal exerts effort to learn about candidates' abilities. Siniscalchi and Veronesi (2021) present a model in which mild population heterogeneity and self-image bias can lead to persistent group differences. Unlike these models, our model's source of bias amplification is the competitive nature of promotion tournaments. Similar to our model, Onuchic and Ray (2023) study a setting where the strategic responses of individuals can amplify small biases. In our model, subtle biases can be either amplified or attenuated depending on the stakes.

In the promotion literature, Lazear and Rosen (1990) provide the first theoretical analysis of promotion gaps. In their model, men have higher promotion rates than equally qualified women because women have better non-work opportunities and are thus more likely to quit after being promoted. In terms of empirical predictions, their model implies small promotion gaps in high-stakes situations (i.e., when workers are more productive upon promotion), while our model predicts the opposite: for a given bias, promotion gaps are larger in high-level jobs. We also show that this result reverses if firms can choose their biases at no cost.

Our setup is similar to that of Prendergast (1993), who proposes a model of promotions in which the firm cannot contractually commit to compensating workers for acquiring firm-specific human capital. Our model differs in two significant aspects: i) promotions are competitive, i.e., candidates compete for a limited number of positions; ii) the principal is subtly biased in favor of candidates from a particular group. More generally, our model relates to the literature on favoritism and other biases in subjective performance evaluations and their consequences for selection and promotion decisions (Prendergast and Topel (1996); MacLeod (2003); Friebel and Raith (2004); Hoffman et al. (2018); Frederiksen et al. (2020); Letina et al. (2020); Frankel (2021); Pagano and Picariello (2022)). In these models, favoritism and other biases have ex-post payoff consequences for the decision-maker. In contrast, in our model, favoritism matters only because it affects ex-ante incentives; it has no ex-post (i.e., after investment) consequences.

Additionally, our paper is related to a small theoretical literature on biased contests (Kawamura and de Barreda (2014); Pérez-Castrillo and Wettstein (2016)). Drugov and Ryvkin (2017) show that under certain conditions, biased contests can be optimal from the organizer's point of view (e.g., total effort maximization) even when contestants are symmetric. In that vein, Nava and Prummer (2022) present a model in which the principal can directly affect the contestants' valuations of the prize (promotion) through work culture. Our paper differs from this literature in many respects, particularly in our focus on subtle discrimination, its empirical implications, and its consequences for different types of firms. Finally, our notion of subtle bias is related to Cunningham and de Quidt's (2022) concept of implicit preferences, defined as preferences toward a trait (like gender or race) that have a stronger effect when mixed with other characteristics. They focus on developing a methodology for identifying such preferences from choices. Our main focus is on how subtle biases (which encompass both their implicit and explicit categories) impact human capital accumulation and firm outcomes.

3 Definitions and Interpretation

A decision-maker needs to choose one of two candidates $i \in \{b, r\}$, called Blue and Red. Blue and Red have objective qualifications – called *skills* – summarized by s_b and s_r, which everyone observes. Without loss of generality, we set $\Delta s \equiv s_r - s_b \geq 0$. The decision-maker privately observes a (subjective) signal x_i for each agent. The signals are independent and identically distributed random variables with support $[\underline{x}, \overline{x}]$. Let $F(.)$ denote the cumulative distribution function of $\Delta x \equiv x_r - x_b$. The decision-maker has a mandate to maximize an observable payoff (e.g., profit), π. If the decision-maker chooses agent i, the payoff is $\pi_i = s_i + \omega x_i + u$, where $\omega > 0$ and u is a zero-mean random variable. We assume that everyone knows ω and holds the same beliefs about $F(.)$ and the distribution of u.

In the absence of biases, after observing s_i and x_i, the decision-maker chooses Blue if $\Delta s + \omega \Delta x < 0$. Thus, the probability of an unbiased decision-maker choosing Blue is $F\left(-\frac{\Delta s}{\omega}\right) = F\left(\frac{s_b - s_r}{\omega}\right)$. In the case of a biased decision-maker, to keep the analysis general, we model the decision-maker's behavior directly without specifying payoffs and beliefs; we discuss different microfoundations below. Let $P(s_b, s_r, \omega)$ denote the probability that the decision-maker chooses Blue given s_b, s_r and ω. We assume that $P(s_b, s_r, \omega)$ is increasing in s_b and decreasing in s_r. While $P(s_b, s_r, \omega)$ cannot be directly observed, the candidates and other observers may form beliefs about it. We assume that such beliefs are correct to avoid introducing additional behavioral considerations.

We define the decision-maker's bias towards Blue as the excess probability of choosing Blue not justified by the qualification gap, $s_b - s_r$, and the importance of subjective signals, ω:

$$b(s_b, s_r, \omega) = P(s_b, s_r, \omega) - F\left(\frac{s_b - s_r}{\omega}\right) \geq 0. \tag{1}$$

There could be several sources for the bias in (1). For example, the decision-maker may prefer blue agents (as in Becker (1957)). Alternatively, the decision-maker may incorrectly believe that blue agents are more productive (Bordalo et al. (2016); Bohren et al. (2023)). Even with correct beliefs, the decision-maker may be better at reading subjective signals from blue agents (e.g., Cornell and Welch (1996); Fershtman and Pavan (2021)). Biases may also be caused by external factors, such as poorly designed incentive structures (Dobbie et al. (2021)). In what follows, we do not take a stand on whether biases are caused by beliefs or preferences, or whether they are intrinsic or extrinsic. Our model can accommodate most of these interpretations.

Next, we define subtle bias:

Definition. *If $F\left(\frac{s_b - s_r}{\omega}\right) > 0$, we say that bias $b(s_b, s_r, \omega)$ is subtle.*

By assumption, Red's objective qualifications are (weakly) better than Blue's qualifications $(\Delta s \geq 0)$. Nevertheless, the decision-maker may still choose Blue, either because the decision-maker is biased towards Blue or because the decision-maker privately observes signal $\Delta x \leq \frac{s_b - s_r}{\omega}$. That is, whenever there exist potential signal realizations that justify choosing Blue, i.e., $F\left(\frac{s_b - s_r}{\omega}\right) > 0$, a biased decision-maker can "plausibly deny" being biased if only a small number of decisions are observed. *Plausible deniability* is the defining property of subtle discrimination. In our formalization, plausible deniability means that if everyone observes Δs and ω and has the same belief about $F(.)$, an act of choosing Blue does not constitute conclusive evidence that the decision-maker is biased. *Subtle discrimination* is thus an act of favored treatment of an agent or group of agents when the decision-maker can resort to a plausible-deniability defense of such an act.

In contrast with subtle biases, we now define *overt bias*:

Definition. *If* $F\left(\frac{s_b - s_r}{\omega}\right) = 0$*, we say that bias* $b(s_b, s_r, \omega)$ *is overt.*

Case $F\left(\frac{s_b - s_r}{\omega}\right) = 0$ occurs when $\Delta s > \omega(\bar{x} - \underline{x})$. In this case, a single act of choosing Blue is incontrovertible evidence of biased decision-making. While a binary classification of biases into subtle or overt is helpful, biases can also vary in "subtlety." In particular, without the bounded support assumption, all biases would be subtle. Still, a subtle bias with very small $F\left(\frac{s_b - s_r}{\omega}\right)$ is, for all practical purposes, an overt bias because one would find it difficult to defend choosing Blue. We can thus think of $F\left(\frac{s_b - s_r}{\omega}\right)$ as a measure of subtlety in discrimination. Bias subtlety is maximized at $s_b = s_r$ (recall that $s_b \leq s_r$ by assumption). Thus, subtle discrimination is most likely to occur when objective differences are small. When agents are observationally equivalent, any choice is rationalizable, even when the importance of subjective information is minimal (i.e., when $\omega \to 0$).

By definition, an individual act of subtle discrimination is not immediately detectable. While outcome realizations (i.e., π) might be informative about underlying biases, they may not be sufficient to reveal subtle biases because of the noise in performance (u). In practice, proving the existence of subtle biases is challenging even after observing a long sequence of decisions. First, for small differences in observables Δs, one would need a large sample to detect subtle discrimination with reasonable confidence. Second, for small Δs, even mild disagreements about ω and $F(.)$ can provide sufficient cover for subtle biases. Thus, subtle discrimination is prevalent in situations where (i) differences in objective qualifications between groups are small, (ii) decisions (like promotions) by a single decision-maker are infrequent, and (iii) observers disagree about the importance of subjective information for forecasting performance.

Our notion of subtle discrimination is also compatible with unconscious biases in decision-making. This interpretation is valid when the decision-maker does not directly benefit from choosing a particular type. One interpretation is that the decision-maker always rationalizes his choice (as in Cherepanov et al. (2013)). In practice, the decision-maker might find it difficult to correct the bias (at least in a finite series of decisions) if he believes his choices are unbiased. Such unconscious biases likely pertain to System 1 thinking, i.e., fast, automatic, and effortless associations (Kahneman (2011)).

While establishing the presence of subtle biases with confidence might be hard in practice, agents and observers may hold (correct or incorrect) beliefs about such biases. Thus, subtle biases may affect agents' decisions to invest in observable skills (s_i). In our application below, we show that subtle biases can be either attenuated or amplified in competitive environments, resulting in substantial differences in outcomes between groups.

4 A Model of Subtle Discrimination in Promotions

In this section, we present an application of our concept of subtle discrimination to a model of promotions. For simplicity, we consider the limiting case in which the principal's subjective information is negligible for the decision: $\omega \to 0$. In that case, subtle discrimination can occur only when the difference in skills, $s_r - s_b = \Delta s$, is zero. We present the setup in Subsection 4.1 and introduce subtle biases in Subsection 4.2. In Subsection 4.3, we describe the first-best solution to serve as a benchmark. We then solve the model for an exogenously given compensation contract in Subsection 4.4. In Subsection 4.5, we let firms choose the compensation contracts optimally. In Subsection 4.6, we endogenize the subtle bias. In the OA, we discuss the robustness of the model to several different assumptions.

4.1 Definitions and Model Setup

At Date 0, a firm hires two ex-ante identical agents – b (*Blue*) and r (*Red*) – for an entry-level position (*job 1*). Red and Blue are payoff-irrelevant labels. The firm needs to fill both vacancies. We assume that the firm does not (or cannot) discriminate at the hiring stage; thus, the 50/50 split between b and r reflects the composition of the candidate pool. That is, we implicitly assume that some frictions prevent firms from picking only one type. This is similar to Athey et al. (2000), who also assume a diverse entry-level workforce in a model of promotions.

At Date 1, the agents simultaneously undertake a non-observable investment (or effort), $e_i \in [0,1]$, $i \in \{b,r\}$, to acquire a "skill." The skill is an observable but unverifiable binary variable:

$s_i \in \{0,1\}$. We interpret skill broadly as any observable evidence that predicts an agent's future performance. We assume that the skill is firm-specific in the sense that it is less valuable to agents who leave the firm; see Sections OA.2 and OA.3 for model extensions in which agents invest in (partially or fully) general skills. Agent i's probability of acquiring the skill is e_i. Both agents are risk-neutral and have the same cost function, $c(e_i)$, which is strictly increasing, strictly convex, and satisfies $c(0) = 0$. That is, agent i's utility is $u_i = w_i - c(e_i)$, where w_i is the agent's monetary compensation.

At Date 2, the principal can choose one of the agents to fill a top position (*job 2*).[4] Agents who are not promoted remain in entry-level jobs; we normalize the revenue generated by these to zero. Agents can be skilled ($s_i = 1$) or unskilled ($s_i = 0$). Promoting an unskilled agent increases the principal's expected payoff by $l > 0$ while promoting a skilled agent increases the payoff by $l + H$, where $H > 0$ denotes the *productivity gain* upon promotion of a skilled agent. That is, a skilled agent is always more productive than an unskilled one when assigned to job 2.[5] We interpret H as a firm characteristic. Larger H means that human capital is more critical at higher hierarchical levels.

Although the principal cannot offer wages contingent on skill acquisition, the principal can commit to a set of non-negative wages (w_1, w_2) for the holders of jobs 1 and 2, respectively. We call $W \equiv w_2 - w_1$ the *promotion premium*. We describe the compensation contract by a vector $\mathbf{w} = (w_1, W)$ representing a basic reward and a promotion premium. We are interested in the case in which contractual discrimination in promotion decisions is not possible. That is, the principal must offer the same contract, $\mathbf{w}$, to both agents. Because $H > 0$, when choosing between two candidates, an unbiased principal always promotes a skilled agent over an unskilled one. As in Prendergast (1993), the principal can effectively commit to rewarding skill acquisition through promotions. In addition, if $l > W$, it is always in the principal's interest to promote one of the

[4] Although we interpret this decision as a promotion, it can also be interpreted as a direct hiring decision for a specific post. See Coate and Loury (1993) for a model that allows for both interpretations. For a related model of worker competition for job slots, see Lazear et al. (2018).

[5] For a theory and recent evidence on the importance of human capital acquisition for career progression inside firms, see Pastorino (2022).

agents, even when both agents are unskilled. As l is a free parameter in the model, we assume $l > W$ for simplicity. Section OA.2 presents a variation of the model in which firms sometimes leave job 2 vacant.

4.2 Subtle Bias and Agents' Beliefs

As in Section 3, we model the principal's behavior by the probability of choosing Blue given ω, s_b and s_r. For simplicity, we consider the limiting case in which the principal's subjective information is negligible for the decision: $\omega \to 0$. Section OA.4 presents a variation of the model where ω is non-negligible. There are three cases for the difference in skills: $s_r - s_b = \Delta s \in \{-1, 0, 1\}$. We assume there is no overt discrimination: if $\Delta s \neq 0$, the principal chooses the skilled agent (Blue if $\Delta s = -1$ and Red if $\Delta s = 1$). However, if $\Delta s = 0$, the principal chooses Blue with probability $0.5 + \beta$, where $\beta \in (0, 0.5]$ is the principal's subtle bias. Using the notation in Section 3, $\beta = \lim_{\omega \to 0} \lim_{\Delta s \to 0} b(s_b, s_b + \Delta s, \omega)$, for any $s_b \in \{0, 1\}$. Thus, given the principal's behavior, the firm's (expected) profit is $\Pi = l + H(e_b + e_r - e_b e_r) - 2w_1 - W$.

In the context of this model, the principal's behavior can be described by a simple lexicographic heuristic: if $s_b \neq s_r$, choose the skilled agent; if $s_b = s_r$, choose Blue with probability $0.5 + \beta$. While we assume that ties are exact, this is not a necessary condition for subtle discrimination (see Section 3).[6] It is also unnecessary for the main results; we only need observable skills to be sufficiently similar so that subtle discrimination is possible. Assuming exact ties is algebraically convenient for solving the firm's maximization problem as described in Subsections 4.5 and 4.6, but otherwise, it is not needed, as we show in Section OA.4.

There are multiple ways to micro-found β. We think of β as the agents' beliefs about the principal's behavior in the case of a tie. Such beliefs may be correct or incorrect. While we will assume that these beliefs are correct for simplicity, most of the results in this paper hold true even when agents' beliefs are incorrect.

[6] Our decision-making heuristic can be mapped into Tversky's (1969) notion of lexicographic semiorder; see also Manzini and Mariotti (2012) for a generalization. In such models, ties can arise with positive probability even when the assessment criteria are continuous.

How would agents acquire correct beliefs without a long series of past promotion decisions by the same firm? Workers observe the behavior of managers and supervisors, the company's culture, and several related decisions on a daily basis. Under the assumption that some of these workplace experiences are correlated with biases in promotion settings, workers may use these observations to form their beliefs about β. Azmat et al. (2020) show evidence of this channel by documenting that negative workplace experiences, such as harassment and demeaning comments, are correlated with women's promotion aspirations in law firms.

A firm may also take actions that signal its bias. As we will show in Section 4.6, if firms can choose their own biases, it is in their interest to communicate their biases truthfully.

Finally, it is also possible to estimate the average subtle bias for a set of firms. Suppose that a fraction 2β of all firms has a slight preference for blue agents. Agents do not know the type of firm they match with. Thus, in the case of a tie, a blue agent expects to be promoted with probability $2\beta + (1-2\beta)\frac{1}{2} = \frac{1}{2} + \beta$. Under this interpretation, an agent can estimate β by observing promotion decisions across several firms. In a large sample of such firms, agents' beliefs can be quite precise. However, no worker can be sure that their particular firm is biased.

4.3 Benchmark: First-best Investment Levels

As a benchmark, we consider the problem of an unbiased social planner who maximizes total surplus. Recall that the principal's private subjective information is payoff-irrelevant (i.e., $\omega \to 0$); thus it does not affect the planner's problem. Define the (expected) social surplus as $S = \Pi + E[u_b] + E[u_r]$. The planner's problem is

$$\max_{(e_b,e_r)\in[0,1]^2} l + H\left(e_b + e_r - e_b e_r\right) - c(e_b) - c(e_r). \tag{2}$$

A trade-off exists between effort duplication and effort sharing. If both agents invest, some skills might be wasted—a duplication cost. Yet, if only one agent invests, her marginal cost of effort is higher than that of the idle agent. Similar to risk sharing under concave utilities, effort sharing (i.e., marginal cost equalization across agents) is efficient under convex costs. The first-best choice thus depends on which of these two effects dominates. The following proposition formalizes this intuition (all proposition proofs are in Appendix A):

Proposition 1. *The first-best investment levels are either (i) $e_b^{FB} = e_r^{FB} = \tilde{e} < 1$ or (ii) $e_i^{FB} > 0$ and $e_{-i}^{FB} = 0$ for some $i \in \{b, r\}$.*

Proposition 1 says that the first-best effort levels can be symmetric or asymmetric. If the benefits from effort sharing are greater than the costs of effort duplication, the social planner chooses the same investment level for both agents (Case (i)). If the duplication costs outweigh the benefits of effort sharing, the social planner asks only one agent to invest in skill acquisition (Case (ii)). As an example, consider $c(e_i) = \frac{ke_i^2}{2}$. If $H \leq k$, treating both agents equally is socially optimal, and the first-best solution is $e_b^{FB} = e_r^{FB} = \tilde{e} = \frac{H}{H+k}$. If $H > k$, the first-best is a corner solution, $e_b^{FB} = 1$ and $e_r^{FB} = 0$ (or $e_b^{FB} = 0$ and $e_r^{FB} = 1$). That is, it is better to treat agents asymmetrically and make only one of them invest in skill acquisition.

For the rest of the paper, we assume a quadratic cost function, $c(e_i) = \frac{ke_i^2}{2}$. This assumption allows us to obtain analytical proofs for most results and better explain the economic forces at play. However, it is not crucial for the main results (see Section OA.5).

4.4 Equilibrium under Exogenous Compensation Contracts

Here, we describe the agents' investment choices under a fixed contract **w**. We assume the contract is individually rational; both Blue and Red accept the contract at Date 0. At Date 1, the agents simultaneously choose their investment levels. At Date 2, investment outcomes are realized, and the principal decides who to promote to the top position. Both agents anticipate that at Date 2, the principal's decision is biased in favor of Blue.

4.4.1 Equilibrium Characterization

We define agent i's expected utility as:

$$U_i(\mathbf{e},\mathbf{w}) \equiv w_1 + W\left[e_i(1-e_{-i}) + \left(\frac{1}{2}+\beta_i\right)(1-e_i-e_{-i}+2e_ie_{-i})\right] - \frac{ke_i^2}{2}, \tag{3}$$

where $\beta_b = -\beta_r = \beta$. The first term is the baseline reward, the second is the promotion premium times the probability of promotion, and the third is the skill acquisition cost. The promotion probability is the sum of the probability of agent i being skilled when agent $-i$ is unskilled and the probability of promotion via a tie-breaking decision.

An agent's problem at Date 1 is to maximize his/her expected utility $U_i(\mathbf{e},\mathbf{w})$ by choosing an investment level e_i taking the contract, $\mathbf{w}$, and the effort of the other agent, e_{-i}, as given. Assuming an interior solution,[7] the agents' reaction functions are

$$e_b = \frac{W}{k}\left(\frac{1}{2}-\beta+2\beta e_r\right) \text{ and } e_r = \frac{W}{k}\left(\frac{1}{2}+\beta-2\beta e_b\right). \tag{4}$$

We define $\sigma \equiv \frac{W}{k}$, i.e., the ratio of the promotion premium to the cost parameter, and call it the *premium-cost ratio*. Higher σ implies a higher net marginal benefit of investment. Intuitively, high σ implies that the gain from promotion, W, is large relative to the cost of investment, which is proportional to k. High σ can thus be interpreted as a "high-stakes" career path, i.e., there is much to gain from investing in skill acquisition. In contrast, if σ is low, agents benefit little from investing; the agents are on a low-stakes career path. Thus, we also informally refer to σ as the "stake" of a career path.

In the baseline case with no subtle bias ($\beta = 0$), the reaction functions in (4) are flat, implying that $e_b^* = e_r^* = \frac{\sigma}{2}$ is the dominant strategy. To understand the intuition for this result, suppose agent i increases her investment by ε. If the other agent is skilled, agent i's promotion probability increases by $\frac{\varepsilon}{2}$ because she can be promoted only when she is skilled, in which case the winner is chosen with equal probabilities. If the other agent is unskilled, agent i increases her promotion

[7] In Subsection 4.5, we show that optimal contracts always imply interior solutions.

probability by $\varepsilon - \frac{\varepsilon}{2} = \frac{\varepsilon}{2}$, because if she is skilled, she is promoted with certainty and if she is unskilled, both agents are equally likely to be promoted. In sum, agent i's promotion probability always increases by half of her investment increase, regardless of whether the other agent is skilled or unskilled.

The introduction of a bias in favor of Blue breaks this symmetry. Now, if Blue increases his effort by ε, his promotion probability increases by $\varepsilon(\frac{1}{2}+\beta)$ in the state where Red is skilled, while in the state where Red is unskilled, his promotion probability increases by less: $\varepsilon(\frac{1}{2}-\beta)$. Thus, the state matters for the investment decision: Blue's marginal benefit of investing is larger when Red is more likely to be skilled. That is, Blue's reaction function is positively sloped. For similar reasons, Red's reaction function is negatively sloped. Intuitively, ties are more valuable to Blue than they are to Red. Thus, Blue wants to imitate Red's behavior, which causes Blue's reaction function to slope upwards. In contrast, Red adopts the opposite strategy in an attempt to avoid ties. The following proposition characterizes the equilibrium investment choices.

Proposition 2. *A unique equilibrium exists. For any $\beta \in [0, 0.5]$, there exists $\overline{\sigma}(\beta) > 1$ (with $\overline{\sigma}(0.5) = \infty$) such that, if $\sigma \leq \overline{\sigma}(\beta)$, the equilibrium is interior and the investment levels are given by:*

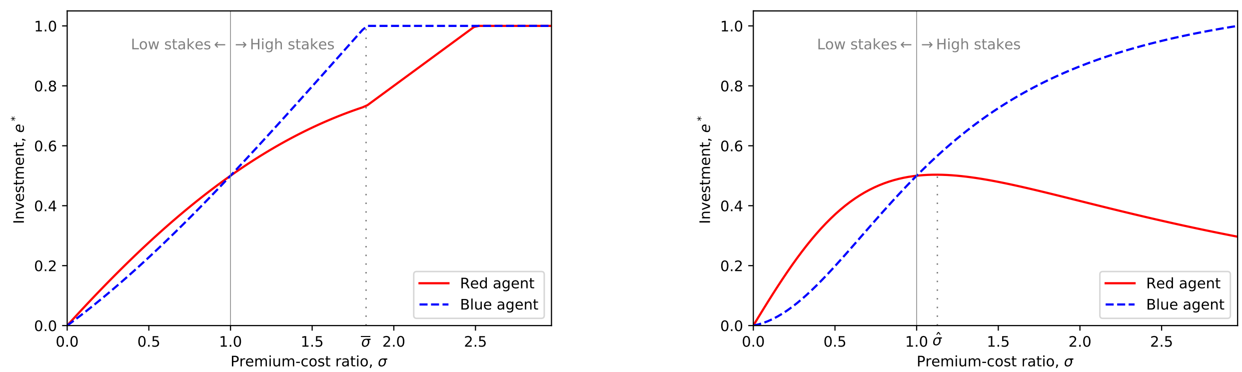
$$e_b^* = \frac{\sigma(0.5-\beta)+2\beta\sigma^2(0.5+\beta)}{1+4\beta^2\sigma^2}; \tag{5}$$

$$e_r^* = \frac{\sigma(0.5+\beta)-2\beta\sigma^2(0.5-\beta)}{1+4\beta^2\sigma^2}. \tag{6}$$

If $\sigma > \overline{\sigma}(\beta)$, $e_b^* = 1$ and $e_r^* = \min\left\{\frac{\sigma(1-2\beta)}{2}, 1\right\}$.

4.4.2 Discouragement versus Overcompensation

Figure 1 shows the equilibrium investment levels as a function of the premium-cost ratio, σ, for two levels of the subtle bias. The figure shows that for low values of σ, Red invests more than Blue, while for high values of σ, it is Blue who invests more. The following corollary formalizes the comparative statics illustrated in Figure 1 (all corollary proofs are in Section OA.1). For simplicity



(a) Small bias, $\beta = 0.1$

(b) Large bias, $\beta = 0.4$

Figure 1: Equilibrium investments, e_b^* and e_r^*, as functions of the premium-cost ratio, σ, for two levels of subtle bias, $\beta_1 = 0.1$ and $\beta_2 = 0.4$.

of exposition, from now on, we assume that the equilibrium is interior.

Corollary 1. *When stakes are low, Red invests more than Blue. When stakes are high, Blue invests more than Red. Formally, $e_r^* \geq e_b^*$ if and only if $\sigma \leq 1$.*

Red's investment decision is shaped by two opposing forces. On the one hand, Red wants to invest heavily in skills to try to separate herself from Blue. We call this force the *overcompensation effect*. Overcompensation may occur because the red agent knows she is held to "higher standards:" unless she is clearly more qualified than Blue, she is viewed less favorably. On the other hand, Red is discouraged from investing because her chances of promotion are slim even if she acquires the skill. We call this force the *discouragement effect*.[8] Parameter σ determines which effect dominates in equilibrium. If the stakes are low, Blue exerts low effort. Thus, Red is willing to overcompensate by investing more, both because the probability of separation is high and because the marginal cost of investing is low under a convex cost function. As the stakes increase, Blue chooses higher levels of investment, discouraging Red from investing. At high investment levels, the probability of separation is low, while the marginal cost of investing is high.

Remark 1. If $l < 0$, job 2 may sometimes stay vacant, and both reaction functions slope downwards (see Section OA.6). In this case, the discouragement effect always dominates the overcom-

[8] See Coate and Loury (1993) and MacLeod (2003) for early models with a similar discouragement effect.

pensation effect, and Red invests less than Blue in equilibrium. Focusing on the $l > 0$ case makes the model richer because either effect may dominate in equilibrium.

Remark 2. The interaction between the overcompensation and discouragement effects is robust in situations where Blue and Red have different beliefs about β. For example, if Red believes that there is subtle discrimination ($\beta > 0$), but Blue thinks that β is zero, we have $e_b^* = \frac{\sigma}{2}$ and $e_r^* = \sigma(\frac{1}{2} + \beta(1-\sigma))$, thus $e_r^* - e_b^* = \beta\sigma(1-\sigma)$, implying that Corollary 1 holds.

Remark 3. Decision-makers may exhibit varying degrees of subtle bias, represented by β_1 when selecting between high-skilled agents ($s_r = s_b = 1$), and β_0 when choosing between low-skilled agents ($s_r = s_b = 0$). For instance, in scenarios involving high-skilled agents, decision-makers might perceive their decisions as subject to heightened scrutiny, leading them to exercise greater caution, thus $\beta_0 \geq \beta_1$. Our findings are robust when considering this extension of the model (see Section OA.7).[9]

Remark 4. The discouragement effect could lead a principal unaware of his bias (and the strategic interaction it creates between the two agents) to misinterpret Red's behavior as a lack of interest in high-paying positions. He might incorrectly "learn" that red and blue agents have different preferences concerning earned income because such a narrative is aligned with his observed data (as in Eliaz and Spiegler (2020)). Such learning might further reinforce the principal's subtle bias or even result in an explicit bias favoring blue agents.

4.4.3 Cost Heterogeneity

To isolate the effect of subtle discrimination on investment incentives from other confounding factors, we have assumed that blue and red agents are identical in all respects except for their labels. However, in reality, favored and unfavored groups often differ systematically in other dimensions, typically due to biases or social norms that influence outcomes even before the promotion decision. For example, women may face higher opportunity costs for investing in skills due to unequal

[9] We thank one of our referees for highlighting this possibility.

distribution of household tasks, such as child care. Additionally, unfavored agents may face higher investment costs because of disparities in mentoring across groups, as in Athey, Avery, and Zemsky (2000). Conversely, there may be instances where unfavored agents may incur lower costs in developing firm-specific skills, particularly if they have higher levels of general human capital from prior educational investments.

How do differential costs affect the discouragement and overcompensation effects? To address this question, let k_i denote agent i's cost parameter. If the equilibrium is interior, the investment levels are given by:

$$e_b^* = \frac{\sigma_b(0.5-\beta)+2\beta\sigma_b\sigma_r(0.5+\beta)}{1+4\beta^2\sigma_b\sigma_r} \tag{7}$$

$$e_r^* = \frac{\sigma_r(0.5+\beta)-2\beta\sigma_b\sigma_r(0.5-\beta)}{1+4\beta^2\sigma_b\sigma_r} \tag{8}$$

where $\sigma_i = \frac{W}{k_i}$. Note that (5) and (6) are special cases of (7) and (8), where $\sigma_b = \sigma_r = \sigma$. We then have the following result:

Corollary 2. *There exists $\widehat{W} \geq 0$ such that $e_r^* \geq e_b^*$ if and only if $W \leq \widehat{W}$.*

Corollary 2 generalizes Corollary 1. Corollary 1 implies that, when $\sigma_b = \sigma_r = \sigma$, we have $\widehat{W} = k$. When costs are different, $\widehat{W}$ may increase or decrease, depending on whether $k_r < k_b$ or $k_r > k_b$. In either case, the discouragement effect eventually dominates the overcompensation effect as the stakes increase.

How do cost differences affect investment incentives? Note that, in the absence of subtle biases, the investment levels are $e_i^* = \frac{\sigma_i}{2}$, implying that the agent with the lowest cost invests more, and the difference between the two agents increases with W. This effect is caused by the productive differences between the two agents; the first-best solution also requires one agent to invest more than the other in this case. If $k_b > k_r$, this *productivity effect* enlarges the region where Red invests more than Blue; conversely, if $k_b < k_r$, it has the opposite effect. If the differences in costs are substantial (in either direction), one of the regions may disappear, as we show in Section OA.8.1. However, we also show that, for any k_b and k_r, there always exists a bias $\beta' \in (0, 0.5]$ such that, for all $\beta > \beta'$, we have $e_r^* > e_b^*$ for low stakes and $e_r^* < e_b^*$ for high stakes. Intuitively, if the subtle bias is sufficiently strong, the discouragement and overcompensation effects always dominate the productivity effect.

An empirical implication is that the productivity effect exacerbates (attenuates) the discouragement effect if Red faces higher (lower) investment costs than Blue. Provided the subtle bias is sufficiently strong, Corollary 2's prediction should be empirically detectable as the stakes change. Therefore, if we observe that Red consistently invests less (or more) than Blue across all stakes, we should conclude that (i) cost differentials are large and (ii) the subtle bias is weak or nonexistent.

While closed-form solutions require quadratic cost functions, in Section OA.8.2 we show numerically that the results in this subsection hold for a large set of convex cost functions. Thus, we confirm that the relationship between stakes and investment differences is not unique to the case of identical cost functions. Section OA.8.3 analyzes the case of general cost functions $c_b(e_b)$ and $c_r(e_r)$. We present conditions under which the results in Corollary 2 hold exactly. We also show that, for some cost functions, a weaker version of Corollary 2 holds: There exist $\widehat{W}_l \leq \widehat{W}_h$ such that $e_r^* \geq e_b^*$ if $W \leq \widehat{W}_l$, and $e_r^* \leq e_b^*$ if $W \geq \widehat{W}_h$. That is, Red invests more than Blue for sufficiently low stakes, and Blue invests more than Red for sufficiently high stakes. In this case, in between $\widehat{W}_l$ and $\widehat{W}_h$, there might exist $\widehat{W}_m$ such that increasing the stakes makes $e_r^* - e_b^*$ switch from negative to positive in an interval around $\widehat{W}_m$. Intuitively, this may happen when, at intermediate investment levels, Red has a low marginal cost of investing, and this marginal cost increases at a lower rate than Blue's marginal cost. In such a case, Red has a productive *comparative advantage* at intermediate investment levels.

In sum, the property that Red invests more (less) than Blue when stakes are low (high) is robust to agents having different cost functions. It holds exactly for quadratic cost functions and several other cases, as shown numerically in Section OA.8.2. In the special case where this property does not hold for intermediate stakes, a weaker version still holds for extreme values of W.

For the remainder of the paper, we assume that agents are identical except for their labels, i.e., $k_b = k_r = k$.

4.4.4 Subtle Discrimination versus Overt Discrimination

To understand how the model's predictions relate to the type of discrimination, we consider how these predictions would change in the presence of overt discrimination. We define the overt bias as $\delta = \lim_{\omega \to 0} b(0,1,\omega)$. That is, the principal promotes an unskilled blue agent, $s_b = 0$, ahead of a skilled red agent, $s_r = 1$, with probability δ. If $\delta > 0$, the reaction functions become

$$e_b = \sigma \left[\frac{1}{2} - \beta + (2\beta - \delta) e_r \right] \text{ and } e_r = \sigma \left[\frac{1}{2} + \beta - \delta - (2\beta - \delta) e_b \right]. \tag{9}$$

Thus far, we have imposed no structure on β and δ other than $\delta \leq 0.5 + \beta$, which follows from $P(s_b, s_r, \omega)$ being decreasing in s_r and increasing in s_b. Without further structure, we do not know whether the reaction functions in (9) have positive or negative slopes. To impose further (but minimal) structure, we can think of overt discrimination as a probability over two states of the world: in the first state, which happens with probability δ, the principal has a large bias towards Blue and, thus, chooses Blue with probability one, while in the second state, which happens with probability $1 - \delta$, the principal's bias is small; thus he still chooses Red if $s_b = 0$ and $s_r = 1$. While in State 2, discrimination does not occur when $s_b = 0$ and $s_r = 1$, it may still happen when $\Delta s = 0$ (both agents have the same skill) because even a small bias might affect decisions in this case. Formally, this implies that δ imposes a lower bound on β: $\frac{1}{2} + \beta \geq \delta + (1 - \delta)\frac{1}{2}$, which implies $2\beta \geq \delta$.

Let $\varepsilon \equiv \beta - \frac{\delta}{2} \geq 0$ denote the *excess subtle bias*. If $\varepsilon = 0$ (no excess subtle bias), the reaction functions in (9) are again flat, implying $e_b = e_r = \frac{\sigma(1-\delta)}{2}$ (for $\delta < 1$). That is, if subtle discrimination is fully "explained" by an overt bias, both agents choose the same investment levels in equilibrium. Asymmetric investment levels occur only when subtle discrimination is stronger than overt discrimination. In other words, overt discrimination moderates the effect of subtle discrimination.

We now generalize Corollary 1 for the case in which overt discrimination is present:

Corollary 3. *If $\delta \geq 0$, $e_r^* \geq e_b^*$ if and only if $\sigma \leq \frac{1}{1-\delta}$.*

Corollary 3 shows that overt discrimination makes it less likely that Red invests more than Blue. Intuitively, under overt discrimination, Red benefits less from separating herself from Blue because, even when Red is more skilled than Blue, she is still passed over for promotion with positive probability. That is, the overcompensation effect is weaker when the principal also overtly discriminates. These results show that subtle discrimination has unique implications. In the context of our model, subtle discrimination creates incentives for separation, while overt discrimination does not. Only in the presence of excess subtle discrimination can the overcompensation effect dominate the discouragement effect. Therefore, our model is particularly relevant when overt discrimination is mild or nonexistent, but subtle discrimination remains.

For the remainder of the paper, we consider only the case of "pure" subtle discrimination, i.e., $\delta = 0$. All results are qualitatively unchanged if we interpret β as excess subtle bias.

4.4.5 Stakes, Investment in Skills, and the Promotion Gap

The next corollary presents further comparative statics results.

Corollary 4. *For $\sigma \leq \overline{\sigma}(\beta)$ (i.e., the equilibrium is interior), we have that*

1. *e_b^* is strictly increasing in σ;*

2. *There exists $\widehat{\sigma}(\beta) \leq \overline{\sigma}(\beta)$ such that e_r^* increases with σ for $\sigma \leq \widehat{\sigma}(\beta)$ and decreases with σ for $\sigma > \widehat{\sigma}(\beta)$.*

3. *$\widehat{\sigma}(\beta)$ is strictly decreasing in β.*

Part 1 shows that Blue's investment in skill acquisition is increasing in the premium-cost ratio. Part 2 shows that Red's investment does not always increase with σ. If the stakes are sufficiently high ($\sigma > \widehat{\sigma}(\beta)$), the discouragement effect dominates, and Red's investment declines with the premium-cost ratio (for this to happen, the subtle bias needs to be sufficiently strong). Part 3 shows that the discouragement effect is stronger when the bias is stronger, implying a lower premium-cost ratio at which Red's investment declines with σ.

Let p_i denote agent i's promotion probability, $i \in \{b, r\}$. The promotion gap between blue and red agents is

$$\Delta p \equiv p_b - p_r = e_b - e_r + [e_b e_r + (1-e_b)(1-e_r)]\, 2\beta. \tag{10}$$

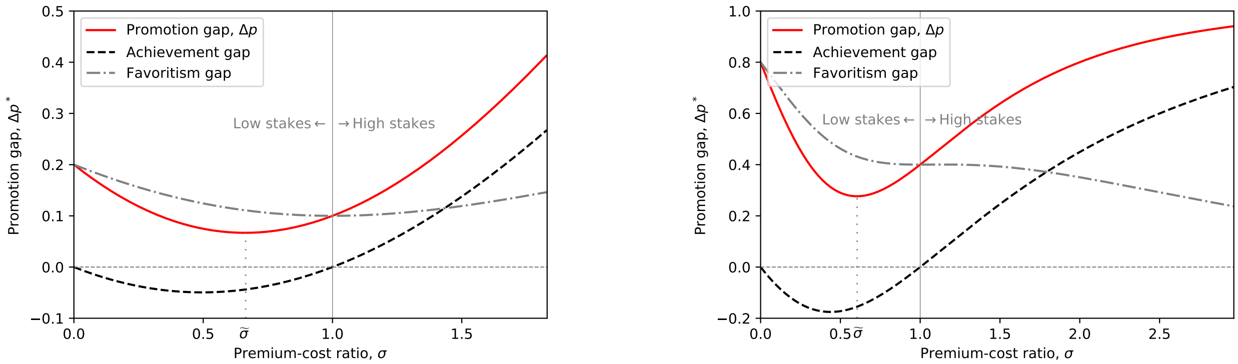
The promotion gap has two terms. The first term, $e_b - e_r$, is the difference in the probabilities of skill acquisition. Given our broad interpretation of skill, we call this difference the *achievement gap*. The second term is the difference in promotion probabilities between Blue and Red that arises as a direct consequence of the subtle bias. It is the promotion gap conditional on a tie times the probability of a tie. We call this term the *favoritism gap*. The subtle bias affects both the achievement gap and the favoritism gap. The following proposition shows how the equilibrium promotion gap varies with the premium-cost ratio, σ.

Proposition 3. *For each $\beta \in (0, 0.5]$, there exists a unique premium-cost ratio $\widetilde{\sigma}(\beta)$ such that the promotion gap decreases in σ for $\sigma < \widetilde{\sigma}(\beta)$ and increases in σ for $\sigma \in (\widetilde{\sigma}(\beta), \overline{\sigma}(\beta))$.*

Figure 2 illustrates how the promotion gap changes with the premium-cost ratio, σ. The promotion gap initially decreases with σ and then increases with σ. For large values of the premium-cost ratio, even a small subtle bias can be significantly amplified through the strategic interactions between the agents. Also, in high-stakes careers, the contribution of the achievement gap to the promotion gap is greater than that of the favoritism gap. That is, differences in "observable" achievements (human capital, performance, experience, effort, etc.) explain most of the promotion gap. In other words, because ties occur less frequently as the stakes increase, the principal is less likely to make biased promotion decisions. In such scenarios, we would expect to find little direct evidence of discrimination. Nevertheless, promotion gaps are large in high-stakes situations.

4.5 Optimal Compensation Contracts

We now allow the principal to design the compensation contract. As is standard in principal-agent problems, we assume that the firm is a monopsonist in the labor market; that is, it has all the bargaining power. The principal cannot discriminate through contracts explicitly; he must offer



(a) Small bias, $\beta = 0.1$

(b) Large bias, $\beta = 0.4$

Figure 2: Equilibrium promotion gap, Δp^*, as a function of the premium-cost ratio, σ, for two levels of subtle bias, $\beta_1 = 0.1$ and $\beta_2 = 0.4$.

the same contract $\mathbf{w} = (w_1, W)$ to both agents. Agents are assumed to be penniless; wages must be non-negative: $w_1 \geq 0$ and $w_1 + W \geq 0$.

To remain in a fully rational world, we assume that the principal knows that the agents behave as if promotions are subject to subtle bias β. One interpretation is that the principal is aware of his own bias. Under this interpretation, the subtle bias may create a dynamic inconsistency problem: the principal could be (in some cases) better off by committing not to discriminate, but no commitment technology is available. In the language of O'Donoghue and Rabin (1999), the principal is a "sophisticate," i.e., someone who understands that they will subtly discriminate and, therefore, can correctly predict their future behavior. A second – and perhaps more empirically relevant – interpretation is that promotion decisions are made by a biased third party (e.g., a direct supervisor), and the principal designs the contract taking into account the third party's bias (see Prendergast and Topel (1996) for a model along these lines).

Agents' outside utilities are normalized to zero. To avoid corner solutions, we assume that the firm pays a fixed entry cost to operate; to save on notation, we assume that this cost is $l + \varepsilon$, with ε arbitrarily small. This assumption implies that the firm chooses to operate if and only if the expected profit after entry is strictly greater than l. The principal is risk-neutral and derives no utility from discrimination. His profit-maximization problem (after entry, i.e., gross of entry costs) is as follows:

$$\max_{w_1 \geq 0, w_1 + W \geq 0} l + H\left(e_b + e_r - e_b e_r\right) - 2w_1 - W, \tag{11}$$

subject to

$$e_i = \arg\max_{e\in[0,1]} eW\left[\left(\frac{1}{2} - \beta_i\right) + 2\beta_i e_{-i}\right] - \frac{ke^2}{2}, \text{ for } i \in \{b, r\}, \tag{12}$$

where $\beta_b = -\beta_r = \beta$. The principal faces two incentive compatibility (IC) constraints in (12). The agents' participation constraints are not binding because $w_1 \geq 0$ and $w_1 + W \geq 0$ imply that agent i can guarantee a non-negative payoff by choosing $e_i = 0$. Because w_1 does not affect the IC constraints, the principal optimally sets $w_1 = 0$. If the principal chooses $w_1 = W = 0$, the agents exert no effort, and the post-entry profit is l. In such a case, the firm's profit from entering the market is $l - l - \varepsilon < 0$. Thus, we use $w_1 = W = 0$ to denote the case in which the firm does not operate.

For any given k, choosing the wage upon promotion, W, is equivalent to choosing the stake, σ. Henceforth, for convenience and clarity, we view the principal's task as selecting σ. Proposition 2 implies that $e_b = 1$ if $\sigma > \overline{\sigma}(\beta)$, thus increasing σ beyond $\overline{\sigma}(\beta)$ has no impact on revenue. That is, in an optimal contract, $\sigma \leq \overline{\sigma}(\beta)$. With these observations, the principal's problem can be written as:

$$\Pi(k, \beta, \theta) = \max_{\sigma \in [0, \overline{\sigma}(\beta)]} k\theta\left(e_b + e_r - e_b e_r\right) - k\sigma, \tag{13}$$

subject to (5) and (6), where $\theta \equiv \frac{H}{k}$ is the *productivity-cost ratio* and $\Pi(k, \beta, \theta)$ is the optimal expected profit net of entry costs (as $\varepsilon \to 0$). Parameter θ can also be interpreted as a measure of the relative importance of human capital at higher hierarchical levels. Thus, for interpretation, we call firms with high θ *human-capital-intensive firms*.

We first solve a baseline case in which $\beta = 0$.

Proposition 4. *If the principal is unbiased ($\beta = 0$), the firm operates if and only if $\theta > 1$ and the optimal stake, $\sigma^* = \frac{2(\theta-1)}{\theta}$, uniquely implements investment levels $e_b^* = e_r^* = \frac{\theta-1}{\theta}$.*

When there is no bias, both agents choose the same investment level in equilibrium. The firm operates only when $\theta > 1$, i.e., the productivity gain for the principal is high relative to the marginal cost of investment for the agents. That is, firms with low productivity-cost ratios prefer to shut down. From a social welfare perspective, all firms should operate because when $e_b = e_r = 0$, the marginal cost of investing is zero, while the marginal social benefit of investing is positive and equal to $H > 0$. Thus, when $\theta \leq 1$, the firm inefficiently stays out of business. Such inefficiency occurs because the non-negative wage constraint prevents the firm from extracting all the surplus from the agents.

An analytical solution is not universally attainable in the general case ($\beta \geq 0$). However, it is possible to prove the existence and uniqueness of the optimal contract:

Proposition 5. *For every set of parameters $(k > 0, \beta \in [0, 0.5], \theta > 0)$, there exists a unique solution $\sigma(k, \beta, \theta)$ to the principal's problem (if the firm chooses not to operate, we set $\sigma = 0$).*

Without loss of generality, from now on, we set $k = 1$. Let $\sigma(\beta, \theta)$ denote the optimal stake. The following result describes the properties of the optimal contract and how it changes with the productivity-cost ratio, θ.

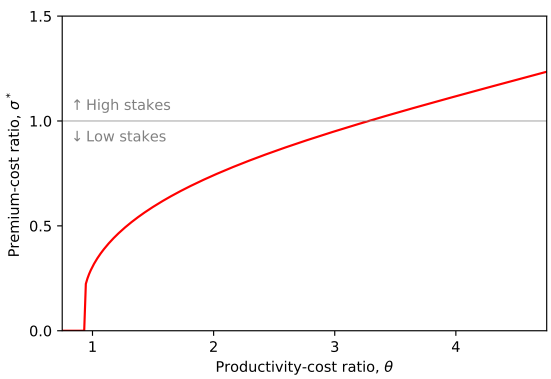
Proposition 6. *For every $\beta \in [0, 0.5]$, there exist values $\underline{\theta}(\beta) < \overline{\theta}(\beta)$ such that:*

1. *If $\theta \leq \underline{\theta}(\beta)$, the optimal stake is $\sigma(\beta, \theta) = 0$ (i.e., the firm does not operate). If $\theta \geq \overline{\theta}(\beta)$, the optimal stake is $\sigma(\beta, \theta) = \overline{\sigma}(\beta)$.*

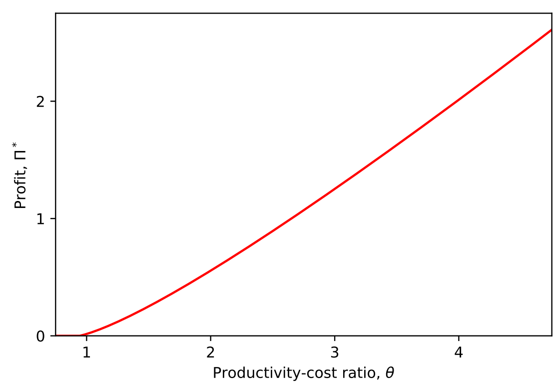
2. *The optimal stake, $\sigma(\beta, \theta)$, is strictly increasing in $\theta \in \left[\underline{\theta}(\beta), \overline{\theta}(\beta)\right]$.*

3. *The firm's profit is strictly increasing in $\theta \geq \underline{\theta}(\beta)$.*

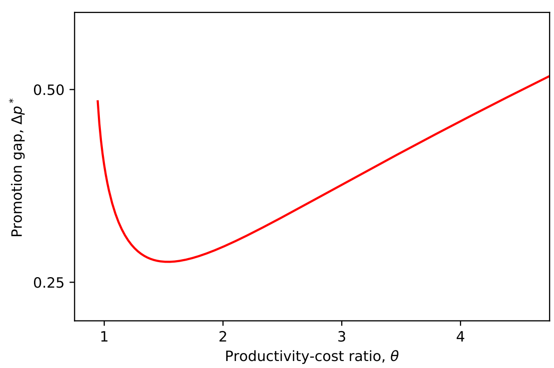
Part 2 of Proposition 6 implies that human-capital-intensive firms (high-θ firms) offer career paths involving higher stakes. Because the optimal stake is increasing in θ, all the comparative statics in the previous subsection are unchanged once we replace σ with θ. In particular, if we define $\widetilde{\theta}(\beta)$ as the value of θ such that the optimal stake is $\sigma(\beta, \widetilde{\theta}(\beta)) = 1$, we again have that Red invests more than Blue when stakes are low ($\theta < \widetilde{\theta}(\beta)$) and Blue invests more than Red when stakes are high ($\theta > \widetilde{\theta}(\beta)$). Finally, Part 3 implies that high-θ firms are more profitable. Thus, we can also use θ as a proxy for firm profitability or productivity. Panels (a) and (b) of Figure 3 illustrate Proposition 6 (for $\beta = 0.4$), while panel (c) shows that similar to Figure 2, the equilibrium promotion gap is U-shaped in the productivity-cost ratio, θ.



(a) Premium-cost ratio, $\sigma(\beta = 0.4, \theta)$



(b) Profit, $\Pi(\beta = 0.4, \theta)$



(c) Promotion gap, $\Delta p(\beta = 0.4, \theta)$

Figure 3: Optimal premium-cost ratio, equilibrium profit, and promotion gap, as functions of the productivity-cost ratio, θ for a given level of subtle bias ($\beta = 0.4$).

4.6 Optimal Anti-Discrimination Policies

Does subtle discrimination benefit or harm firms? To see how subtle discrimination affects profits, we now consider the problem of a principal who can choose both the compensation contract and the firm's own subtle bias. We have in mind a situation in which the firm chooses an optimal anti-discrimination policy. For example, the firm can set up processes that lead to the selection of supervisors with high or low subtle biases. Similarly, the firm can invest in a corporate culture that is either friendly or hostile to diversity goals. Firms can become conservative by adopting policies associated with high β. Similarly, firms can become progressive by adopting policies associated with low β.

Suppose the principal can select policies associated with bias β at no cost (another interpretation is that market forces drive firms with suboptimal biases out of the market). Which β would the principal choose? The principal's problem is

$$\Pi(\theta) = \max_{(\sigma,\beta)\in[0,\overline{\sigma}(\beta)]\times[0,0.5]} \theta\left(e_b + e_r - e_b e_r\right) - \sigma, \tag{14}$$

subject to (5) and (6). Let $\beta(\theta)$ denote the profit-maximizing subtle bias and $\sigma(\theta)$ the corresponding optimal stake. Define $\overline{\theta}$ as the lowest value of θ such that $\sigma(\theta) = \overline{\sigma}(\beta(\theta))$. That is, the optimal stake is strictly interior if and only if $\theta \leq \overline{\theta}$. From now on, we focus on strictly interior solutions for σ.

One might think that firms would obviously choose to have no bias, i.e., $\beta = 0$. Indeed, in the unconstrained first-best (assuming both workers need to be hired), both workers invest the same amount. However, the unconstrained first-best is typically not achievable because the non-negative wage constraint prevents the firm from extracting all the surplus from the workers. While non-negative constraints are just a modeling feature here, in the real world, workers face borrowing constraints. Thus, workers typically need to earn a minimum (non-negative) wage. This constraint is more likely to be binding in low-stakes situations. Due to this friction, it is no longer evident that the firm prefers to treat all agents equally, i.e., to set $\beta = 0$. The contest literature shows that asymmetric contests may be optimal even when agents are identical (see Kawamura and de Barreda (2014) and Drugov and Ryvkin (2017)). However, most papers in this literature do not explicitly model the frictions that may lead to the optimality of biased contests. In particular, no paper in this literature distinguishes between subtle and overt biases. The next proposition shows how a subtle bias might be optimal when firms cannot extract all the surplus through low initial wages:

Proposition 7. *There exists $\theta' < \overline{\theta}$ such that*

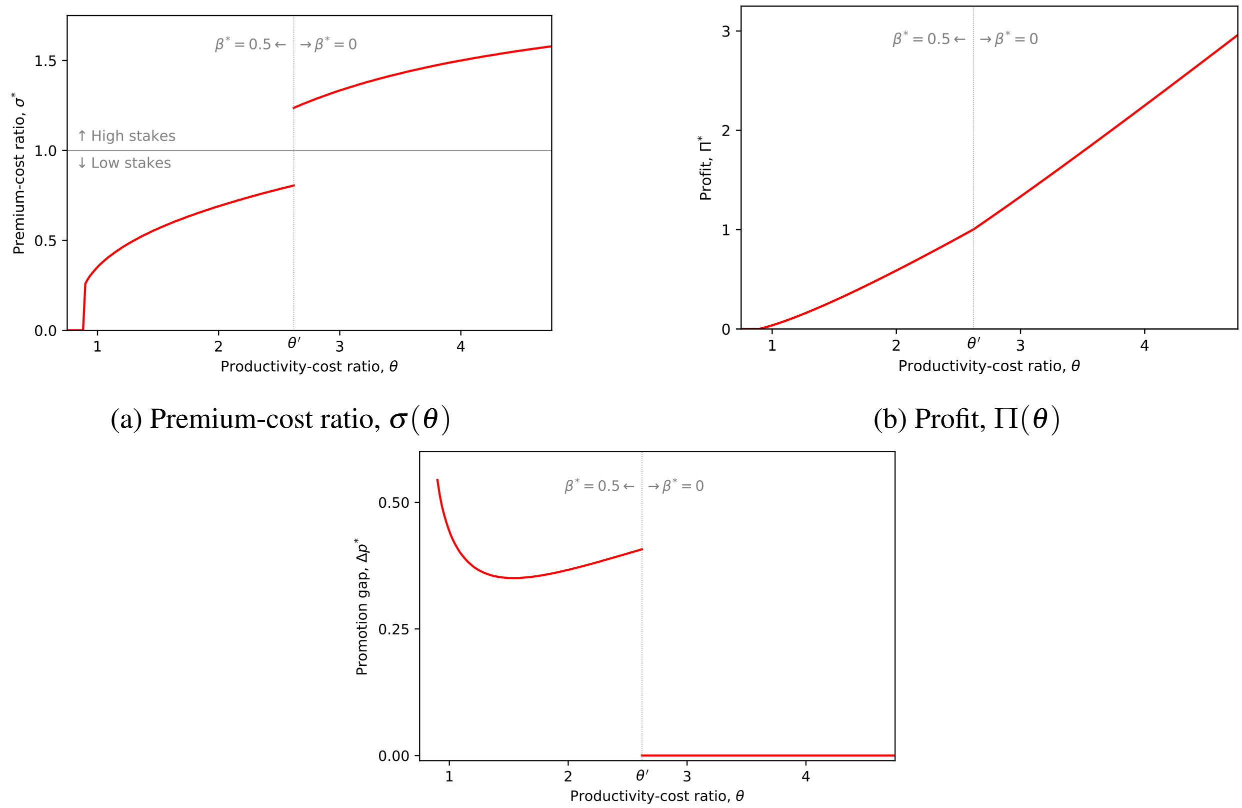
$$\beta(\theta) = \begin{cases} 0.5 & \text{if } \theta \in (0, \theta'] \\ 0 & \text{if } \theta \in [\theta', \overline{\theta}] \end{cases}. \tag{15}$$

Furthermore, $\sigma(\theta) < 1$ *if* $\theta \in (0, \theta']$ *and* $\sigma(\theta) > 1$ *if* $\theta \in [\theta', \overline{\theta}]$.

This proposition shows that if the principal could optimally choose his subtle bias (or, equivalently, a supervisor with a given bias) at no cost, he would always choose a corner solution for the bias: either no bias or the maximum bias. This choice is determined by the productivity-cost ratio, θ. Figure 4 illustrates the optimal stake, profit, and the resulting promotion gap as functions of θ. For less productive firms, i.e., firms with low θ, profits increase with subtle discrimination. Thus, firm profit is maximized at $\beta^* = 0.5$. Such firms also optimally offer low-stakes careers, $\sigma(\theta) < 1$. Intuitively, subtle discrimination is profitable for firms that offer low-stakes careers because the overcompensation effect improves the performance of discriminated agents. Thus, in less productive (or less human-capital-intensive) sectors, firms perform better when they discriminate.

In contrast, high-θ firms greatly profit from promoting skilled agents and are therefore willing to offer high stakes to incentivize agents to invest in their skills. However, under high stakes, the discouragement effect hinders the investment of discriminated agents. Thus, high-θ firms prefer to choose robust anti-discrimination policies to counter the discouragement effect. That is, in high-θ firms, the maximal profit is achieved with zero subtle bias. This result implies that in sectors with high θ, discriminating firms are less profitable than non-discriminating firms, and thus, they are more likely to be driven out by competition.

When firms optimally choose their biases, high-θ firms do not have a promotion gap (see Figure 4c). That is, such firms have more diversity at the top. In contrast, low-θ firms have positive promotion gaps. Thus, our model predicts a particular type of firm polarization, in which high- and low-productivity firms choose different policies regarding discrimination and diversity. Note that our model predicts a large promotion gap for high-productivity firms <u>conditional on a given bias</u>. When firms can choose their biases, high-productivity firms have small promotion gaps.



(a) Premium-cost ratio, $\sigma(\theta)$

(b) Profit, $\Pi(\theta)$

(c) Promotion gap, $\Delta p(\theta)$

Figure 4: Optimal premium-cost ratio, equilibrium profit, and promotion gap, as functions of the productivity-cost ratio, θ.

In Section OA.7, we show that if firms can choose their biases separately for low-skilled and high-skilled agents, then for low-skill ties (i.e., $s_b = s_r = 0$), they always choose the highest possible bias, $\beta_0^* = 0.5$. This bias provides the strongest incentives for discriminated agents to invest in their human capital to avoid ties at zero. In this case, the firm maximizes its benefit from the overcompensation effect. For high-skill ties (i.e., $s_b = s_r = 1$), the optimal bias is again high for low-θ firms ($\beta_1^* = 0.5$) and zero for high-θ firms. In this case, the promotion gap is positive for low-θ firms and negative for high-θ firms, due to a large negative achievement gap. Thus, by separately choosing biases for low-skilled and high-skilled agents, firms choose β_0 to maximize their benefits from the overcompensation effect (by minimizing ties at zero skill) and β_1 to trade off the overcompensation and discouragement effects.

The results concerning firm policies can be summarized as follows. High-productivity firms prefer to promote a discrimination-free work environment. They strive to be perceived as "progressive" and "activist." If successful, they would have more diversity at the top (i.e., a smaller promotion gap). These firms also offer careers with higher stakes and are likely to be large, profitable, and human capital-intensive. In contrast, low-productivity firms do not take actions to counter subtle discrimination. They do not mind being perceived as "conservative" and have less diversity at the top. They offer careers with low stakes and are smaller, less profitable, and less human capital-intensive than "progressive" firms. Note that polarization implies that two firms with productivity-cost ratios $\theta' + \eta$ and $\theta' - \eta$ adopt very different anti-discrimination policies even if their differences in productivity are negligible (i.e., as $\eta \to 0$).

Firms want workers to learn about β because the benefits from choosing an optimal β arise from the workers' ex-ante investments. Thus, firms have incentives to communicate truthfully and use credible signals of their biases. For example, a company for which $\beta^* = 0.5$ may want to signal this bias by distorting its promotion decisions.[10] Even if the principal is "intrinsically" unbiased, it is in his interest to appear biased. Thus, the principal discriminates in promotion decisions to influence agents' beliefs in future decisions. In this example, the principal is "strategically biased:" the source of bias is extrinsic rather than intrinsic. As discussed in Section 3, our definition of subtle bias is agnostic concerning the sources of bias.

4.7 Welfare and Robustness

Section OA.10 shows that a larger bias can increase or decrease social welfare. Perhaps surprisingly, for low-θ firms, a larger bias might be Pareto improving. In such firms, even red agents can be better off under a larger β because the firm offers a higher promotion premium in an attempt to exploit the overcompensation effect.

In Section OA.11, we analyze a hard quota's impact on protecting red agents and show that the quota has its desired effect only when the bias is sufficiently high. However, despite equal

[10] In practice, the principal would prefer to maintain plausible deniability to avoid potential lawsuits. He would not explicitly state that in ties, he prefers blue agents over red because such statements can be used in court.

outcomes, red agents still experience lower expected utility than blue agents because firms lower their promotion stakes under the quota, leading to red agents working harder than blue agents.

In Sections OA.2-OA.9, we show that our results are robust and not driven by several assumptions used in the main text to ensure tractability. In particular, our results are not driven by skills being firm-specific (see OA.2 and OA.3) or being discrete or binary (see OA.4 and OA.9), firms having all the bargaining power (see OA.2), the promotion premium being independent of skill level (see OA.2), the effort cost function being quadratic (see OA.5) or the same for two agents (see OA.8), firms always having to promote an agent to the top position (see OA.6), or subtle bias being the same for all skill levels (see OA.7).

5 Further Applications

This section briefly explores how our framework can be applied to other settings. We show through simple examples how applying our notion of subtle bias to other contexts can explain existing findings and deliver new empirical predictions.

5.1 Equity and Credit Analysts

Hong and Kubik (2003) show that equity analysts are rewarded with better career outcomes for both accuracy and optimism in their forecasts. Kisgen et al. (2020) show that credit rating analysts face similar incentives: they are rewarded for being accurate and avoiding rating downgrades.

Forecasting accuracy is a relatively objective performance metric. Optimism, on the other hand, is an implicit goal; investment banks and rating agencies are unlikely to acknowledge a trade-off between accuracy and optimism. Thus, accuracy is a hard metric from an analyst's perspective, while optimism is implicit and subjective.

If firms are biased against some types of analysts, overt and subtle biases may have different implications. Suppose firms assess an analyst's performance by metric $y = y_1 + y_2$, where y_1 is accuracy (objective) and y_2 is optimism (subjective). Assume that the probability of promotion increases in y and let W denote the promotion premium. An analyst of type i can improve his/her accuracy by investing in human capital at cost $k\frac{e_i^2}{2}$. Set $y_{1i} = e_i$ and $y_{2i} = B - ay_{1i}$, where $B > 0$ and $0 < a < 1$. That is, there is a trade-off between accuracy and optimism. A biased firm uses metric $\hat{y}_i = (1-\delta_i)(y_{1i} + (1-\beta_i)y_{2i})$ to assess a type-i analyst, where $\delta_i \in [-1,1]$ is the firm's overt bias and $\beta_i \in [-1,1]$ is the firm's subtle bias against type i. As before, let $\delta_r = -\delta_b = \delta \geq 0$ and $\beta_r = -\beta_b = \beta \geq 0$. Assuming for simplicity that the probability of promotion is $\hat{y}_i$, red and blue analysts choose to invest

$$e_r^* = \sigma(1-\delta)\big(1-a(1-\beta)\big) \text{ and } e_b^* = \sigma(1+\delta)\big(1-a(1+\beta)\big), \tag{16}$$

where, as before, $\sigma = W/k$. Equation (16) implies that a larger overt bias decreases Red's investment and increases Blue's investment, while a larger subtle bias increases Red's investment and decreases Blue's investment. As before, a subtle bias incentivizes Red to overcompensate, while an overt bias discourages Red. An implication is that, in the absence of overt biases, red analysts invest more than blue analysts to compensate for the subtle bias, implying that red analysts are more accurate than blue analysts. This simple model can thus rationalize the evidence that female analysts are more accurate than male analysts (Kumar (2010)); female analysts overinvest in forecasting skills as a rational response to subtle biases. This example illustrates the relevance of our classification for models that combine objective and subjective performance measures.[11]

5.2 Lending

There is an important literature on discrimination in lending decisions, in particular for personal loans (see, e.g., Dobbie et al. (2021); Bartlett et al. (2022); Butler et al. (2023); and Frame et al. (2023)). Due to the prevalence of automated lending decisions based on hard information, the scope for overt discrimination is limited. Yet, when these automated systems do not make definitive recommendations, loan officers' discretion allows subtle biases to influence their decisions.

[11]Many other professionals face a similar incentive structure. For example, investment and commercial bankers, beyond deal closure and loan underwriting, must nurture client relationships to maximize future revenue potential.

Subtle biases may help explain some of the puzzling findings in this literature. For example, Frame et al. (2023) find that white loan officers are less skilled at screening minority applicants than white applicants with similar observable characteristics. While they find that minority applicants are less likely to be approved than white applicants working with the same white loan officer, approved minority borrowers have higher default rates ex-post. To see why this finding is puzzling, consider a simple example in which there are two (payoff-relevant) types of applicants: G (low default probability) and B (high default probability). Suppose the proportions of G and B are the same across blue and red applicants. The bank wants to lend to G but not to B. However, the proportion of G is sufficiently large that, without screening, the bank would like to approve all applicants. In that scenario, an unbiased loan officer who can screen blue applicants but not red applicants would approve all type-G blue applicants and reject all type-B blue applicants. The same officer would approve all red applicants. Thus, red applicants would have higher approval rates than blue applicants and, conditional on approval, higher default rates than blue borrowers. Thus, the puzzling piece of evidence in Frame et al. (2023) is the lower approval rate of minority applicants. Suppose the loan officer is subtly biased: he trusts his soft signal x_r despite being uninformative (i.e., $\omega = 0$). In this case, the officer arbitrarily rejects several red applicants without improving the expected default rate for that group. Thus, subtle discrimination could explain low approval rates for minority applicants, even when loan officers cannot distinguish between good and bad applicants within that group.

The discouragement effect of subtle discrimination can also help explain why minority applicants may exert less effort to complete their loan applications (Frame et al. (2023)) or to shop around for better loan terms. These effects may be further amplified if minority applicants receive lower quality financial services, e.g., less help to produce the paperwork necessary for completing their applications because they tend to live in low-income areas (Huang et al. (2024)).

5.3 Fund Flows

Evidence shows that investors consider fund managers' personal characteristics when choosing among funds. Kumar et al. (2015) show that mutual fund managers with foreign-sounding names have lower fund flows despite having similar performance. Niessen-Ruenzi and Ruenzi (2019) show similar results when comparing funds managed by men and women. Cassel et al. (2022) find that flows are less sensitive to the performance of Black and Hispanic-owned funds in private capital markets. They also find that minority-owned funds have similar performance to white-owned funds. In all these papers, the authors interpret their findings as evidence of investor discrimination. To address the possibility that true differences in skill are masked by decreasing returns to scale, as the Berk and Green (2004) model predicts, Kumar et al. (2015) and Niessen-Ruenzi and Ruenzi (2019) show experimental evidence that investors are more likely to choose *index* funds managed by non-foreign-sounding or male managers. In addition, Kumar et al. (2015) show that arguably exogenous shocks to biases affect fund flows. Similarly, Cassel et al. (2022) find that diverse-fund flows become more sensitive to performance during periods of high racial awareness.[12]

Despite the direct evidence of discrimination, one could still argue that the lack of significant differences in fund performance is evidence *against* taste-based discrimination. The gold standard for distinguishing between taste-based discrimination (including stereotypes/incorrect beliefs) and rational statistical discrimination is the "Becker marginal outcome test" (Becker (1957, 1993)). The idea is that if one group is discriminated against, its marginally treated members should outperform those from the favored group. Thus, under the null hypothesis of no discrimination (or, more generally, rational statistical discrimination), there should be no group differences in the performance of marginally treated agents.

Suppose that investor discrimination is subtle. Thus, it can be modeled as a lexicographic choice: when selecting funds, investors first consider past performance and then use payoff-irrelevant manager characteristics to choose among funds with similar past performance. That

[12]In contrast, Evans and Young (2022) find that ethnically diverse mutual fund managers have lower performance rankings. Conditional on performance, they are still less likely to be promoted than local managers. These effects are stronger in countries with higher levels of nationalism.

is, if fund managers are close substitutes, a subtle bias towards one group may not affect an investor's return.[13] In this case, an outcome test cannot reject its null hypothesis; unlike overt biases, subtle biases are harder to detect with outcome tests. We conclude that subtle discrimination should feature alongside statistical discrimination as the null hypothesis in Becker outcome tests.

6 Testing and Further Evidence

There are several ways in which one can test for subtle discrimination. One is to identify a direct shock to the bias. According to our model, an ex-post, unanticipated preference shock would change promotion gaps but would not impact firm performance in the short run. As an example of this approach, Ronchi and Smith (2021) find evidence that an exogenous shock to male managers' gender attitudes – the birth of a daughter as opposed to a son – increases managers' propensity to hire female workers. They also find that the shock does not affect firm performance, which is explained by managers replacing men with women with comparable qualifications, experience, and earnings. Overall, the evidence is consistent with subtle discrimination affecting gender gaps but not profits in the short run.

An emerging literature documents gender promotion gaps in the banking industry. Using data from a financial institution, Bircan et al. (2023) find evidence of subtle mechanisms that disadvantage women in promotions: differential assignment of men and women to leadership roles in team projects. Individual managers play a crucial role: those with children tend to favor women, while those who take longer to become directors favor men. Ceccarelli et al. (2024) also highlight the importance of individuals rather than institutions in explaining the gender promotion gap among senior commercial bankers. They also find that banks with larger promotion gaps perform worse. However, they do not compare the effects of promotion gaps at the bottom and the top of the hierarchy. Our model predicts that larger gaps at the top are costlier to firms. Huang et al. (2023)

[13] Cassel et al. (2022) explicitly make this point. They argue that biases can explain their findings if many funds deliver the same returns. In that case, bias-driven discrimination by investors would not affect their returns. This result follows because investors choose among funds that are effectively "tied" in expected performance, as in our main model.

document a gender gap in promotions among mortgage loan officers driven by biased beliefs about women's managerial ability. If this bias is overt, then both male and female officers should decrease their efforts, while a subtle bias results in a differential effort response, as discussed in Section 4.4.4. Huang et al. (2023) find that, unconditionally, female officers have worse performance in terms of loan volume than male officers, which is consistent with the discouragement effect.

Our model can also be applied to understand the choices faced by potential founders competing for a scarce pool of capital (e.g., angel or VC investors). Significant gender and racial gaps are found at every stage of the startup life cycle (Ewens (2023)). Several papers in this literature show evidence suggesting that such gaps are driven at least partially by biases (see, e.g., Ewens and Townsend (2020) and Hebert (2023)). The typical startup path has many stages; in every one of them, there are opportunities for biases to influence decisions. Initially, a founder may need to invest in education or acquire work experience (see Ewens (2023)). Thus, expected discrimination in later stages may affect a founder's human capital investments. Does discrimination discourage such investments? Or do discriminated founders overcompensate by investing more than the average? As our model illustrates, the theoretical predictions depend largely on whether biases are subtle or overt. Thus, our model generates new predictions for this rapidly growing empirical literature.

Academia is yet another domain where large disparities in the representation of women and minorities have been documented, especially in STEM (see, for example, Llorens et al. (2021)). Evaluations of academic success often include a significant subjective element, making them susceptible to subtle biases. For example, Witteman et al. (2019) analyze a natural experiment of research grant applications with two treatments: one with an explicit review focusing on the principal investigator's caliber and one without such a focus. While women receive similar scores for research quality compared to men, they receive less favorable assessments as principal investigators in the "focus on the scientist" treatment, contributing to a gender gap in application success. In a study of careers of academic economists, Sarsons et al. (2021) show that while both men and women benefit equally from solo authorship, co-authorship harms women's chances of being tenured. This evidence is compatible with our notion of subtle discrimination: employers are likelier to "break the tie" in favor of male co-authors when attributing credit for joint work. Hengel (2022) provides evidence of the overcompensation effect in academic writing: papers written by women are better written than equivalent papers written by men.

7 Conclusion

We contribute to the discrimination literature in economics by introducing and formalizing the concept of subtle discrimination. Because subtle discrimination leaves no trace, it is subject to plausible deniability. Our leading example of subtle discrimination is the use of biased tie-breaking rules in promotion contests. We also discuss how differentiating between subtle and overt discrimination is helpful when analyzing evidence from other contexts, such as analysts' careers, lending discrimination, and fund management. Generally, we find that subtle and overt discrimination have markedly different empirical predictions.

Our model of promotions generates several novel predictions. The model shows that when promotion stakes are high, unfavored agents are discouraged from investing in human capital. While it is not always clear how to measure "promotion stakes," the gain from promotion is likely related to the importance of human capital for performing a task. For example, promotion benefits are widely perceived to be high in professional services careers, such as consulting, law, and finance. Azmat et al. (2020) find that the differences in promotion rates between men and women in law firms are explained by men working more hours (i.e., exerting more effort) than women in entry-level positions. Such evidence is consistent with a discouragement effect in high-stakes careers. In contrast, Benson et al. (2024) find that women on management-track careers in retail have better pre-promotion performance than men. This finding is consistent with an overcompensation effect that dominates in low-stakes situations. Our model also predicts that, in competitive contexts, observable skills are more valuable to unfavored groups. Consistent with this prediction, Niessen-

Ruenzi and Zimmerer (2023) find that women's career outcomes are more sensitive to observable skills than men's.

Our model also explains why some firms invest in building a "progressive" corporate culture while others are content to maintain a "conservative" image. Subtle discrimination is detrimental to high-productivity firms because discriminated workers are discouraged from investing in valuable skills. Thus, such firms prefer to foster equality as a means to incentivize a diverse workforce. In contrast, low-productivity firms benefit from holding discriminated workers to higher standards, as these employees overcompensate by working harder. Consistent with our predictions, Edmans et al. (2023) find that employees' perception of diversity, equity and inclusion is stronger in growing, high-valuation, and financially strong firms. Similarly, a robust empirical finding is that, in the cross-section, large and high-performing firms have more women on their boards (see, e.g., Adams and Ferreira (2009)). Consistent with these cross-sectional correlations, Gao et al. (2023) find that firms experiencing a shock that lowers their cost of capital reduce their racial promotion gaps for mid- and high-skill workers. We are unaware of theoretical work explaining this body of evidence. Subtle discrimination can explain these findings.

References

ADAMS, R. B. AND D. FERREIRA (2009): "Women in the boardroom and their impact on governance and performance," *Journal of Financial Economics*, 94, 291–309.

ARROW, K. J. (1973): "The theory of discrimination," in *Discrimination in Labor Markets*, ed. by O. Ashenfelter and A. Rees, Cambridge MA: Princeton University Press, 3–33.

ATHEY, S., C. AVERY, AND P. ZEMSKY (2000): "Mentoring and diversity," *American Economic Review*, 90, 765–786.

AZMAT, G., V. CUÑAT, AND E. HENRY (2020): "Gender promotion gaps: Career aspirations and workplace discrimination," Working paper, Sciences Po.

Bartlett, R., A. Morse, R. Stanton, and N. Wallace (2022): “Consumer-lending discrimination in the FinTech era,” *Journal of Financial Economics*, 143, 30–56.

Bartoš, V., M. Bauer, J. Chytilová, and F. Matějka (2016): “Attention discrimination: Theory and field experiments with monitoring information acquisition,” *American Economic Review*, 106, 1437–75.

Becker, G. (1957): *The economics of discrimination*, Chicago: University of Chicago Press.

Becker, G. S. (1993): “Nobel lecture: The economic way of looking at behavior,” *Journal of Political Economy*, 101, 385–409.

Benson, A., D. Li, and K. Shue (2024): ““Potential” and the gender promotion gap,” Working paper, University of Minnesota.

Berk, J. B. and R. C. Green (2004): “Mutual fund flows and performance in rational markets,” *Journal of Political Economy*, 112, 1269–1295.

Bircan, Ç., G. Friebel, and T. Stahl (2023): “Knowledge teams, careers, and gender,” Working paper, Goethe-University Frankfurt.

Bohren, J. A., K. Haggag, A. Imas, and D. G. Pope (2023): “Inaccurate statistical discrimination: An identification problem,” *Review of Economics and Statistics*, 1–45.

Bordalo, P., K. Coffman, N. Gennaioli, and A. Shleifer (2016): “Stereotypes,” *The Quarterly Journal of Economics*, 131, 1753–1794.

Bosquet, C., P.-P. Combes, and C. García-Peñalosa (2019): “Gender and promotions: Evidence from academic economists in France,” *The Scandinavian Journal of Economics*, 121, 1020–1053.

Butler, A. W., E. J. Mayer, and J. P. Weston (2023): “Racial disparities in the auto loan market,” *The Review of Financial Studies*, 36, 1–41.

Cassel, J., J. Lerner, and E. Yimfor (2022): “Racial diversity in private capital fundraising,” Working Paper, National Bureau of Economic Research.

Ceccarelli, M., C. Herpfer, and S. Ongena (2024): “Gender, performance, and promotion in the labor market for commercial bankers,” Swiss Finance Institute Research Paper.

CHAUDHURI, S. AND R. SETHI (2008): "Statistical discrimination with peer effects: Can integration eliminate negative stereotypes?" *Review of Economic Studies*, 75, 579–596.

CHEREPANOV, V., T. FEDDERSEN, AND A. SANDRONI (2013): "Rationalization," *Theoretical Economics*, 8, 775–800.

CLERMONT, K. M. AND S. J. SCHWAB (2009): "Employment discrimination plaintiffs in federal court: From bad to worse," *Harv. L. & Pol'y Rev.*, 3, 103.

COATE, S. AND G. C. LOURY (1993): "Will affirmative-action policies eliminate negative stereotypes?" *The American Economic Review*, 1220–1240.

CORNELL, B. AND I. WELCH (1996): "Culture, information, and screening discrimination," *Journal of Political Economy*, 104, 542–571.

CUNNINGHAM, T. AND J. DE QUIDT (2022): "Implicit preferences inferred from choice," Working paper, Stockholm University.

DAVIES, S., E. VAN WESEP, AND B. WATERS (2021): "On the magnification of small biases in decision-making," Working paper, University of Colorado at Boulder.

DHANANI, L. Y., J. M. BEUS, AND D. L. JOSEPH (2018): "Workplace discrimination: A meta-analytic extension, critique, and future research agenda," *Personnel Psychology*, 71, 147–179.

DIPBOYE, R. L. AND S. K. HALVERSON (2004): *Subtle (and not so subtle) discrimination in organizations*, vol. 16 of *The dark side of organizational behavior*, 131–158.

DOBBIE, W., A. LIBERMAN, D. PARAVISINI, AND V. PATHANIA (2021): "Measuring bias in consumer lending," *The Review of Economic Studies*, 88, 2799–2832.

DOVIDIO, J. F. AND S. L. GAERTNER (1986): *Prejudice, discrimination, and racism.*, Academic Press.

DRUGOV, M. AND D. RYVKIN (2017): "Biased contests for symmetric players," *Games and Economic Behavior*, 103, 116–144.

EDGECLIFFE-JOHNSON, A. (2022): "The war on 'woke capitalism'," *Financial Times*.

EDMANS, A. (2020): *Grow the pie: How great companies deliver both purpose and profit–updated and revised*, Cambridge University Press.

EDMANS, A., C. FLAMMER, AND S. GLOSSNER (2023): "Diversity, equity, and inclusion," Working paper, National Bureau of Economic Research.

ELIAZ, K. AND R. SPIEGLER (2020): "A model of competing narratives," *American Economic Review*, 110, 3786–3816.

ESSED, P. (1991): *Understanding everyday racism: An interdisciplinary theory*, vol. 2, Sage.

EVANS, R. B. AND M. YOUNG (2022): "Nationalism, subtle bias, and labor outcomes: Evidence from global mutual funds," Working paper, University of Missouri.

EWENS, M. (2023): "Gender and race in entrepreneurial finance," *Handbook of the Economics of Corporate Finance: Private Equity and Entrepreneurial Finance*, 239.

EWENS, M. AND R. R. TOWNSEND (2020): "Are early stage investors biased against women?" *Journal of Financial Economics*, 135, 653–677.

FANG, H. AND A. MORO (2011): "Chapter 5 - Theories of statistical discrimination and affirmative action: A survey," in *Handbook of Social Economics*, ed. by J. Benhabib, A. Bisin, and M. O. Jackson, North-Holland, vol. 1, 133–200.

FERSHTMAN, D. AND A. PAVAN (2021): ""Soft" affirmative action and minority recruitment," *American Economic Review: Insights*, 3, 1–18.

FRAME, W. S., R. HUANG, E. JIANG, Y. LEE, W. LIU, E. J. MAYER, AND A. SUNDERAM (2023): "The impact of minority representation at mortgage lenders," *The Journal of Finance, Forthcoming*.

FRANKEL, A. (2021): "Selecting applicants," *Econometrica*, 89, 615–645.

FREDERIKSEN, A., L. B. KAHN, AND F. LANGE (2020): "Supervisors and performance management systems," *Journal of Political Economy*, 128, 2123–2187.

FRIEBEL, G. AND M. RAITH (2004): "Abuse of authority and hierarchical communication," *RAND Journal of Economics*, 224–244.

FRYER JR., R. G. (2007): "Belief flipping in a dynamic model of statistical discrimination," *Journal of Public Economics*, 91, 1151–1166.

GAERTNER, S. L. AND J. F. DOVIDIO (2005): "Understanding and addressing contemporary racism: From aversive racism to the common ingroup identity model," *Journal of Social Issues*, 61, 615–639.

GAO, J., W. MA, AND Q. XU (2023): "Access to financing and racial pay gap inside firm," Working paper, Georgetown University.

HEBERT, C. (2023): "Gender stereotypes and entrepreneur financing," Working Paper, University of Toronto.

HEBL, M., S. K. CHENG, AND L. C. NG (2020): "Modern discrimination in organizations," *Annual Review of Organizational Psychology and Organizational Behavior*, 7, 257–282.

HENGEL, E. (2022): "Are women held to higher standards? Evidence from peer review," *The Economic Journal*.

HOFFMAN, M., L. B. KAHN, AND D. LI (2018): "Discretion in hiring," *The Quarterly Journal of Economics*, 133, 765–800.

HONG, H. AND J. D. KUBIK (2003): "Analyzing the analysts: Career concerns and biased earnings forecasts," *The Journal of Finance*, 58, 313–351.

HOSPIDO, L., L. LAEVEN, AND A. LAMO (2019): "The gender promotion gap: Evidence from central banking," *Review of Economics and Statistics*, 1–45.

HUANG, R., J. S. LINCK, E. J. MAYER, AND C. A. PARSONS (2024): "Can human capital explain income-based disparities in financial services?" *The Review of Financial Studies*, hhae004.

HUANG, R., E. J. MAYER, AND D. P. MILLER (2023): "Gender bias in promotions: Evidence from financial institutions," *The Review of Financial Studies*, hhad079.

JONES, K. P., D. F. ARENA, C. L. NITTROUER, N. M. ALONSO, AND A. P. LINDSEY (2017): "Subtle discrimination in the workplace: A vicious cycle," *Industrial and Organizational Psychology*, 10, 51–76.

KAHNEMAN, D. (2011): *Thinking, fast and slow*, Macmillan.

KAWAMURA, K. AND I. M. DE BARREDA (2014): "Biasing selection contests with ex-ante identical agents," *Economics Letters*, 123, 240–243.

KISGEN, D. J., J. NICKERSON, M. OSBORN, AND J. REUTER (2020): "Analyst promotions within credit rating agencies: accuracy or bias?" *Journal of Financial and Quantitative Analysis*, 55, 869–896.

KUMAR, A. (2010): "Self-selection and the forecasting abilities of female equity analysts," *Journal of Accounting Research*, 48, 393–435.

KUMAR, A., A. NIESSEN-RUENZI, AND O. G. SPALT (2015): "What's in a name? Mutual fund flows when managers have foreign-sounding names," *The Review of Financial Studies*, 28, 2281–2321.

LANG, K. AND J.-Y. K. LEHMANN (2012): "Racial discrimination in the labor market: Theory and empirics," *Journal of Economic Literature*, 50, 959–1006.

LANG, K., M. MANOVE, AND W. T. DICKENS (2005): "Racial discrimination in labor markets with posted wage offers," *American Economic Review*, 95, 1327–1340.

LAZEAR, E., K. SHAW, AND C. STANTON (2018): "Who gets hired? The importance of competition among applicants," *Journal of Labor Economics*, 36, S133–S181.

LAZEAR, E. P. AND S. ROSEN (1990): "Male-female wage differentials in job ladders," *Journal of Labor Economics*, 8, S106–S123.

LETINA, I., S. LIU, AND N. NETZER (2020): "Delegating performance evaluation," *Theoretical Economics*, 15, 477–509.

LINOS, E., S. MOBASSERI, AND N. ROUSSILLE (2023): "Asymmetric peer effects at work: How White coworkers shape the careers of "people of color"," Working paper, Harvard.

LLORENS, A., A. TZOVARA, L. BELLIER, I. BHAYA-GROSSMAN, A. BIDET-CAULET, W. K. CHANG, Z. R. CROSS, R. DOMINGUEZ-FAUS, A. FLINKER, Y. FONKEN, ET AL. (2021): "Gender bias in academia: A lifetime problem that needs solutions," *Neuron*, 109, 2047–2074.

LUNDBERG, S. AND J. STEARNS (2019): "Women in economics: Stalled progress," *Journal of Economic Perspectives*, 33, 3–22.

MACLEOD, W. B. (2003): "Optimal contracting with subjective evaluation," *American Economic Review*, 93, 216–240.

MAILATH, G. J., L. SAMUELSON, AND A. SHAKED (2000): "Endogenous inequality in integrated labor markets with two-sided search," *American Economic Review*, 90, 46–72.

MANZINI, P. AND M. MARIOTTI (2012): "Choice by lexicographic semiorders," *Theoretical Economics*, 7, 1–23.

MORO, A. AND P. NORMAN (2004): "A general equilibrium model of statistical discrimination," *Journal of Economic Theory*, 114, 1–30.

NAVA, F. AND A. PRUMMER (2022): "Profitable Inequality," Working paper, Queen Mary University of London.

NIESSEN-RUENZI, A. AND S. RUENZI (2019): "Sex matters: Gender bias in the mutual fund industry," *Management Science*, 65, 3001–3025.

NIESSEN-RUENZI, A. AND L. ZIMMERER (2023): "The importance of signaling for women's careers," Working paper, University of Mannheim.

O'DONOGHUE, T. AND M. RABIN (1999): "Doing it now or later," *American Economic Review*, 89, 103–124.

ONUCHIC, P. (2023): "Recent contributions to theories of discrimination," Working paper, Oxford University.

ONUCHIC, P. AND D. RAY (2023): "Signaling and discrimination in collaborative projects," *American Economic Review*, 113, 210–252.

PAGANO, M. AND L. PICARIELLO (2022): "Corporate governance, favoritism and careers," Working paper, University of Naples Federico II.

PASTORINO, E. (2022): "Careers in firms: The role of learning and human capital," *Journal of Political Economy*, forthcoming.

PÉREZ-CASTRILLO, D. AND D. WETTSTEIN (2016): "Discrimination in a model of contests with incomplete information about ability," *International Economic Review*, 57, 881–914.

PHELPS, E. S. (1972): "The statistical theory of racism and sexism," *The American Economic Review*, 62, 659–661.

PRENDERGAST, C. (1993): "The role of promotion in inducing specific human capital acquisition," *The Quarterly Journal of Economics*, 108, 523–534.

PRENDERGAST, C. AND R. H. TOPEL (1996): "Favoritism in organizations," *Journal of Political Economy*, 104, 958–978.

RONCHI, M. AND N. SMITH (2021): "Daddy's girl: Daughters, managerial decisions, and gender inequality," Working paper, Bocconi University.

SARSONS, H., K. GËRXHANI, E. REUBEN, AND A. SCHRAM (2021): "Gender differences in recognition for group work," *Journal of Political Economy*, 129, 101–147.

SELMI, M. (2000): "Why are employment discrimination cases so hard to win," *La. L. Rev.*, 61, 555.

SHERMAN, M. G. AND H. E. TOOKES (2022): "Female representation in the academic finance profession," *The Journal of Finance*, 77, 317–365.

SINISCALCHI, M. AND P. VERONESI (2021): "Self-image bias and lost talent," Working paper, NBER.

TVERSKY, A. (1969): "Intransitivity of preferences," *Psychological Review*, 76, 31.

VAN LAER, K. AND M. JANSSENS (2011): "Ethnic minority professionals' experiences with subtle discrimination in the workplace," *Human Relations*, 64, 1203–1227.

WITTEMAN, H. O., M. HENDRICKS, S. STRAUS, AND C. TANNENBAUM (2019): "Are gender gaps due to evaluations of the applicant or the science? A natural experiment at a national funding agency," *The Lancet*, 393, 531–540.

Appendix

A Proofs

Proof of Proposition 1. Suppose first that both agents undertake strictly positive investments in skill acquisition in the first-best solution. The first-order conditions (FOCs) for an interior solution are $H(1-e_j)-c'(e_i)=0$, for $i \neq j \in \{b,r\}$. Under $c''(e_i)>0$, an interior solution must be unique, which implies that the solution is symmetric and given by $\tilde{e}$, where $\tilde{e}=1-\frac{c'(\tilde{e})}{H}$. Note that $\tilde{e}$ is well defined as long as $H>c'(0)$. We extend the definition of $\tilde{e}$ so that $\tilde{e}=0$ if $H \leq c'(0)$. We can then calculate the surplus associated with $\tilde{e}$: $\tilde{S} \equiv H\tilde{e}(2-\tilde{e})-2c(\tilde{e})$.

Consider now the case in which only one agent, say b, is requested to exert effort. If $H>c'(0)$, the optimal investment is given by $\hat{e}_b=\min\{c'^{-1}(H),1\}$. If $H \leq c'(0)$, we set $\hat{e}_b=0$. The surplus associated with $\hat{e}_b$ is $\hat{S} \equiv H\hat{e}_b - c(\hat{e}_b)$.

The first-best investment levels can take one of two forms. If $\tilde{S} \geq \hat{S}$, the gains from sharing effort are greater than the losses from effort duplication, in which case we have $e_b^{FB}=e_r^{FB}=\tilde{e}$. If, instead, $\tilde{S}<\hat{S}$, effort duplication is too costly, thus the first-best solution is $e_b^{FB}=\hat{e}_b$ and $e_r^{FB}=0$. $\square$

Proof of Proposition 2. Equations (5) and (6) represent the unique solution to the system of equations given by (4). From (5), we find that $e_b^* \leq 1$ requires $f_b(\sigma)=\beta\,(2\beta-1)\,\sigma^2-(0.5-\beta)\,\sigma+1 \geq 0$. Function f_b is strictly concave and has a unique positive root, $\overline{\sigma}(\beta) \equiv \frac{\beta-0.5+\sqrt{\frac{1}{4}+3\beta-7\beta^2}}{2\beta(1-2\beta)} \geq 0$, for all $\beta \in (0,0.5)$. Thus, $e_b^* \leq 1$ if and only if $\sigma \leq \overline{\sigma}(\beta)$.

To show that $\overline{\sigma}(\beta)>1$, note that

$$\beta-0.5+\sqrt{\frac{1}{4}+3\beta-7\beta^2}=\beta-0.5+\sqrt{(\beta-0.5)^2+4\beta\,(1-2\beta)}=-a+\sqrt{a^2+2b},$$

where $a=0.5-\beta>0$ and $b=2\beta(1-2\beta)>0$. Then, $-a+\sqrt{a^2+2b}>b \iff a^2+2b>a^2+2ab+b^2 \iff 2>b+2a \iff 2>2\beta-4\beta^2+1-2\beta \iff 1>-4\beta^2$.

Similarly, $e_r^* \leq 1$ requires $f_r(\sigma)=\beta\,(2\beta+1)\,\sigma^2-(0.5+\beta)\,\sigma+1 \geq 0$. Function f_r is strictly convex. If f_r has no real root, then trivially $e_r^*<1$ for any value of σ. A pair of real roots exists when $\beta \in \left(0,\frac{1}{14}\right]$. In this case, the smallest real root is: $\frac{0.5+\beta-\sqrt{\frac{1}{4}-3\beta-7\beta^2}}{2\beta(2\beta+1)}$, which is greater than 1; this is shown by setting $a=0.5+\beta$, $b=2\beta(1+2\beta)$ and verifying that $a-\sqrt{a^2-2b}>b$.

Note also that $f_r(\sigma)-f_b(\sigma)=2\beta\sigma(\sigma-1)$, which is positive if $\sigma>1$. Thus, if $\sigma \leq \overline{\sigma}(\beta)$, then both e_b^* and e_r^* are interior. If $\sigma>\overline{\sigma}(\beta)$, then we must have $e_b^*=1$, which implies $e_r^*=$

$\min\left\{\frac{\sigma(1-2\beta)}{2},1\right\}$. Notice that if $\beta=0.5$, then $\overline{\sigma}\to\infty$, and the solution is interior for any σ. $\square$

Proof of Proposition 3. The equilibrium promotion gap is

$$\Delta p(\sigma)=2\beta\frac{1+2\beta^2(1+4\beta^2)\sigma^4+(\frac{3}{2}+2\beta^2)\sigma^2-2\sigma}{(1+4\beta^2\sigma^2)^2}.$$

Its derivative with respect to σ is

$$\frac{\partial\Delta p}{\partial\sigma}=2\beta\frac{-2+3\left(1-4\beta^2\right)\sigma+24\beta^2\sigma^2+4\beta^2\left(4\beta^2-1\right)\sigma^3}{(1+4\beta^2\sigma^2)^3}.$$

Define the function $A(\sigma)$ as the numerator of the expression above. $A(\sigma)$ is a third-degree polynomial of σ, thus, for $\sigma\in\mathbb{R}$, it has three (real or complex) roots (r_1,r_2,r_3), a local minimum, and a local maximum. Consider its first derivative: $A'(\sigma)=3(1-4\beta^2)+48\beta^2\sigma+12\beta^2\left(4\beta^2-1\right)\sigma^2$. The roots for $A'(\sigma)=0$ are $\sigma^m=\frac{4\beta-\sqrt{16\beta^2+(1-4\beta^2)^2}}{2\beta(1-4\beta^2)}$ and $\sigma^M=\frac{4\beta+\sqrt{16\beta^2+(1-4\beta^2)^2}}{2\beta(1-4\beta^2)}$. Notice that $\sigma^m<0$ and $\sigma^M>0$. At $\sigma=0$, we have $A(0)=-2<0$ and $A'(0)=3(1-4\beta^2)>0$. Thus, $A(\sigma^m)$ must be a local minimum and $A\left(\sigma^M\right)$ a local maximum. Thus, $A(\sigma)$ has one negative real root ($r_1<\sigma^m$), while $r_2\leq r_3$ must be positive if they are real numbers.

Notice that at $\sigma=1$, the solution is interior, and we have $A(1)=1+8\beta^2+16\beta^4>0$. That is, $\frac{\partial\Delta p}{\partial\sigma}$ is strictly positive at $\sigma=1$. Thus, a real root $r_2\in(0,1)$ must exist; $r_3>r_2$ must also be a real number. We then have that $\frac{\partial\Delta p}{\partial\sigma}<0$ for $\sigma\in(0,r_2)$, $\frac{\partial\Delta p}{\partial\sigma}>0$ for $\sigma\in(r_2,r_3)$, and $\frac{\partial\Delta p}{\partial\sigma}<0$ for $\sigma>r_3$. Because σ^M is a local maximum, $\sigma^M<r_3$. Brute force comparison reveals that $\sigma^M>\overline{\sigma}$ for all $\beta\in(0,0.5]$. Thus, σ^M cannot be an interior solution $\Rightarrow r_3>\sigma^M$. Thus, in an interior solution, $\frac{\partial\Delta p}{\partial\sigma}<0$ for $\sigma<r_1$ and $\frac{\partial\Delta p}{\partial\sigma}>0$ for $\sigma>r_2$. We thus have that $\Delta p(\sigma)$ reaches a minimum at $\min\{r_2,\overline{\sigma}\}\equiv\widetilde{\sigma}$. $\square$

Proof of Proposition 4. If $\beta=0$, we have that, in an interior solution, $e_r=e_b=\frac{\sigma}{2}$. The principal's problem is $\max_{\sigma\in[0,\overline{\sigma}(0)]}\theta\left[1-\left(1-\frac{\sigma}{2}\right)^2\right]-\sigma$. The FOC for an interior solution is $\theta\left(1-\frac{\sigma}{2}\right)-1=0$. The SOC holds (the problem is globally concave): $-\frac{\theta}{2}<0$. Thus, we have $\sigma^*=2\frac{\theta-1}{\theta}, e^*=\frac{\theta-1}{\theta}$. Notice that for all $\theta\geq1$, the solution is interior, and for all $\theta<1$ the principal does not operate the firm. $\square$

Proof of Proposition 5. Notice that the firm can guarantee a non-negative profit by choosing $\sigma = 0$. Because the objective function is continuous in σ and $[0, \overline{\sigma}(\beta)]$ is a compact set, a maximum always exists. An optimal σ^* is generically unique because the objective function is a function of polynomials and thus has no flat regions in the interior of $[0, \overline{\sigma}(\beta)]$. The uniqueness here is generic; multiple solutions may arise for measure-zero combinations of parameters (k, β, θ). $\square$

Proof of Proposition 6. Define $\sigma(\beta, \theta) \equiv \arg\max_{\sigma \in [0, \overline{\sigma}(\beta)]} \theta f(\sigma, \beta) - \sigma$, where $f(\sigma, \beta) = e_b(\sigma, \beta) + e_r(\sigma, \beta) - e_b(\sigma, \beta) e_r(\sigma, \beta)$, where $e_b(\sigma, \beta)$ and $e_r(\sigma, \beta)$ are given by (5) and (6), respectively. From Proposition 5, the optimal σ is generically unique, thus $\sigma(\beta, \theta)$ is well-defined (except perhaps for a measure-zero combination of parameters (β, θ)). The maximum profit is thus defined as $\Pi(\beta, \theta) \equiv \theta f(\sigma(\beta, \theta), \beta) - \sigma(\beta, \theta)$.

First notice that, for $\sigma(\beta, \theta) > 0$, we have that the optimal profit strictly increases with θ (by the Envelope Theorem):

$$\frac{\partial \Pi}{\partial \theta} = f(\sigma(\beta, \theta), \beta) > 0. \tag{17}$$

To prove Part 1, notice first that at $\theta = 0$, trivially, $\sigma(\beta, 0) = 0$ and the profit is zero. For $\theta = \overline{\sigma}(\beta) + \varepsilon$ (when $\beta < 0.5$), where $\varepsilon > 0$, if the principal chooses $\sigma = \overline{\sigma}(\beta)$ we have $f(\overline{\sigma}(\beta), \beta) = 1$ and the profit is strictly positive. Thus, we know that there exists $\underline{\theta}(\beta)$ such that $\sigma(\beta, \theta) > 0$ (and the profit is strictly positive) if and only if $\theta > \underline{\theta}(\beta)$. Now define $\overline{\theta}(\beta)$ as $\overline{\theta}(\beta) \equiv \frac{1}{f_\sigma(\overline{\sigma}(\beta), \beta)}$, where f_σ denotes the derivative with respect to σ (note that $f(\sigma, \beta)$ is differentiable in σ in the interior of $[0, \overline{\sigma}(\beta)]$). We then have $\sigma\left(\beta, \overline{\theta}(\beta) + \varepsilon\right) = \overline{\sigma}(\beta)$ for all $\varepsilon \geq 0$. This proves Part 1.

To prove Part 2, consider $\theta \in \left(\underline{\theta}(\beta), \overline{\theta}(\beta)\right)$, that is, the values for θ such that $\sigma(\beta, \theta)$ is interior, i.e., $\sigma(\beta, \theta) \in (0, \overline{\sigma}(\beta))$. Thus, the FOC at $\sigma^* = \sigma(\beta, \theta)$ must hold: $\frac{\partial \Pi}{\partial \sigma} = \theta f_\sigma(\sigma^*, \beta) - 1 = 0$, as well as the second order condition: $\frac{\partial^2 \Pi}{\partial \sigma^2} = \theta f_{\sigma\sigma}(\sigma^*, \beta) < 0$. We have that (by implicit differentiation of the FOC): $\frac{\partial \sigma}{\partial \theta} = -\frac{f_\sigma(\sigma^*, \beta)}{\theta f_{\sigma\sigma}(\sigma^*, \beta)} = -\frac{1}{\theta^2 f_{\sigma\sigma}(\sigma^*, \beta)} > 0$, proving Part 2. Part 3 follows from (17). $\square$

Proof of Proposition 7. For $e_b^* < 1$ (i.e., a strictly interior solution for effort levels), define $f(\sigma, \beta)$

as

$$f(\sigma,\beta) = e_b^* + e_r^* - e_b^* e_r^* = \frac{\sigma + 4\beta^2\sigma^2}{1+4\beta^2\sigma^2} - \frac{4\beta^2\sigma^3 + (1-4\beta^2\sigma^2)(\frac{1}{4}-\beta^2)\sigma^2}{(1+4\beta^2\sigma^2)^2}.$$

If $e_b^* < 1$ we can write the profit as $\Pi(\sigma,\beta,\theta) = \theta f(\sigma,\beta) - \sigma$. We then have

$$\frac{\partial \Pi}{\partial \beta} = -\frac{2\beta\theta\sigma^2(\sigma-1)\left\{4\sigma^2(\sigma+1)\beta^2 - 3\sigma + 5\right\}}{(4\sigma^2\beta^2+1)^3},$$

which has non-negative roots at $\beta = 0$ and $\beta_{root}(\sigma) = \frac{1}{2\sigma}\sqrt{\frac{3\sigma-5}{\sigma+1}}$. Note that for $\sigma < 1$, $\frac{\partial \Pi}{\partial \beta}$ is strictly positive for $\beta > 0$, implying that the optimal bias is $\beta = 0.5$. At $\sigma \in (1, \frac{5}{3})$, the derivative is strictly negative for $\beta > 0$, implying that the optimal bias is $\beta = 0$. For $\sigma > 5/3$, $\frac{\partial \Pi}{\partial \beta}$ is positive for $\beta < \beta_{root}(\sigma)$ and negative for $\beta > \beta_{root}(\sigma)$, implying that the optimal bias is $\beta_{root}(\sigma)$. Define the following:

$$\sigma(\beta,\theta) \equiv \arg \max_{\sigma \in [0,\overline{\sigma}(\beta)]} \Pi(\sigma,\beta,\theta),$$

$$\beta(\sigma,\theta) \equiv \arg \max_{\beta \in [0,0.5]} \Pi(\sigma,\beta,\theta), \text{ and}$$

$$\sigma(\theta) \equiv \arg \max_{\sigma \in [0,\overline{\sigma}(\beta(\sigma,\theta))]} \Pi(\sigma,\beta(\sigma,\theta),\theta).$$

For now we assume that θ is such that $\sigma(\theta) < 5/3$, so that the optimal profit is

$$\Pi(\theta) = \max\{\Pi(\sigma(0,\theta),0,\theta), \Pi(\sigma(0.5,\theta),0.5,\theta)\}.$$

Define $\Delta(\theta) \equiv \Pi(\sigma(0,\theta),0,\theta) - \Pi(\sigma(0.5,\theta),0.5,\theta)$, and let θ' denote an element of $\{\theta : \Delta(\theta) = 0\}$. We know that at least one such θ' exists because: (i) $\Pi(\sigma(0.5,1),0.5,1) \geq \Pi(0.5,0.5,1) = 0.02 > \Pi(\sigma(0,1),0,1) = 0$ (see Proposition 4) and (ii) $\Pi(\sigma(0.5,4),0.5,4) = 2 < \Pi(\sigma(0,4),0,4) = 2.25$.

By continuity there must be a $\theta' \in (1,4)$ (numerically, we obtain that $\theta' \approx 2.62054$) such that $\Pi(\sigma(0.5,\theta'),0.5,\theta') = \Pi(\sigma(0,\theta'),0,\theta')$. We need to show that θ' is unique. By the Envelope Theorem,

$$\frac{\partial \Delta(\theta)}{\partial \theta} = f(\sigma(0,\theta),0) - f(\sigma(0.5,\theta),0.5).$$

If $\frac{\partial \Delta(\theta')}{\partial \theta} > 0$ for all $\theta' \in \{\theta : \Delta(\theta) = 0\}$, then θ' is unique. To show that this is indeed the case, note first that at θ', it must be that $\sigma(0.5, \theta') \leq 1$, otherwise $\frac{\partial \Pi}{\partial \beta} < 0$ and thus $\Pi(\sigma(0, \theta'), 0, \theta') - \Pi(\sigma(0.5, \theta'), 0.5, \theta') > 0$. Similar reasoning implies that $\sigma(0, \theta') \geq 1$. We then have that $\Delta(\theta') = 0$ implies $f(\sigma(0, \theta'), 0) - f(\sigma(0.5, \theta'), 0.5) = \frac{\sigma(0,\theta') - \sigma(0.5,\theta')}{\theta'} > 0$. Thus, θ' is unique. Notice that $\theta' < \overline{\theta}$. If not, at $\overline{\theta}$ we have $\Delta(\overline{\theta}) < 0$, i.e., the optimal bias is $\beta = 0.5$. From Proposition 2, $\sigma(0.5, \overline{\theta}) > 1$. But then $\frac{\partial \Pi}{\partial \beta}$ is strictly negative for $\beta > 0$, thus the optimal bias cannot be $\beta = 0.5$.

Consider now values for θ such that $\sigma(\theta) \geq 5/3$. In any strictly interior solution for e_b^*, we have $f(\sigma, \beta_{root}(\sigma)) = \frac{\sigma^2 + 2\sigma + 25}{32}$, and thus $\Pi(\sigma, \beta_{root}(\sigma), \theta) = \theta \frac{\sigma^2 + 2\sigma + 25}{32} - \sigma$ and

$$\frac{\partial \Pi(\sigma, \beta_{root}(\sigma), \theta)}{\partial \sigma} = \theta \frac{2\sigma + 2}{32} - 1,$$

implying that $\Pi(\sigma, \beta_{root}(\sigma), \theta)$ has a global minimum at $\sigma = 16 - \theta$. At any $\sigma \geq \frac{5}{3}$ with $e_b^* < 1$, the principal prefers either to increase or decrease σ. Thus, there is no strictly interior solution in which $\sigma(\theta) \geq \frac{5}{3}$. That is, $\sigma(\overline{\theta}) < \frac{5}{3}$. $\square$

Internet Appendix
Subtle Discrimination

Elena S. Pikulina   Daniel Ferreira

OA.1 Proof of corollaries

Proof of Corollary 1. Note that, from (5) and (6), $e_r^* = e_b^* \frac{(0.5+\beta)-2\beta\sigma(0.5-\beta)}{(0.5-\beta)+2\beta\sigma(0.5+\beta)}$. It follows that $e_r^* \geq e_b^*$ if and only if $\sigma \leq 1$. □

Proof of Corollary 2. The achievement gap is

$$e_b^* - e_r^* = \frac{0.5(\sigma_b - \sigma_r) - \beta(\sigma_b + \sigma_r) + 2\beta\sigma_b\sigma_r}{1 + 4\beta^2\sigma_b\sigma_r}$$

Define

$$n(W) = 0.5W(\frac{1}{k_b} - \frac{1}{k_r}) - \beta W(\frac{1}{k_b} + \frac{1}{k_r}) + 2W^2\beta\frac{1}{k_bk_r}$$

$$d(W) = 1 + 4W^2\beta^2\frac{1}{k_bk_r}$$

Note that $e_b^* - e_r^* = \frac{n(W)}{d(W)} = 0$ has at most two roots, one of them being $W_1 = 0$. Then we have

$$\frac{\partial(e_b^* - e_r^*)}{\partial W} = \frac{n'(W)d(w) - d'(W)n(W)}{d(W)^2}.$$

At $W = W_1 = 0$, we have $e_b^* - e_r^* = 0$ and

$$\frac{\partial(e_b^* - e_r^*)}{\partial W}(W = 0) = \frac{0.5 - \beta}{k_b} - \frac{0.5 + \beta}{k_r}.$$

Thus, we have that $e_b^* - e_r^*$ is decreasing at $W = 0$ if and only if

$$\frac{k_r}{k_b} < \frac{0.5 + \beta}{0.5 - \beta}.$$

The second root, W_2, is given by

$$\frac{0.5 - \beta}{k_b} - \frac{(0.5 + \beta)}{k_r} + \frac{2W\beta}{k_bk_r} = 0,$$

implying a unique solution

$$W_2 = \frac{k_b(\beta + 0.5) - k_r(0.5 - \beta)}{2\beta}$$

Thus, $W_2 > 0$ if and only if $e_b^* - e_r^*$ is decreasing at $W = 0$. Define $\widehat{W} \equiv \max\{W_2, 0\}$. Then, we have that $e_b^* - e_r^* \leq 0$ iff $W \leq \widehat{W}$, proving the result. $\square$

Proof of Corollary 3. Solving (7) yields

$$e_b^* = \frac{\sigma(0.5-\beta) + (2\beta-\delta)\sigma^2(0.5+\beta-\delta)}{1+(2\beta-\delta)^2\sigma^2}, \tag{OA.1}$$

$$e_r^* = \frac{\sigma(0.5+\beta-\delta) - (2\beta-\delta)\sigma^2(0.5-\beta)}{1+(2\beta-\delta)^2\sigma^2}, \tag{OA.2}$$

and $e_r^* = e_b^* \frac{(0.5+\beta-\delta)-(2\beta-\delta)\sigma(0.5-\beta)}{(0.5-\beta)+(2\beta-\delta)\sigma(0.5+\beta-\delta)}$. It follows that $e_r^* \geq e_b^*$ if and only if $\sigma(1-\delta) \leq 1$. $\square$

Proof of Corollary 4. 1. Differentiating (5) with respect to σ yields

$$\frac{\partial e_b^*}{\partial \sigma} = \frac{0.5-\beta+4\beta\sigma\left[(0.5+\beta)-\beta\sigma(0.5-\beta)\right]}{\left(1+4\beta^2\sigma^2\right)^2} > 0,$$

because $e_r^* \geq 0$ implies $0.5+\beta-2\beta\sigma(0.5-\beta) \geq 0 \Rightarrow 0.5+\beta-\beta\sigma(0.5-\beta) > 0$.
2. Differentiating (6) with respect to σ yields

$$\frac{\partial e_r^*}{\partial \sigma} = \frac{(0.5+\beta)-4\beta\sigma\left[(0.5-\beta)+\beta\sigma\,(0.5+\beta)\right]}{\left(1+4\beta^2\sigma^2\right)^2}.$$

Note that $\frac{\partial e_r^*}{\partial \sigma} > 0$ for $\sigma = 0$ and the numerator is strictly decreasing in σ (with limit at $-\infty$). Solving for the unique positive root for the numerator yields

$$\sigma_{root}(\beta) \equiv \frac{\sqrt{(0.5-\beta)^2+(0.5+\beta)^2}-(0.5-\beta)}{2\beta\,(0.5+\beta)} > 0.$$

We then define $\widehat{\sigma}(\beta) \equiv \min\{\sigma_{root}(\beta), \overline{\sigma}(\beta)\}$.

3.

$$\frac{\partial \sigma_{root}(\beta)}{\partial \beta} = \frac{\left(1+2\beta\left(\frac{1}{2}+2\beta^2\right)^{-\frac{1}{2}}\right)\left(\beta+2\beta^2\right)-(1+4\beta)\left[\left(\frac{1}{2}+2\beta^2\right)^{\frac{1}{2}}-(0.5-\beta)\right]}{\left(\beta+2\beta^2\right)^2}.$$

The numerator is negative for $\beta = 0$ and decreasing in β for $\beta \in (0, 0.5]$. Thus, we have $\frac{\partial \sigma_{root}}{\partial \beta} < 0$, that is, the region in which e_r^* declines starts earlier for larger values of β. $\square$

OA.2 Market equilibrium under perfect competition

In this section, we show the robustness of the main results to different assumptions. The modification presented here has the following important features: (i) Firms compete for workers and, thus, have no bargaining power; (ii) Workers invest in general skills before being hired and their skills are observable to all; (iii) Compensation may depend on skills. We consider a stylized labor market with search frictions and asymmetric employer learning. Despite such frictions, in the absence of a subtle bias β, the first-best investment levels are achieved. We then show that a positive subtle bias distorts investment decisions in a similar way as in the main model. Unlike the main model, firms have zero profits in equilibrium regardless of their biases. Thus, they have no incentives to change their biases, i.e., market forces do not eliminate biases.

Consider an economy with mass L of agents, half of them blue and the other half red. Each agent lives for two periods (we can think of overlapping generations that enter the market in each period with mass L). There is a mass F of firms with infinite lifespans. Each firm is endowed with a technology that allows them to employ two workers in each period. Workers can be skilled, $s_i = 1$, or unskilled, $s_i = 0$. Skill is observed at the time of hiring and is general in the sense that it is valuable to all companies.

After an agent works for a firm for one period, the firm learns about her managerial ability, $m_i \in \{0,1\}$. As in asymmetric employer learning models (see, e.g., Waldman (1984) and Greenwald (1986)), this information is private to the firm. After learning the managerial abilities of its two workers, the firm decides which worker (if any) to promote to the top position, job 2. We assume that promoting an agent with no managerial ability leads to large losses. Thus, firms never promote a worker unless they are sure that $m_i = 1$. Non-promoted agents remain at job 1 or leave the firm (for simplicity, we assume that the productivity of an "old" worker is zero at job 1). Formally, with probability μ, agent i generates a signal $m_i = 1$, indicating suitability for managerial positions. Among those with $m_i = 1$, agents can be skilled ($s_i = 1$) or unskilled ($s_i = 0$). Promoting an unskilled agent increases the principal's expected payoff by $l > 0$ while promoting a skilled agent increases the payoff by $l + H$, where $H > 0$. That is, a skilled agent is always more productive than an unskilled one once assigned to job 2. Note that the main model in the paper is a special case of this one in which $\mu = 1$.

The timing is as follows. Within Period 1 (when agents are young), there are three dates. At Date 0, young agents first decide how much to invest in skills, $e_i \in [0,1]$ at cost $c(e_i) = \frac{e_i^2}{2}$ (i.e., we set $k = 1$ for simplicity). As in the main model, e_i is the probability of acquiring the skill.

At Date 1, each firm searches for workers to fill its two job 1 vacancies. Search is costly in the sense that each firm can find only two workers per period. Each worker can be matched with multiple firms at this stage; workers join firms that offer them the best conditions. Workers cannot observe the identities of agents who hold job offers. We assume that $2F > L$, that is, there are more vacancies than workers. Random search implies that each worker receives more than one offer to choose from. Firms offer an entry-level wage $w_1(s_i)$ to a candidate with skill s_i, conditional on both candidates accepting the offer. At this stage, there is no need for firms to promise a wage upon promotion in Period 2; this wage is determined later in the labor market. Firms must hire two workers, otherwise, they cannot produce.

At Date 2, after m_i is internally revealed to each firm, firms make their promotion decisions. Promotion decisions are visible, thus firms with vacancies can make poaching offers $w_p(s_i)$ to promoted workers from other firms. To retain such workers, incumbent firms must match any poaching offer.[1] Non-promoted workers are never poached because they might have no managerial abilities.

Finally, in Period 2, promoted agents work at job 2 and receive wages $w_p(s_i)$ and non-promoted workers stay at job 1 and continue receiving wage $w_1(s_i)$. For simplicity, we assume no discounting between periods.

We characterize the equilibrium by working backwards. At the end of Date 2, mass $\frac{(1-\mu)^2 L}{2}$ of active firms have vacancies because none of their workers has managerial ability. Such firms try to poach promoted workers from other firms. Thus, workers with managerial abilities are in short supply; their wages must be $w_p(s_i = 1) = l + H$ and $w_p(s_i = 0) = l$. This implies that all active firms make zero profit when employing old workers in job 2. We assume that firms do not overtly discriminate and, thus, always promote skilled agents ahead of unskilled agents.

At Date 1, firms anticipate they will make zero profits from either type of worker and are thus indifferent between $s_i = 0$ and $s_i = 1$ workers. This implies that they are also indifferent between blue and red workers. Firms then offer $w_1(s_i) = 0$ to all workers as the base wage.

At Date 0, workers make investment decisions knowing that (i) wages upon promotion are $w_p(s_i = 1) = l + H$ and $w_p(s_i = 0) = l$, (ii) the probability of being paired with either blue or red agents is 0.5, and (iii) firms have a subtle bias β towards blue agents (or, equivalently, 2β is the proportion of firms with subtle bias 0.5).

In a rational expectations equilibrium, agents maximize utility taking the equilibrium effort of all other blue and red agents as given, e_b^* and e_r^*, here conjectured to be the same for all agents of

[1] For a model of poaching along these lines, see Ferreira and Nikolowa (2023).

the same type. A blue agent's expected utility is:

$$\begin{aligned}
&\mu(1-\mu)(l{+}e_bH)+\\
&\frac{\mu^2}{2}\left[(l+H)e_b(1-e_r^*)+(l+H)\left(\frac{1}{2}+\beta\right)e_be_r^*+l\left(\frac{1}{2}+\beta\right)(1-e_b)(1-e_r^*)\right]+\\
&\frac{\mu^2}{2}\left[(l+H)e_b(1-e_b^*)+(l+H)\frac{1}{2}e_be_b^*+l\frac{1}{2}(1-e_b)(1-e_b^*)\right]-\frac{e_b^2}{2},
\end{aligned}\tag{OA.3}$$

and a red agent's expected utility is:

$$\begin{aligned}
&\mu(1-\mu)(l{+}e_rH)+\\
&\frac{\mu^2}{2}\left[(l+H)e_r(1-e_b^*)+(l+H)\left(\frac{1}{2}-\beta\right)e_re_b^*+l\left(\frac{1}{2}-\beta\right)(1-e_r)(1-e_b^*)\right]+\\
&\frac{\mu^2}{2}\left[(l+H)e_r(1-e_r^*)+(l+H)\frac{1}{2}e_re_r^*+l\frac{1}{2}(1-e_r)(1-e_r^*)\right]-\frac{e_r^2}{2}.
\end{aligned}\tag{OA.4}$$

Note that if the premium is the same for all skill levels ($H=0$), $\mu=1$, and workers pair with a different type with probability one, we are back in the case of the main model.

The first-order conditions are:

$$\begin{aligned}
\mu(1-\mu)H+\frac{\mu^2}{2}\left[2(l+H)-l(1+\beta)+e_r^*\left(l(\frac{1}{2}+\beta)-(l+H)(\frac{1}{2}-\beta)\right)-H\frac{e_b^*}{2}\right]-e_b=0,\\
\mu(1-\mu)H+\frac{\mu^2}{2}\left[2(l+H)-l(1-\beta)+e_b^*\left(l(\frac{1}{2}-\beta)-(l+H)(\frac{1}{2}+\beta)\right)-H\frac{e_r^*}{2}\right]-e_r=0.
\end{aligned}$$

Note that while the reaction function of red agents is always downward-sloping, the reaction function of blue agents is upward-sloping if and only if $2l\beta>H(0.5-\beta)$. This is always true for sufficiently high β. Thus, a large subtle bias is sufficient for the blue agents' reaction functions to slope upwards.

We can rewrite the equilibrium conditions as

$$\sigma_l(1-\beta+2\beta e_r^*)+\sigma_h\left[2-(0.5-\beta)\mu e_r^*\right]=e_b^* \tag{OA.5}$$

$$\sigma_l(1+\beta-2\beta e_b^*)+\sigma_h\left[2-(0.5+\beta)\mu e_b^*\right]=e_r^*, \tag{OA.6}$$

where $\sigma_l = \frac{2\mu^2 l}{4+\mu^2 H}$ and $\sigma_h = \frac{2\mu H}{4+\mu^2 H}$. The unique solution (if interior) is given by

$$e_b^* = \frac{\sigma_l(1-\beta)+2\sigma_h+[2\beta\sigma_l-\mu(0.5-\beta)\sigma_h][\sigma_l(1+\beta)+2\sigma_h]}{1+[2\beta\sigma_l-\mu(0.5-\beta)\sigma_h][2\beta\sigma_l+\mu(0.5+\beta)\sigma_h]}; \tag{OA.7}$$

$$e_r^* = \frac{\sigma_l(1+\beta)+2\sigma_h-[2\beta\sigma_l+\mu(0.5+\beta)\sigma_h][\sigma_l(1-\beta)+2\sigma_h]}{1+[2\beta\sigma_l-\mu(0.5-\beta)\sigma_h][2\beta\sigma_l+\mu(0.5+\beta)\sigma_h]}. \tag{OA.8}$$

We have that $e_r^* \geq e_b^*$ if and only if $2\sigma_l(1-2\sigma_l) \geq [(8+\mu)\sigma_l+4\mu\sigma_h]\sigma_h$. To interpret this condition, note that, in this version of the model, we have two measures for the stakes: l and H, which are proportional to σ_l and σ_h. If $\sigma_l > 0.5$, the condition for $e_r^* \geq e_b^*$ does not hold. If $\sigma_l < 0.5$, the condition holds only if σ_h is sufficiently small. Thus, as in the main model, the overcompensation effect dominates for low stakes (low σ_l and σ_h), while the discouragement effect dominates for high stakes (high σ_l and/or high σ_h).

How do we know that there exist parameters that lead to either case? Notice that if $H \to 0$, the unique (interior) equilibrium converges to

$$e_b^* = \frac{\sigma_l(1-\beta)+2\beta\sigma_l^2(1+\beta)}{1+4\beta^2\sigma_l^2} \quad \text{and} \quad e_r^* = \frac{\sigma_l(1+\beta)-2\beta\sigma_l^2(1-\beta)}{1+4\beta^2\sigma_l^2}, \tag{OA.9}$$

with the condition for Red to invest more becoming simply $\sigma_l \leq 0.5$, which is equivalent to $\mu^2 l \leq 1$. If $\mu = 1$ as in the main model, the condition is thus identical to Corollary 1, because σ in this context is identical to $l = W/k$.

OA.3 Skills not fully firm-specific

Here we show that the skill does not have to be fully firm-specific for our results to go through. Suppose that by acquiring the skill an agent improves her outside wage by $w_0 > 0$, such that $W > w_0$, i.e., the skill is still more valuable to the worker inside the firm than outside. Then, the equilibrium investment levels (assuming an interior solution) are

$$e_b^* = \frac{\sigma(0.5-\beta)+2\beta\sigma^2(0.5+\beta)+\sigma\sigma_0(0.25-\beta^2)}{1+4\beta^2\sigma(\sigma-\sigma_0)-\sigma_0^2(0.25-\beta^2)}; \tag{OA.10}$$

$$e_r^* = \frac{\sigma(0.5+\beta)-2\beta\sigma^2(0.5-\beta)+\sigma\sigma_0(0.25-\beta^2)}{1+4\beta^2\sigma(\sigma-\sigma_0)-\sigma_0^2(0.25-\beta^2)}, \tag{OA.11}$$

where $\sigma_0 = w_0/k$. Note that Corollary 1 is unchanged: $e_r^* \geq e_b^*$ if and only if $\sigma \leq 1$.

Even if the skill is fully general, it could be more valuable to the worker inside the firm if there are switching (or search) costs. Alternatively, in the presence of other realistic frictions, skill specificity does not matter, as we showed in Section OA.2.

OA.4 Continuous skill levels

Our model is not limited to scenarios where skill levels are discrete and workers can be equally qualified with positive probability. Subtle discrimination may occur whenever a decision-maker can credibly claim to use private information for choosing between candidates. This holds true even when observable differences in skills are large, as long as the decision-maker can plausibly deny being biased.

Suppose agents can choose to acquire a skill level in the unit interval, $s_i \in [0,1]$. For simplicity, we assume that $e_i = s_i$, i.e., effort deterministically translates into skill. Thus, the difference in observable skill is $\Delta s = e_r - e_b$. As in Section 3, let $F(.)$ denote the c.d.f. of Δx. For simplicity, assume that $\underline{x} = -1$, $\bar{x} = 1$, $F(.)$ is uniform, and $\omega = 1$. That is, for *any* differences in skill Δs, the principal can plausibly justify promoting Blue. Thus, subtle discrimination may occur for any Δs.[2]

Consider first how an unbiased decision-maker would choose between the two candidates. For given e_b and e_r, the probability that Blue is chosen is given by $P(e_b, e_r) = \frac{1}{2}(1 + e_b - e_r)$. In the contest literature, $P(e_b, e_r)$ is known as the *contest success function*, or CSF. Below we show that the key to our results is how subtle biases affect the CSF.

Under the unbiased CSF, the agents' expected utilities are (assuming $k = 2$ for simplicity) $U_b = \sigma \frac{1}{2}(1 + e_b - e_r) - e_b^2$ and $U_r = \sigma\left[1 - \frac{1}{2}(1 + e_b - e_r)\right] - e_r^2$, leading to first-order conditions $\frac{\sigma}{2} - e_b^* = 0$ and $\frac{\sigma}{2} - e_r^* = 0$, i.e., $e_b^* = e_r^* = \frac{\sigma}{2}$, as in our main model.

Suppose now the principal is subtly biased. Then, we can write the CSF as $P(e_b, e_r) = \min\left\{\frac{1}{2}(1 + e_b - e_r) + b(e_b, e_r), 1\right\}$, where $b(e_b, e_r)$ is the bias in favor of Blue. The bias $b(e_b, e_r)$ may depend on the observed skill levels because the cost to the principal from making a biased decision may depend on the skill levels. How subtle discrimination distorts effort levels thus depends on the particular functional form of $b(e_b, e_r)$.

Consider first the case in which $b(e_b, e_r) = \beta e_b$, for $\beta \in (0, 0.5]$. This may be the case when the principal finds it easier to justify choosing Blue when Blue is more objectively qualified. Assuming an interior solution, the first-order conditions are $\sigma\left(\frac{1}{2} + \beta\right) - e_b^* = 0$ and $\frac{\sigma}{2} - e_r^* = 0$, which implies $e_b^* > e_r^*$. (An interior solution always obtains for $\sigma < \frac{2}{1+2\beta}$). Thus, subtle discrimination produces an *encouragement effect* for Blue. If $\sigma > \frac{2}{1+2\beta}$, the equilibrium is $e_b^* = 1$ and $e_r^* = 0$. That is, in high-stakes situations, subtle discrimination leads to a discouragement effect for Red.

Suppose now that $b(e_b, e_r) = \beta(1 - e_r)$, for $\beta \in (0, 0.5]$. Intuitively, this is the case when the

[2] This is obviously not necessary for the results, but it helps make the point that there is nothing special about the "ties" in our main model.

principal finds it easier to justify choosing Blue when Red is less objectively qualified. Assuming an interior solution, the first-order conditions are $\frac{\sigma}{2} - e_b^* = 0$ and $\sigma\left(\frac{1}{2} + \beta\right) - e_r^* = 0$, which implies $e_b^* < e_r^*$. Thus, subtle discrimination produces an *overcompensation effect* for Red. (Again, an interior solution always obtains for $\sigma < \frac{2}{1+2\beta}$).

These examples show that with continuous skill levels and subtle discrimination, either discouragement or overcompensation effects can arise depending on the chosen bias function $b(e_b, e_r)$. While the above simple functions produce only one effect at a time, complex ones can yield either, depending on parameters. However, justifying more complex functions based only on intuition or introspection is challenging. Our main model with discrete skills and arbitrarily small subtle bias ($\omega \to 0$) provides a microfoundation for the bias function

$$b(e_b, e_r) = \beta(1 - e_b - e_r + 2e_b e_r). \tag{OA.12}$$

Thus, we can replicate all the results in the paper using the continuous approach developed here if we use the bias function in Eq. (OA.12).

To conclude, we show that the equilibrium effort levels depend on how the subtle bias affects the contest success function. Our main model shows a natural example where the subtle bias may lead to either discouragement or overcompensation. Importantly, our model has predictions for when each of these effects is likely to dominate. The simple examples in this Appendix section show that models with continuous skills and non-infinitesimal subtle biases can also generate overcompensation and discouragement.

OA.5 General functional form of investment cost

Here, we show that our main results hold when we use a more general cost function $c(e)$, such that $c(0) = 0$, $c(1) \to \infty$, $c'(.) > 0$, $c''(.) > 0$. In particular, we (numerically) solve the agents' and the principal's problems for the following cost function $c(e) = \frac{k}{\alpha}\frac{e^\alpha}{1-e^\gamma}$.

This form has several advantages. First, for $\alpha = 2$ and $\gamma \to \infty$, it converges to the quadratic cost function $\frac{ke^2}{2}$ used in the main text. Second, for the agent's problem, it guarantees an interior solution for any value of the premium-cost ratio, $\sigma \equiv \frac{W}{k}$. Finally, we confirm that for a sizable interval of parameters values α and γ and for any value of the productivity-cost ratio θ, the social welfare is maximized when both agents are treated symmetrically, that is, invest in their human capital. In particular, we define social surplus under the asymmetric treatment (only one agent invests in her human capital) as $s_{1a}(\theta; \alpha, \gamma) = \max_e \theta e - \frac{k}{\alpha}\frac{e^\alpha}{1-e^\gamma}$ and under the symmetric treatment (both agents invest) as $s_{2a}(\theta; \alpha, \gamma) = \max_e \theta e(2-e) - 2 \times \frac{k}{\alpha}\frac{e^\alpha}{1-e^\gamma}$.

Then, for $\alpha \in [2.0, 10.0]$, we compute $\bar{\gamma}$ such that for all $\gamma < \bar{\gamma}$ (see Figure OA.1), the social welfare is maximized when both agents invest in their human capital for all values of θ. That is, for all values of θ, $s_{2a}(\theta;\alpha,\gamma) > s_{1a}(\theta;\alpha,\gamma)$ as long as $\alpha \in [2.0, 10.0]$ and $\gamma < \bar{\gamma}$. Thus, we confirm that our results are not driven by the fact that under a quadratic cost function, for $\theta > 1$ it is more socially efficient to treat agents asymmetrically. In the remainder of this section we assume $\alpha = 2.0$ ($\bar{\gamma} \approx 18.5013$ for $\alpha = 2.0$).

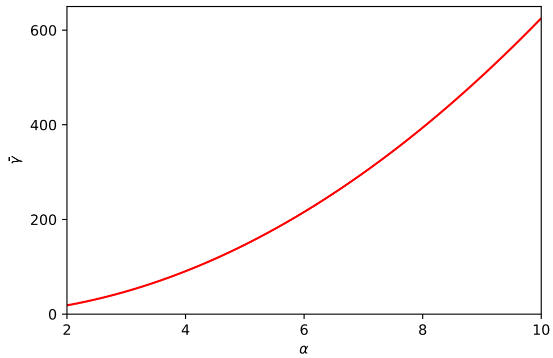


Figure OA.1: $\bar{\gamma}$ as a function of the cost function parameter α

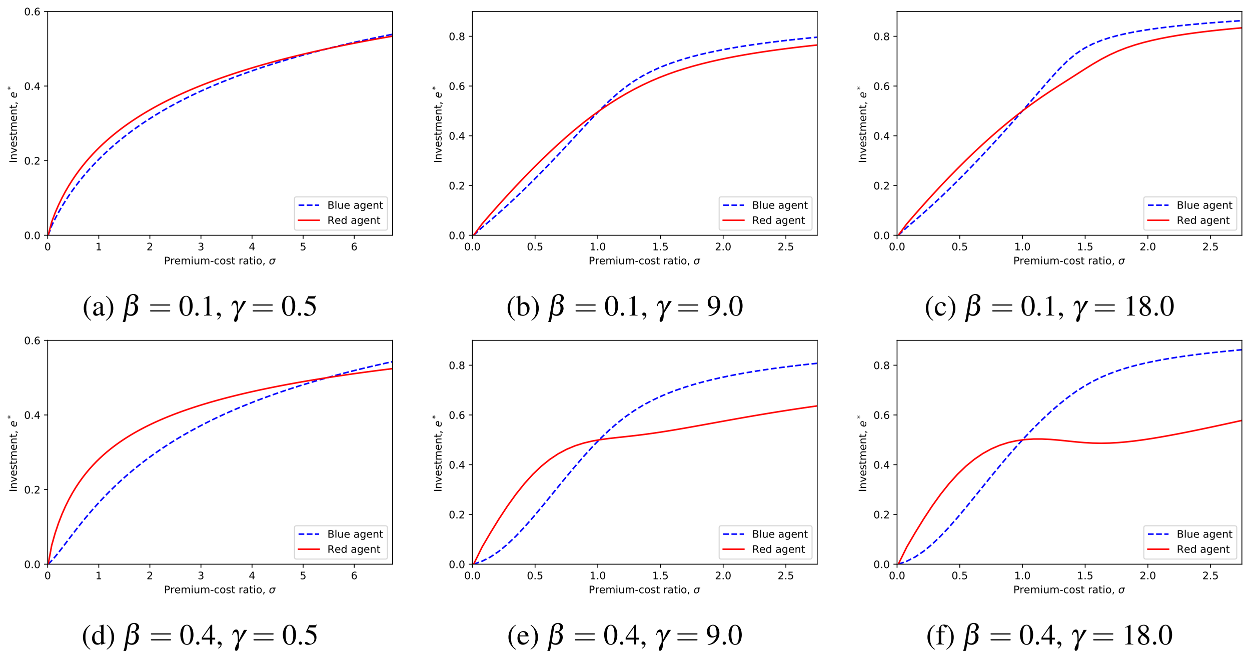
Figure OA.2 shows the equilibrium investment levels as a function of the premium-cost ratio, σ, for two levels of the subtle bias, $\beta \in \{0.1, 0.4\}$ and the following levels of the cost function parameters, $\gamma \in \{0.5, 9.0, 18.0\}$. Note that for $\alpha = 2.0$, all these values of γ are below $\bar{\gamma}$. Figure OA.2 confirms the results in Proposition 2 and Figure 1 of the main text. In particular, it shows that for low values of σ, Red invests more than Blue, while for high values of σ, it is Blue who invests more. Therefore, the existence of the overcompensation and the discouragement effects is not driven by our choice of the quadratic cost function.

Figure OA.3 shows the equilibrium promotion gaps for the same values of β and γ as functions of the premium-cost ratio σ (it replicates the results in Proposition 3 and Figure 2 of the main text). Again, we confirm that for a wide range of parameters, the promotion gap has U-shape and that under high values of σ, the contribution of the achievement gap to the promotion gap is lower than under low values of σ.

Figure OA.4 shows the solution for the principal's problem, the optimal premium-cost ratio, $\sigma^*(\theta;\beta,\gamma,\alpha)$, the optimal profit, $\Pi^*(\theta;\beta,\gamma,\alpha)$, and the resulting promotion gap, $\Delta p^*(\theta;\beta,\gamma,\alpha)$ as functions of the productivity-cost ratio θ and for several values of the subtle bias β and parameters γ and α, $\beta \in \{0.1, 0.4\}$, $\gamma \in \{0.5, 9.0, 18.0\}$ and $\alpha = 2.0$.

Next, we consider a case when a firm optimally chooses its subtle bias. We confirm that there exists θ' such that for all $\theta < \theta'$, the firm prefers $\beta^* = 0.5$ and for all $\theta > \theta'$, the firm prefers $\beta^* = 0$. Figure OA.5 shows θ' as a function of γ, where $\gamma \in [0.5, 18.0]$ and $\alpha = 2$. Thus, firm polarization occurs even under a more general cost function.

Finally, in Figure OA.6 we replicate the results in Figure 4 from the main text for $\gamma = 9.0$ and

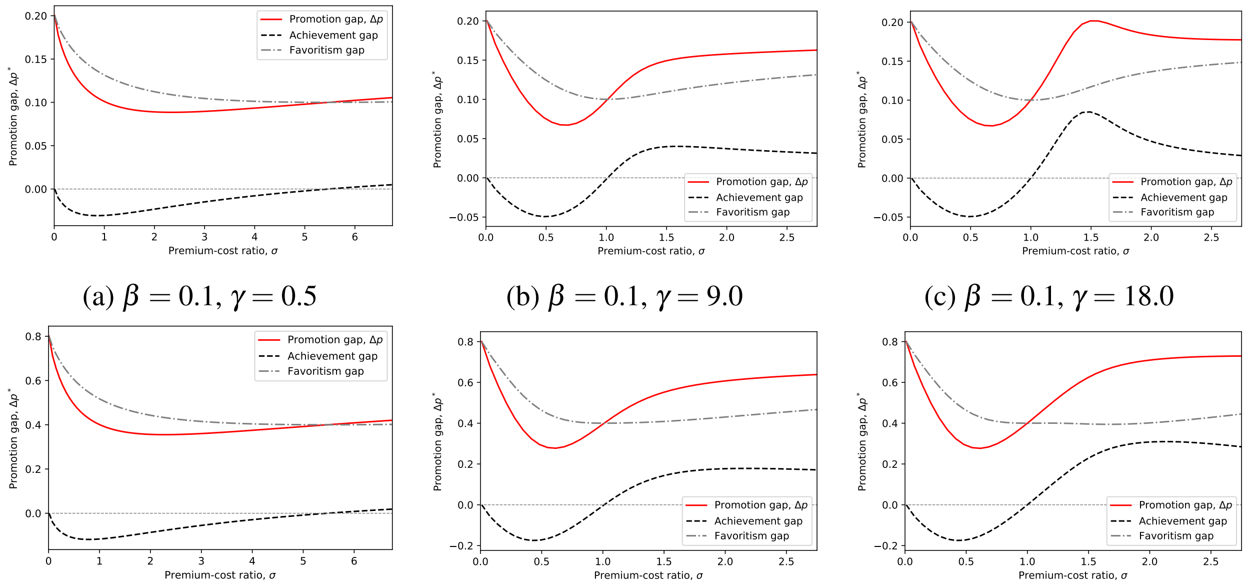


(a) $\beta = 0.1$, $\gamma = 0.5$ (b) $\beta = 0.1$, $\gamma = 9.0$ (c) $\beta = 0.1$, $\gamma = 18.0$

(d) $\beta = 0.4$, $\gamma = 0.5$ (e) $\beta = 0.4$, $\gamma = 9.0$ (f) $\beta = 0.4$, $\gamma = 18.0$

Figure OA.2: Equilibrium investments of blue and red agents, e_b^* and e_r^*, as functions of the premium-cost ratio, σ, for different values of the subtle bias β and the cost function parameter γ.

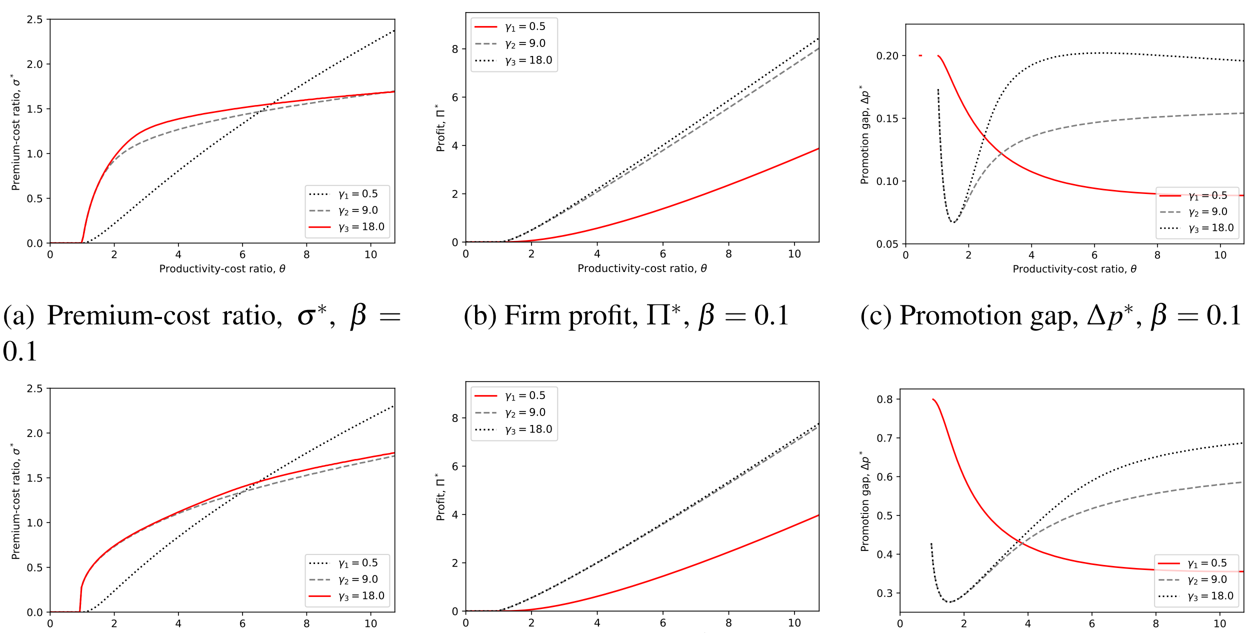
$\alpha = 2.0$. For other values of the cost function parameters, γ and α, the optimal stake, σ^*, the resulting profit, Π^*, and the promotion gap, Δp^* have similar shapes.



(a) $\beta = 0.1$, $\gamma = 0.5$ (b) $\beta = 0.1$, $\gamma = 9.0$ (c) $\beta = 0.1$, $\gamma = 18.0$

(d) $\beta = 0.4$, $\gamma = 0.5$ (e) $\beta = 0.4$, $\gamma = 9.0$ (f) $\beta = 0.4$, $\gamma = 18.0$

Figure OA.3: Equilibrium promotion gap, Δp^*, as a function of the premium-cost ratio, σ, for different values of the subtle bias β and the cost function parameter γ.



(a) Premium-cost ratio, σ^*, $\beta = 0.1$ (b) Firm profit, Π^*, $\beta = 0.1$ (c) Promotion gap, Δp^*, $\beta = 0.1$

(d) Premium-cost ratio, σ^*, $\beta = 0.4$ (e) Firm profit, Π^*, $\beta = 0.4$ (f) Promotion gap, Δp^*, $\beta = 0.4$

Figure OA.4: Optimal premium-cost ratio, σ^*, firm profit, Π^*, and promotion gap, Δp^*, as a function of the productivity-cost ratio, θ, for different values of the subtle bias β and the cost function parameter γ.

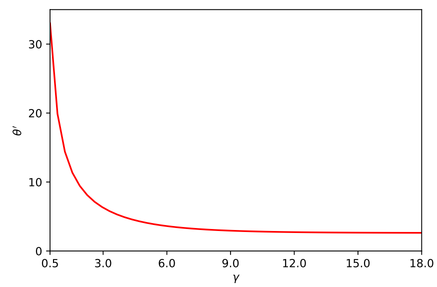
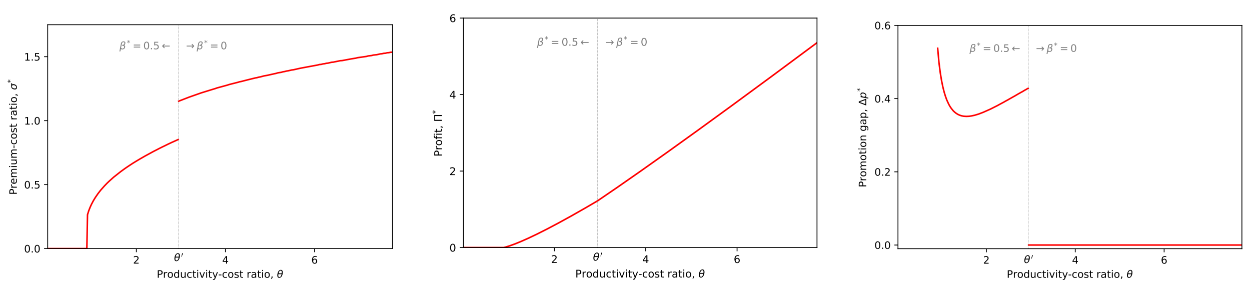


Figure OA.5: θ' as a function of the cost function parameter γ



(a) Premium-cost ratio, $\sigma(\theta)$ (b) Profit, $\Pi(\theta)$ (c) Promotion gap, $\Delta p(\theta)$

Figure OA.6: Optimal premium-cost ratio, equilibrium profit, and promotion gap, as functions of the productivity-cost ratio, θ, for $\gamma = 9.0$ and $\alpha = 2.0$.

OA.6 Filling all slots and promoting low-skill workers

In Section OA.2, we showed an example in which job-2 slots are sometimes left vacant. This has no effect on our analysis, as long as low-skilled workers have some probability of being promoted.

In the main model, our assumption that $l > 0$ implies that the principal always prefers to promote a worker instead of leaving the post vacant (or employing a worker from the outside). Suppose instead that $l < 0$. Then, both reaction functions are downward sloping, and the equilibrium investment levels (assuming an interior solution) are

$$e_b^* = \frac{\sigma - \sigma^2(0.5 - \beta)}{1 - \sigma^2(0.25 - \beta^2)} \quad \text{and} \quad e_r^* = \frac{\sigma - \sigma^2(0.5 + \beta)}{1 - \sigma^2(0.25 - \beta^2)}. \tag{OA.13}$$

Note that $e_b^* > e_r^*$ in any interior solution. Thus, for the overcompensation effect to dominate, we need that when both agents are unskilled, the firm promotes from the inside *with positive probability*. Intuitively, Red's overcompensation only pays off when Blue choose low effort. If Blue cannot be promoted unless he has the skill, Blue does not slack off, and the overcompensation effect is always dominated by the discouragement effect.

OA.7 Multiple levels of subtle bias

Here we show that our results are robust to allowing the principal to have different levels of subtle bias: $\beta_1 \in [0, 0.5]$ for high-skilled agents ($s_r = s_b = 1$) and $\beta_0 \in [0, 0.5]$ for low-skilled agents.

The utility functions for Blue and Red become:

$$U_b = w_1 + W\left[e_b(1-e_r) + \left(\frac{1}{2} + \beta_0\right)(1-e_b)(1-e_r) + \left(\frac{1}{2} + \beta_1\right)e_b e_r\right] - \frac{ke_b^2}{2}; \tag{OA.14}$$

$$U_r = w_1 + W\left[e_r(1-e_b) + \left(\frac{1}{2} - \beta_0\right)(1-e_b)(1-e_r) + \left(\frac{1}{2} - \beta_1\right)e_b e_r\right] - \frac{ke_r^2}{2}. \tag{OA.15}$$

The reaction functions for Blue and Red are

$$e_b = \sigma\left(\frac{1}{2} - \beta_0 + (\beta_0 + \beta_1)\, e_r\right); \tag{OA.16}$$

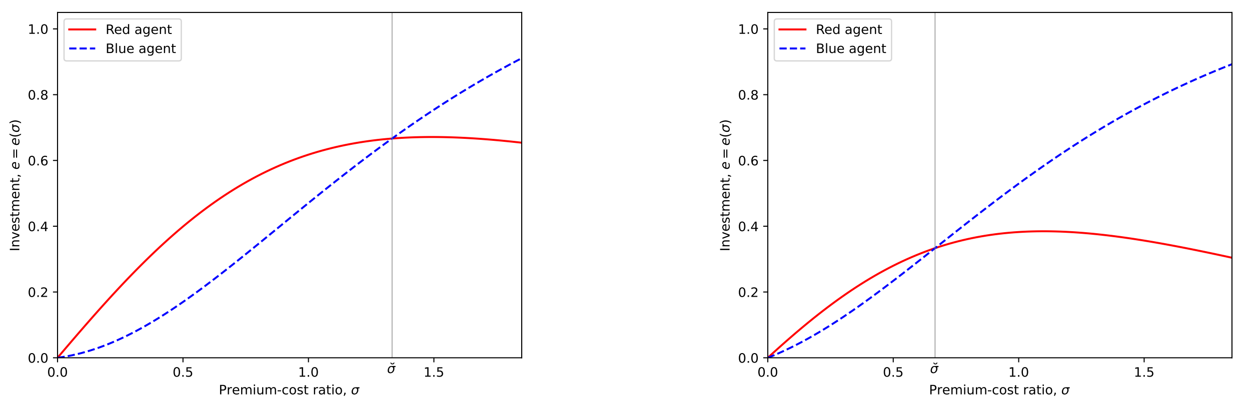
$$e_r = \sigma\left(\frac{1}{2} + \beta_0 - (\beta_0 + \beta_1)\, e_b\right), \tag{OA.17}$$

where $\sigma \equiv \frac{W}{k}$ is the premium-cost ratio. Then, the resulting optimal effort levels (under an internal solution assumption) are:

$$e_b^* = \frac{\sigma(0.5-\beta_0)+\sigma^2(\beta_0+\beta_1)(0.5+\beta_0)}{1+\sigma^2(\beta_0+\beta_1)^2}; \tag{OA.18}$$

$$e_r^* = \frac{\sigma(0.5+\beta_0)-\sigma^2(\beta_0+\beta_1)(0.5-\beta_0)}{1+\sigma^2(\beta_0+\beta_1)^2}. \tag{OA.19}$$

The overcompensation region, $e_r^* > e_b^*$, exists as long as $\sigma < \breve{\sigma} = \frac{2\beta_0}{\beta_0+\beta_1}$. In our original model, the two levels of subtle bias are the same, $\beta_0 = \beta_1 = \beta$, and as a result $\breve{\sigma} = 1$. If the subtle bias is stronger for low-skilled workers, $\beta_0 > \beta_1$, then the overcompensation region expands, i.e., $\breve{\sigma} > 1$. If the bias is stronger for high-skilled workers, $\beta_0 < \beta_1$, then the overcompensation region shrinks, i.e., $\breve{\sigma} < 1$. These comparative statics are illustrated in Figure OA.7. The resulting equilibrium promotion gap, Δp^*, is presented in Figure OA.8.



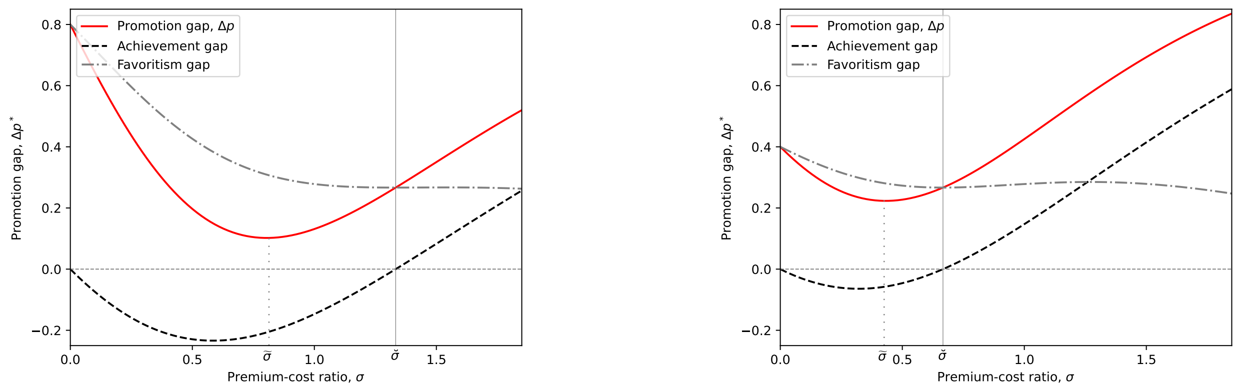
(a) $\beta_0 = 0.4 > \beta_1 = 0.2$

(b) $\beta_0 = 0.2 < \beta_1 = 0.4$

Figure OA.7: Equilibrium investments of blue and red agents, e_b^* and e_r^*, as functions of the premium-cost ratio, σ, for different values of the subtle bias, β_0 and β_1.

OA.7.1 Optimal compensation contracts

We now allow the principal to design the compensation contract, taking into account his subtle bias for low-skilled agents, β_0, and for high-skilled agents, β_1, as given. As in the main text, we assume that the principal knows that the agents have correct beliefs about β_0 and β_1. The principal's problem and its solution are analogous to those in the main text. In Figure OA.9, we illustrate the optimal premium-cost ratio, σ^*, the resulting profit, Π^*, and the promotion gap, Δp^*, as functions of the productivity-cost ratio, θ for two cases. Case A: the subtle bias is stronger for low-skilled workers, $\beta_0 = 0.4 > \beta_1 = 0.2$. Case B: the subtle bias is stronger for high-skilled workers, $\beta_0 = 0.2 < \beta_1 = 0.4$. Our results are qualitatively similar to those in the main text.



(a) $\beta_0 = 0.4 > \beta_1 = 0.2$

(b) $\beta_0 = 0.2 < \beta_1 = 0.4$

Figure OA.8: Equilibrium promotion gap, Δp^*, as a function of the premium-cost ratio, σ, for different values of the subtle bias, β_0 and β_1.

OA.7.2 Optimal anti-discrimination policies

In this section, we investigate the problem of a principal who can simultaneously choose both the compensation contract, σ, and the two levels of subtle bias, β_0 and β_1.

Suppose, as in the main text, the principal can choose policies associated with the two levels of bias, β_0 and β_1 at no cost. Will the principal choose the same level of bias both for skilled and unskilled agents, $\beta_0 = \beta_1$? Or will he prefer to differentiate, $\beta_0 \neq \beta_1$? If so, how does this choice depend on the productivity-cost ratio, θ? The principal's problem is

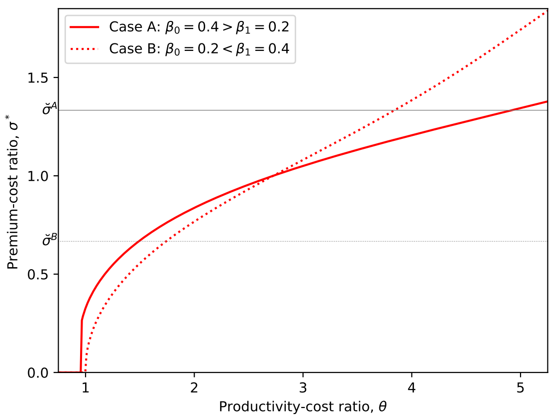
$$\Pi(\theta) = \max_{(\sigma,\beta_0,\beta_1)\in[0,\overline{\sigma}(\beta_0,\beta_1)]\times[0,0.5]\times[0,0.5]} \theta\left(e_b + e_r - e_b e_r\right) - \sigma, \qquad \text{(OA.20)}$$

subject to (OA.18) and (OA.19).

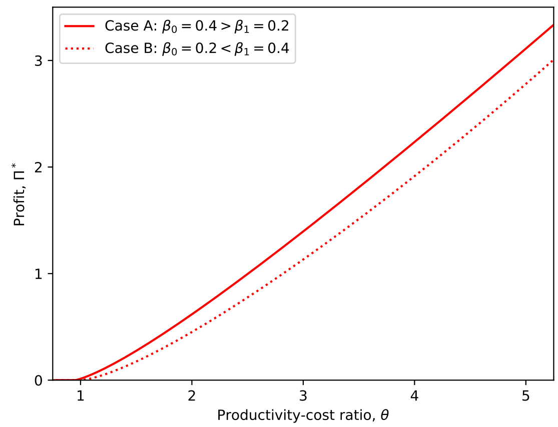
We solve the principal's problem numerically. For a given value of the productivity-cost ratio θ, we fix a pair of values of β_0 and β_1 from the interval intersection $[0,0.5]\times[0,0.5]$, and choose the value of the premium-cost ratio, σ, that maximizes the principal's profit, Π. Then, we maximize the principal's profit across all pairs of values of β_0 and β_1 and obtain the optimal profit $\Pi^*(\theta)$ as a function of the productivity-cost ratio, as well as the optimal values of the parameters, $\beta_0^*(\theta)$, $\beta_1^*(\theta)$, and $\sigma^*(\theta)$. Figure OA.10 presents the solution.

The optimal subtle bias for low-skilled agents does not depend on the productivity-cost ratio (see Figure OA.10a). The principal always sets it to its highest possible value, $\beta_0^* = 0.5$ for all $\theta > \underline{\theta}$, where $\underline{\theta} \approx 0.891$, when he operates the firm.[3] This way, the principal discourages the red agent from tying with the blue agent at zero, thereby maximizing gains from the overcompensation

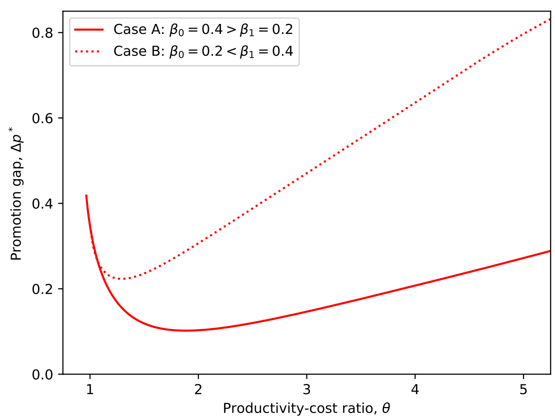
[3]Remember that in the absence of the subtle bias, $\beta_0 = \beta_1 = 0$, the principal chooses zero stakes and does not operate the firm for all $\theta < 1$.



(a) Premium-cost ratio, $\sigma(\theta)$



(b) Profit, $\Pi(\theta)$



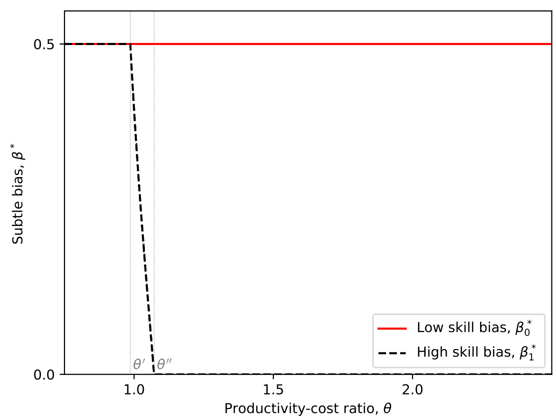
(c) Promotion gap, $\Delta p(\theta)$

Figure OA.9: Optimal premium-cost ratio, equilibrium profit, and promotion gap, as functions of the productivity-cost ratio, θ, for different values of the subtle bias, β_0 and β_1.

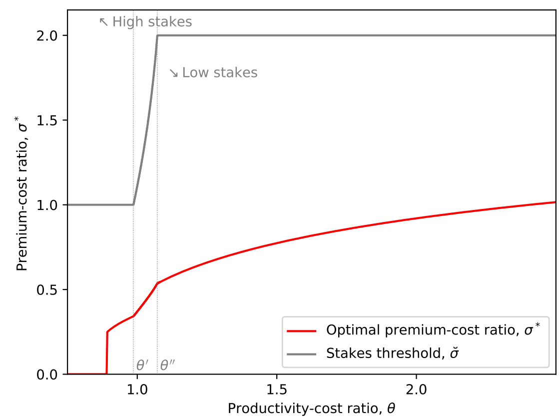
effect.

For high-skilled agents, the optimal subtle bias varies with the productivity-cost ratio (see Figure OA.10a). For low levels of the productivity-cost ratio, $\theta \in [\underline{\theta}, \theta']$, where $\theta' \approx 0.98625$, the principal sets the subtle bias for high-skilled agents to its highest possible value, $\beta_1^* = 0.5$. On the interval $[\theta', \theta'']$, where $\theta'' \approx 1.07195$, the optimal subtle bias for high-skilled agents β_1^* is decreasing in the productivity-cost ratio, θ. Finally, the optimal bias is zero, $\beta_1^* = 0$ for all $\theta > \theta''$. In other words, the firm does not discriminate between high-skilled workers in order to preserve their incentives to invest in their human capital.

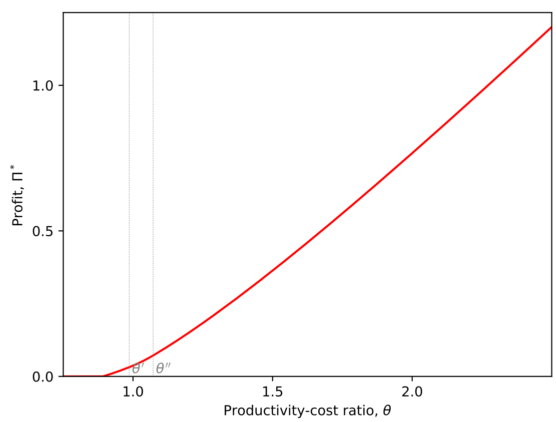
Figure OA.10b depicts the optimal stakes, σ^* as a function of the productivity-cost ratio, θ (solid red line) and the stakes threshold, $\breve{\sigma}^* = \frac{2\beta_0^*}{\beta_0^* + \beta_1^*}$ (solid grey line). Note that the firm consistently offers low-stakes careers in order to benefit from the overcompensation effect and the improved performance of the discriminated agents. By setting different levels of the subtle bias for



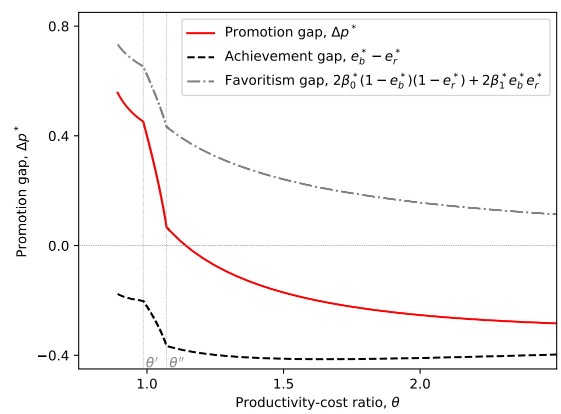
(a) Subtle bias, $\beta(\theta)$



(b) Premium-cost ratio, $\sigma(\theta)$



(c) Profit, $\Pi(\theta)$



(d) Promotion gap, $\Delta p(\theta)$

Figure OA.10: Optimal subtle bias, premium-cost ratio, equilibrium profit, and promotion gap, as functions of the productivity-cost ratio, θ.

high- and low-skill workers, the principal can take full advantage of the overcompensation effect while avoiding the costs of the discouragement effect.

Finally, under the optimal levels of the subtle bias for high- and low-skill workers, the promotion gap can be negative for large enough values of the productivity-cost ratio, θ (see Figure OA.10d). The promotion gap is negative because the achievement gap, $e_b^* - e_r^*$, is negative and large, while the favoritism gap, $2\beta_0^*(1-e_b^*)(1-e_r^*) + 2\beta_1^* e_b^* e_r^*$, is relatively small.

OA.8 Cost heterogeneity

Here we provide an extended analysis of the case where agents face different costs.

OA.8.1 Subtle bias and the productivity effect

Let k_b and k_r denote the cost parameters. Define $\kappa(\beta) = \frac{0.5+\beta}{0.5-\beta}$. If the ratio of marginal costs, $\frac{k_r}{k_b}$, is greater than $\kappa(\beta)$, then $\widehat{W} = 0$ and Blue always works harder than Red (see the proof of Corollary 2, in OA.1). Intuitively, if Red's investment disadvantage is too large, the overcompensation effect cannot outweigh Blue's productivity advantage. Similarly, if Red's marginal costs are low enough, Red always invests more than Blue.

Suppose now that $\frac{k_r}{k_b} < \kappa(\beta)$. To find a threshold $\widehat{W} > 0$, suppose we have $e_b^* = e_r^* = e$. Then, the reaction functions imply

$$W = \frac{1}{2\beta}\frac{(k_b - k_r)e}{2e-1} \equiv W(e).$$

Note that W is positive for $\frac{k_r}{k_b} < 1$ and $e > 0.5$ (case 1); and $\frac{k_r}{k_b} > 1$ and $e < 0.5$ (case 2).

Case 1: $W(e)$ is decreasing in e, for $e \in (0.5, 1]$. For $e < 1$, we need $W(e) = W_2$, which implies $W(1) < W_2$. We obtain

$$\frac{k_r}{k_b} > \frac{0.5-\beta}{0.5+\beta} = \frac{1}{\kappa(\beta)}.$$

Thus,

$$\frac{k_r}{k_b} \in \left(\frac{1}{\kappa(\beta)}, 1\right).$$

Case 2: $W(e)$ is increasing in e, for $e \in [0, 0.5)$. For $e > 0$, we need $W(e) = W_2$, which implies $W(0) < W_2$. We obtain

$$\frac{k_r}{k_b} < \frac{0.5+\beta}{0.5-\beta} = \kappa(\beta).$$

Thus,

$$\frac{k_r}{k_b} \in (1, \kappa(\beta)).$$

Now, we can combine the above two equations and obtain that an interior switching point W_2 exists if and only if

$$\frac{k_r}{k_b} \in \left(\frac{1}{\kappa(\beta)}, \kappa(\beta)\right).$$

Because $\kappa(\beta)$ is strictly positive for $\beta \in (0, 0.5]$, strictly increasing, and $\lim_{\beta \to 0.5} \kappa(\beta) = \infty$, for any k_b and k_r there exists β' such that $\widehat{W}$ is strictly interior for any $\beta > \beta'$.

OA.8.2 Non-quadratic convex cost functions

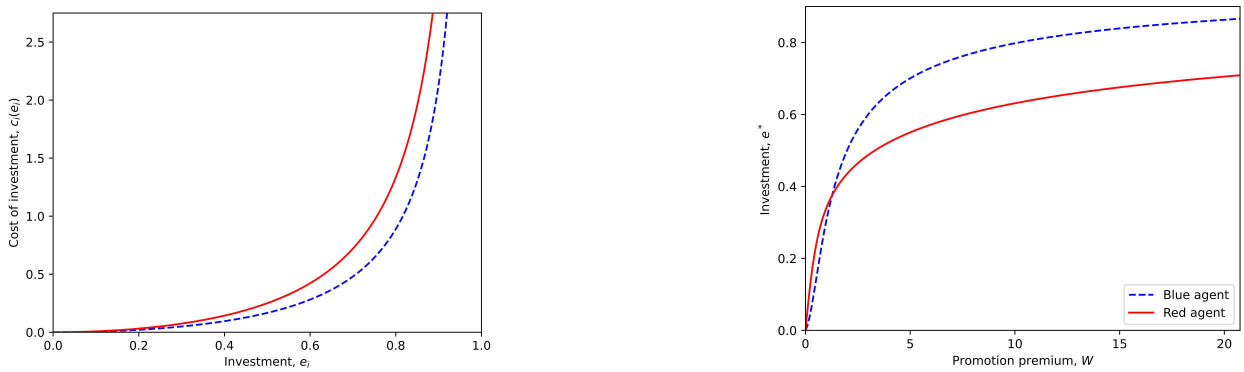
As in OA.5, we consider the following cost function: $c(e_i) = \frac{k_i}{\alpha_i}\frac{e^{\alpha_i}}{1-e^{\gamma_i}}$, where parameters k_i, α_i, and γ_i are agent specific, $i = r, b$. This functional form is flexible, allowing us to study several interesting scenarios. For example, it accommodates situations where one agent consistently has a cost advantage for all values of investment, as well as cases where an agent benefits at low levels of e_i but faces a disadvantage for higher levels of e_i. As before, we assume that $\alpha_r = \alpha_b = 2$ and consider values of k_i and γ_i such that the social welfare is maximized when both agents are allowed to invest in their human capital. The agents' reaction functions are

$$\frac{k_b}{2}\frac{e_b\left((\gamma_b-2)e_b^{\gamma_b}+2\right)}{\left(1-e_b^{\gamma_b}\right)^2} = W\left(\frac{1}{2}-\beta+2\beta e_r\right), \tag{OA.21}$$

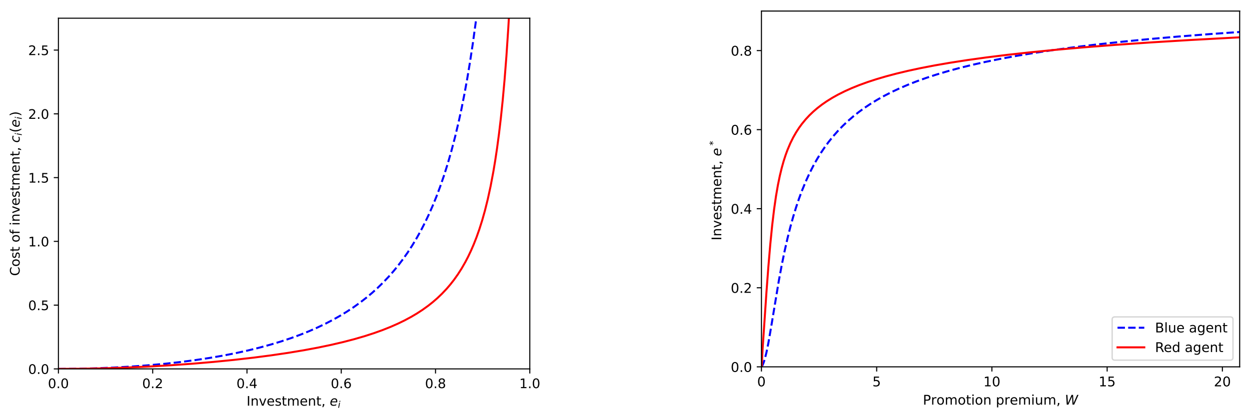
$$\frac{k_r}{2}\frac{e_r\left((\gamma_r-2)e_r^{\gamma_r}+2\right)}{\left(1-e_r^{\gamma_r}\right)^2} = W\left(\frac{1}{2}+\beta-2\beta e_b\right). \tag{OA.22}$$

We solve the above system of the agents' reactions functions with asymmetric costs functions to show that our main results are not driven by the symmetry of the cost functions assumed in the main text.

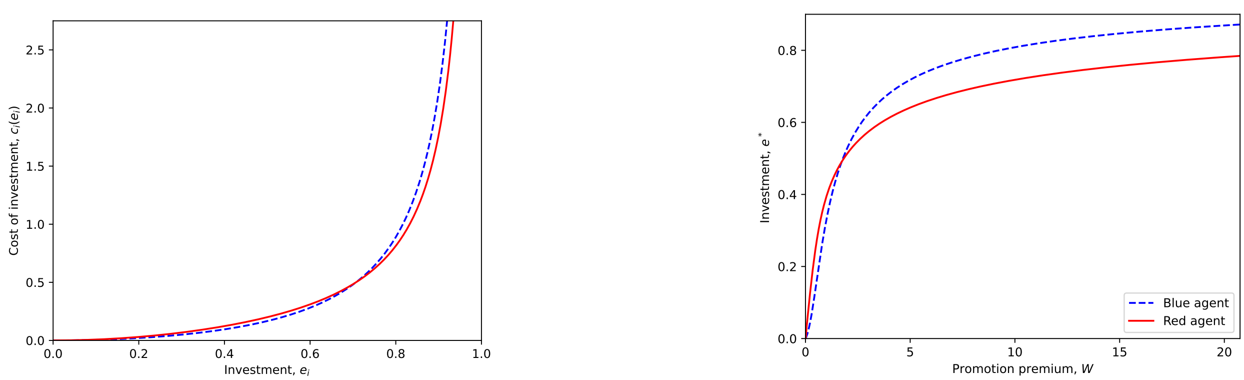
Figure OA.11 shows the investment costs for Blue and Red as functions of investment levels (on the left hand side) and the corresponding equilibrium investments levels (on the right-hand side) for different values of parameters k_i and γ_i, with $\beta = 0.4$. By adjusting the values of k_i and γ_i, we can alter the relative advantage in investment costs between the agents. For example, when $k_b = 1.0$, $k_r = 1.5$, $\gamma_b = 2.0$, $\gamma_r = 2.0$, Blue consistently incurs lower investment costs than Red (see Figure OA.11a). Conversely, when $k_b = 1.5$, $k_r = 1.0$, $\gamma_b = 2.0$, $\gamma_r = 4.0$, Red consistently incurs lower costs than Blue (see Figure OA.11b). In contrast, when $k_b = 1.0$, $k_r = 1.5$, $\gamma_b = 2.0$, $\gamma_r = 4.0$, neither agent has a consistent advantage in investment costs (see Figure OA.11c), as Blue has an advantage at low investment levels, while Red has an advantage at high investment levels. Nevertheless, in all three cases, Red invests more than Blue for low levels of W (the *overcompensation effect*) and invests less than Blue for high levels of W (the *discouragement effect*). Thus, our main results are robust to deviations from cost function symmetry. Of course, large deviations could lead to one agent always investing more than the other. They could also exacerbate our results, for example, if Red has a cost advantage at low investment levels but faces a disadvantage at high investment levels.



(a) Cost advantage for Blue: $k_b = 1.0$, $k_r = 1.5$, $\gamma_b = 2.0$, $\gamma_r = 2.0$



(b) Costs advantage for Red: $k_b = 1.5$, $k_r = 1.0$, $\gamma_b = 2.0$, $\gamma_r = 4.0$



(c) No consistent cost advantage: $k_b = 1.0$, $k_r = 1.5$, $\gamma_b = 2.0$, $\gamma_r = 4.0$

Figure OA.11: Asymmetric costs of investment, $c_i(e_i)$, and equilibrium investments of blue and red agents, e_b^* and e_r^*, for $\beta = 0.4$ and different values of parameters k_i and γ_i.

OA.8.3 The general case

In this subsection, we consider the case of general cost functions. Our goal is to establish the conditions under which the results in Corollaries 1 and 2 generalize to other cost functions.

Let $c_i : [0,1] \to \Re^+$ denote the cost function of agent $i \in \{b,r\}$. We maintain the assumptions that $c_i(.)$ is class-C^2 differentiable, $c_i'(e) > 0$ for all $e > 0$, $c_i''(e) > 0$ everywhere (which implies strict convexity), and $c_i(0) = c_i'(0) = 0$.[4]

For simplicity, we consider only the limiting case where $\beta \to 0.5$, for three reasons. First, setting $\beta = 0.5$ considerably simplifies the algebra without changing the economics. Second, it eliminates corner solutions for the promotion premium thresholds, reducing the number of cases to consider without affecting any of the economic insights. Third, the model is continuous in β, thus all results discussed here hold exactly for β close to 0.5, i.e., when the subtle bias is sufficiently strong. Meaningful differences between e_b^* and e_r^* arise only when β is sufficiently large (see, e.g., Figure 1, (a) and (b)). Thus, the case of a large β is the empirically relevant one. We note, however, that the results in this subsection generalize to any β, at the cost of more complexity.

For $\beta = 0.5$, the reaction functions are implicitly defined by

$$c_b'(e_b) - We_r = 0; \tag{OA.23}$$

$$c_r'(e_r) - W(1-e_b) = 0. \tag{OA.24}$$

The reaction functions are well-defined because the cost functions are strictly increasing and convex. Because $e_b(e_r)$ is strictly increasing and $e_r(e_b)$ strictly decreasing, there is a unique equilibrium (e_b^*, e_r^*). For sufficiently low W, the equilibrium is interior (i.e., both e_b^* and e_r^* are lower than 1).

Note that at $W = 0$, we must have $e_r^* = e_b^* = 0$. Is there a positive promotion premium at which both types exert the same amount of effort in equilibrium? The next result shows that the answer is positive.

Result 1. *There exists $\widehat{W} > 0$ such that the equilibrium is interior and $e_b^*(\widehat{W}) - e_r^*(\widehat{W}) = 0$.*

Proof. Suppose $e_b = e_r = e > 0$. The reaction functions imply $c_b'(e) + c_r'(e) = W$. If both agents exert the same amount of effort e at some positive W, we then need

$$c_b'(e) + c_r'(e) = \frac{c_b'(e)}{e}, \tag{OA.25}$$

[4] Notation $c_i'(0)$ is a shortcut for $\lim_{e\to 0} c'(e)$. Similarly, $c_i'(1) = \lim_{e\to 1} c'(e)$. The same applies for second derivatives.

where the right-hand side comes from (OA.23). At $e = 0$, we have

$$\lim_{e \to 0} c_b'(e) + c_r'(e) - \frac{c_b'(e)}{e} = -c_b''(0) < 0 \tag{OA.26}$$

(using L'Hôpital's Rule).[5] At $e = 1$, we have

$$\lim_{e \to 1} c_b'(e) + c_r'(e) - \frac{c_b'(e)}{e} = c_r'(1) > 0. \tag{OA.27}$$

By continuity, there must exist at least one fixed point $\hat{e} < 1$ such that (OA.25) holds. □

This result shows that there must be at least one $\widehat{W} > 0$ such that Blue and Red invest the same amount. The next result shows that if $\widehat{W}$ is unique, we get the results in Corollaries 1 and 2.

Result 2. *Suppose there is a unique $\widehat{W} > 0$ such that $e_b^*(\widehat{W}) - e_r^*(\widehat{W}) = 0$. Then, Red invests more than Blue if $W < \widehat{W}$ and Blue invests more than Red if $W > \widehat{W}$.*

Proof. At an interior equilibrium, differentiating the implicit function $e_b^*(W)$ and using $c_b'(e_b^*) = We_r^*$, we get

$$\frac{de_b^*(W)}{dW} = \frac{e_r^* c_r''(e_r^*) + W(1 - e_b^*)}{c_b''(e_b^*) c_r''(e_r^*) + W^2}. \tag{OA.28}$$

Thus, $e_b^*(W)$ is strictly increasing for $W > 0$ (this is the counterpart of Corollary 4, Part 1). Thus, we have

$$\frac{de_b^*(0)}{dW} = 0. \tag{OA.29}$$

Differentiating the implicit function $e_r^*(W)$ and using $c_r'(e_r^*) = W(1 - e_b^*)$, we get

$$\frac{de_r^*(W)}{dW} = \frac{(1 - e_b^*) c_b''(e_b^*) - We_r^*}{c_b''(e_b^*) c_r''(e_r^*) + W^2}, \tag{OA.30}$$

implying

$$\frac{de_r^*(0)}{dW} = \frac{1}{c_r''(0)} > 0. \tag{OA.31}$$

We know that at $W = 0$, $e_b^* - e_r^* = 0$. Thus, for $W = \varepsilon > 0$ arbitrarily small, $e_b^* - e_r^* < 0$. This means that Red always invests more than Blue for sufficiently low W. At $e_b = 1$, we must have $e_r = 0$, thus Red always invests less than Blue for sufficiently high W. Because there is a unique $\widehat{W} > 0$ such $e_b^*(\widehat{W}) - e_r^*(\widehat{W}) = 0$, then Red invests more (less) than Blue if $W < (>) \widehat{W}$. □

[5] The assumption that $c_i''(e) > 0$ everywhere is convenient but not necessary. This result goes through under the much weaker (but less intuitive) assumption that $c_r'(e)/c_b'(e) \leq 1/e - 1$ for all $e < \varepsilon$ for some arbitrarily small $\varepsilon > 0$. If $c_b''(0) > 0$, this condition is trivially satisfied.

We conclude that we obtain Corollaries 1 and 2 in the main text because there is a unique "crossing point" when the cost functions are quadratic.[6] Subsection OA.8.2 shows that this property is not unique to quadratic functions. Result 1 implies that at least one crossing point must exist (for sufficiently high β). Next, we present the most general result, allowing for multiple crossing points.

Result 3. *There exist $\widehat{W}_h \geq \widehat{W}_l > 0$ such that $e_b^*(\widehat{W}_l) - e_r^*(\widehat{W}_l) = e_b^*(\widehat{W}_h) - e_r^*(\widehat{W}_h) = 0$, and Red invests more than Blue if $W < \widehat{W}_l$ and Blue invests more than Red if $W > \widehat{W}_h$.*

Proof. The proof of Result 1 shows that function $c_b'(e) + c_r'(e)$ must cross $\frac{c_b'(e)}{e}$ at least once as e varies from 0 to 1. Equation (OA.26) shows that $c_b'(e) + c_r'(e)$ is lower than $\frac{c_b'(e)}{e}$ for e arbitrarily close to zero. Equation (OA.27) shows that $c_b'(e) + c_r'(e)$ is greater than $\frac{c_b'(e)}{e}$ for e arbitrarily close to one. Thus, if these functions cross more than once, there must be an odd number of such crossing points. $\widehat{W}_h$ is defined as the largest crossing point and $\widehat{W}_l$ as the lowest (strictly positive) crossing point. The proof of Result 2 implies that Red invests more than Blue for sufficiently low values of W and Blue invests more than Red for sufficiently high values of W. $\square$

Result 3 generalizes Corollaries 1 and 2 for general increasing and strictly convex cost functions. If $\widehat{W}_h = \widehat{W}_l$, there is a unique value of $W > 0$ such that $e_b^* - e_r^*$ changes sign. In this case, Red invests more (less) than Blue if $W < (>) \widehat{W}$. If $\widehat{W}_h > \widehat{W}_l$, there could also exist points at which $e_b^* - e_r^*$ changes sign from positive to negative. Importantly, such points are always sandwiched between $\widehat{W}_l$ and $\widehat{W}_h$. We note that the behavior of $e_b^* - e_r^*$ close to the boundaries (0 and 1) is of particular importance because it is such behavior that creates incentives for the principal to choose β optimally at the boundaries, as shown in Proposition 7.

To summarize the results, define a *C-point* (for "crossing point") as $W_C > 0$ such that $e_b^*(W_C) - e_r^*(W_C) = 0$ and the sign of $e_b^* - e_r^*$ changes from negative to positive at this point. Define an *R-point* (for "reverse crossing point") as $W_R > 0$ such that $e_b^*(W_R) - e_r^*(W_R) = 0$ and the sign of $e_b^* - e_r^*$ changes from positive to negative at this point. Then the results can be summarized as follows: (i) A C-point always exists (for sufficiently high β); (ii) if an R-point exists, there exist at least two C-points, and the R-point lies between two C-points. Thus, R-points are "less common" than C-points.

In our examples (the quadratic case and the non-quadratic examples in Subsection OA.8.2) we always found a unique C-point and no R-points. This suggests that costs functions must have specific properties for R-points to arise. The next result shows necessary conditions for R-points to exist.

[6] If $\beta < 0.5$ there might be no crossing point, as shown in Subsection OA.8.1. In such a case, either $e_b^* - e_r^* < 0$ or $e_b^* - e_r^* > 0$ for all values of $W > 0$ that lead to an interior solution. This is the only difference between Corollary 1 and Corollary 2. The important result is that, under quadratic costs, there is no crossing point that shows "reversal," where Red invests more than Blue for high W but not for low W.

Result 4. *Suppose there exists $e = e_b^* = e_r^* > 0$ and let $\widehat{W} > 0$ denote the promotion premium associated with these equilibrium effort levels. Then, $\widehat{W}$ is an R-point if and only if*

$$\frac{c_r'(e)}{1-e} + ec_r''(e) < (1-e)c_b''(e). \tag{OA.32}$$

Proof. If $e = e_b^* = e_r^* > 0$, then

$$c_b'(e) + c_r'(e) = \frac{c_b'(e)}{e}. \tag{OA.33}$$

For this to be an R-point, the right-hand side curve must cross the (strictly increasing) left-hand side curve from below, implying

$$c_b''(e) + c_r''(e) < \frac{ec_b''(e) - c_b'(e)}{e^2}. \tag{OA.34}$$

Rearranging it yields (OA.32). □

Condition (OA.32) helps with the intuition for many results. First, note that Condition (OA.32) does not hold when $e \to 1$. While (OA.32) can hold as $e \to 0$, condition (OA.33) cannot (under the maintained assumption that $c''(0) > 0$). Thus, R-points – if they exist – must involve intermediate effort levels. Second, if both agents have the same cost function, *R-points do not exist*. This result follows because, with identical functions, condition (OA.33) implies $e^* = 0.5$ and condition (OA.32) does not hold if $e = 0.5$ and $c_b''(e) = c_r''(e)$. Thus, identical cost functions always imply a unique C-point and no R-points.

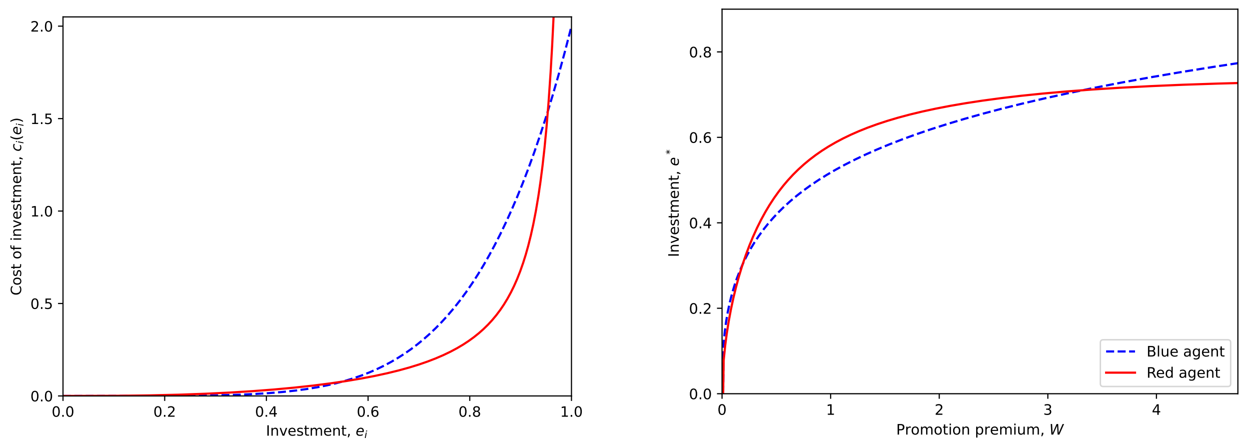
Intuitively, for Red to respond more strongly than Blue to an increase in stakes (when stakes are no longer small), her marginal cost must grow at a significantly lower rate than Blue's marginal cost. Economically, this implies that Red has a *comparative advantage* over Blue in the cost of effort provision for a range of intermediate effort levels. While this can happen for some cost functions, the empirical relevance of this case is questionable. If costs differ, we would normally expect the unfavored agents to have higher marginal costs, either because of discrimination in the provision of training, differences in mentoring, or differences in opportunity costs. If Red has sufficiently high marginal costs, Condition (OA.32) does not hold.

Result 4 implies that R-points are theoretically possible but require some complex conditions on the shape of both cost functions. While C-points are easy to find (see the previous subsection), we have not been able to find an R-point in numerical examples in which our assumptions hold. We have been able to find an example of an R-point with the following functions:

$$c_b(e_b) = A_0 e_b^{B_0} + A_1 e_b^{B_1},$$
$$c_r(e_r) = \frac{k_r}{\alpha_r} \frac{e_r^{\alpha_r}}{1 - e_r^{\gamma_r}},$$

where $A_0 = 1.0$, $A_1 = 1.0$, $B_0 = 5.0$, $B_1 = 6.0$, and for $k_r = 1.0$, $\alpha_r = 2.7$, $\gamma_r = 5.0$. In this case, the agents' cost functions cross three times (including at zero); see Figure OA.12a.

Note that, in this example, $c_b''(0) = 0$, which violates our assumption that $c''(e)$ is strictly positive everywhere. In this case, it is possible that an R-point comes before a C-point, as we can see in Figure OA.12b.



(a) Cost functions of Blue and Red

(b) Optimal efforts for Blue and Red, for $\beta = 0.5$

Figure OA.12: Cost functions cross multiple times

In this example, the R-point happens in regions where both cost functions are fairly flat. Before the R-point, $e_b^* - e_r^*$ is positive but very small.

Motivated by this example, here we offer some comments on the assumption that $c_i''(e) > 0$ everywhere. This assumption implies, in particular, that $c_b''(0) > 0$ and $c_r''(0) > 0$. Nothing would change in our analysis if $c_r''(0) = 0$. However, some of our proofs used the fact that $c_b''(0) > 0$. Thus, how do the results change if we allow for $c_b''(0) = 0$? Consider the following assumption:

Assumption 1. *$c_r'(e)/c_b'(e) \leq \frac{1}{e} - 1$ for all $e < \varepsilon$ for some arbitrarily small $\varepsilon > 0$.*

Note that $c_b''(0) > 0$ implies Assumption 1, but not vice versa. If $c_b''(0) = 0$ but Assumption 1 holds, all previous results hold exactly. Instead, if $c_b''(0) = 0$ but Assumption 1 does not hold, the results change in the following way:

1. There may not exist $\widehat{W} > 0$ such that $e_b^*(\widehat{W}) - e_r^*(\widehat{W}) = 0$. In this case, Blue always invests more than Red.

2. If such $\widehat{W} > 0$ exists, there are at least two of such points. The lowest is an R-point and the highest is a C-point. The number of R-points equals the number of C-points.

Figure OA.12b illustrates a case in which $c_b''(0) = 0$ *and* Assumption 1 does not hold. In such a case (and only in this case), we can find an R-point that comes before a C-point, as the Figure illustrates. But it is not possible to find only an R-point and no C-point.

OA.9 Non-binary skill and endogeneous skill assessment precision

Here we show that our results are robust to relaxing the assumption that the acquired skill is a binary variable. We extend the model to three levels of skill, $s_i = \{0, 0.5, 1\}$.

OA.9.1 Agent's problem

As in the main model, at Date 1, the agents simultaneously undertake a nonverifiable investment (or effort), $e_i \in [0,1]$, $i \in b, r$, in firm-specific human capital. Both agents are risk-neutral and have the same skill-acquisition cost function, $c(e_i)$, strictly increasing and convex and such that $c(0) = 0$. Agent i's probability of acquiring the lower skill level, $s_i = 0.5$, is e_i and the higher skill level, $s_i = 1$, is αe_i, where $\alpha \in [0,1)$. For example, in project management, project planning, scheduling, and budgeting are considered parts of the foundational skill level ($s_i = 0.5$ in our interpretation), while vision and goal-setting for projects as well as ability to align project objectives with organizational strategy are considered parts of more advanced skill level ($s_i = 1$ in our model). For completeness, we assume that the probability of an agent remaining unskilled is $1 - e_i - \alpha e_i \geq 0$, which implies that the following condition is necessary for an interior solution: $e_i \leq \frac{1}{1+\alpha}$.

First, we compute probabilities of three disjoint outcomes:

1. Both agents are equally skilled with probability $p_{tie} = e_b e_r + \alpha^2 e_b e_r + (1 - \alpha e_b - e_b)(1 - \alpha e_r - e_r) = 1 + 2 e_b e_r (1 + \alpha + \alpha^2) - (e_b + e_r)(1 + \alpha)$;

2. Blue is more skilled with probability $p_{btop} = \alpha e_b (1 - \alpha e_r) + e_b (1 - e_r - \alpha e_r)$;

3. Red is more skilled with probability $p_{rtop} = \alpha e_r (1 - \alpha e_b) + e_r (1 - e_b - \alpha e_b)$.

Now we can write down the maximization problem for Blue and Red under the assumption that the effort cost function is quadratic, $c(e_i) = \frac{ke_i^2}{2}$:

$$\max_{e_b} \sigma \left(p_{btop} + (\frac{1}{2} + \beta) p_{tie} \right) - \frac{e_b^2}{2}; \tag{OA.35}$$

$$\max_{e_r} \sigma \left(p_{rtop} + (\frac{1}{2} - \beta) p_{tie} \right) - \frac{e_r^2}{2}. \tag{OA.36}$$

The first order conditions imply the following reaction functions for Blue and Red $e_b = \sigma\big((\alpha + 1)(\frac{1}{2} - \beta) + 2\beta e_r(\alpha^2 + \alpha + 1)\big)$ and $e_r = \sigma\left((\alpha+1)(\frac{1}{2}+\beta) - 2\beta e_r(\alpha^2+\alpha+1)\right)$. Note that in the absence of subtle bias $\beta = 0$, the reaction functions are flat, just as in the main text: $e_b^* = e_r^* = \frac{\sigma}{2}(\alpha+1)$.

If $\alpha = 0$, the optimal investment levels are the same as in the main text (see Eq. (8) and (9) in the main text). For $\alpha \in (0,1)$, we solve for e_b^* and e_r^* and obtain the solution:

$$e_b^* = \sigma(1+\alpha)\frac{0.5 - \beta + 2\beta\sigma(0.5+\beta)(1+\alpha+\alpha^2)}{1 + 4\sigma^2\beta^2(\alpha^2+\alpha+1)^2} \tag{OA.37}$$

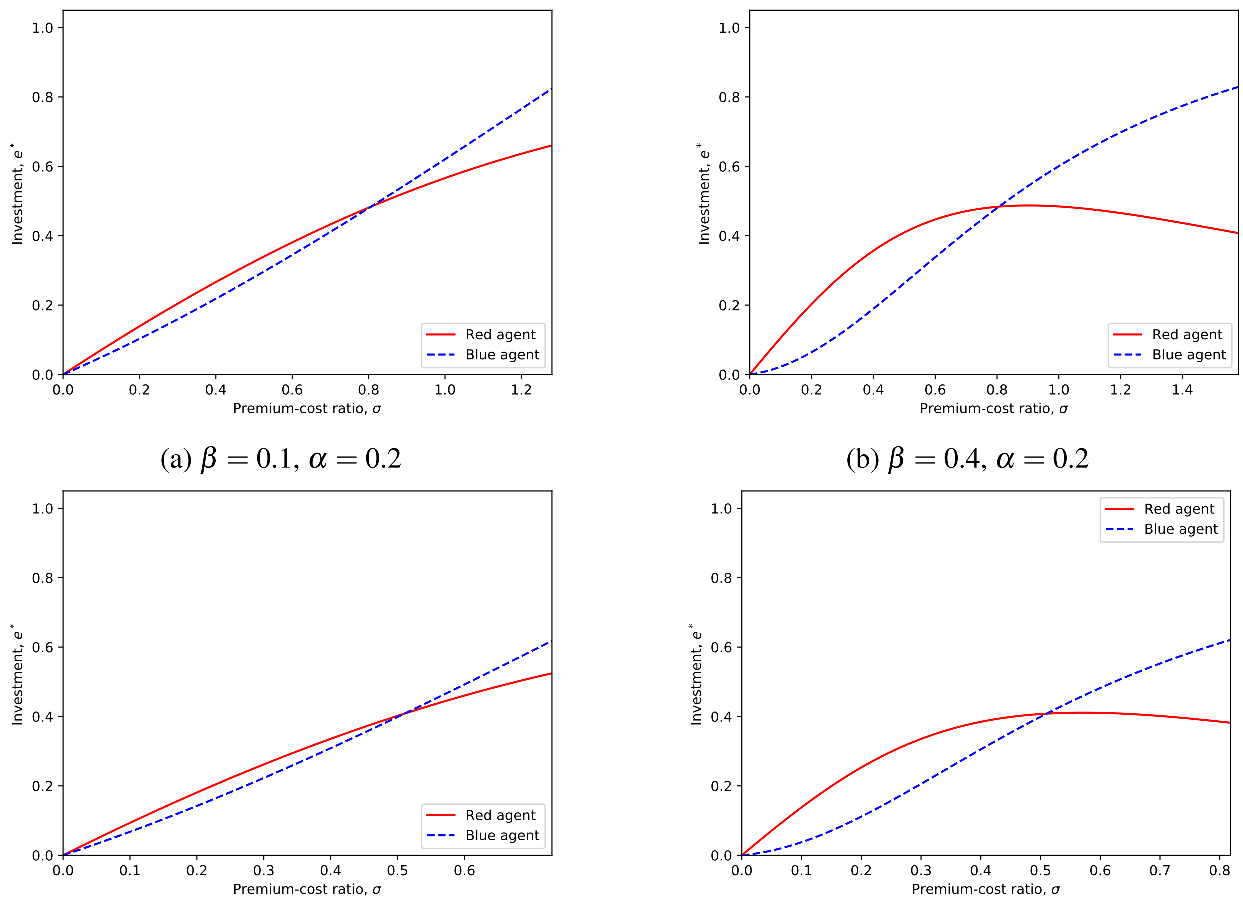
$$e_r^* = \sigma(1+\alpha)\frac{0.5 + \beta - 2\beta\sigma(0.5-\beta)(1+\alpha+\alpha^2)}{1 + 4\sigma^2\beta^2(\alpha^2+\alpha+1)^2} \tag{OA.38}$$

The equilibrium investment levels of Blue and Red for $\alpha \in (0,1)$ have the same functional form as in the main version of the model. Figure OA.13 shows the equilibrium investment levels of Blue and Red as functions of the premium-cost ratio, σ, for two levels of the subtle bias β and parameter α. As in the main text, in low-stakes situations, the overcompensation effect dominates and in high-stakes situations, the discouragement effect dominates. Therefore, firms that offer low-stakes careers (or less human-capital-intensive) find it profitable to engage in subtle discrimination in order to benefit from the overcompensation effect. In contrast, firms that offer high-stakes careers (and are more human-capital-intensive) prefer not to discriminate.

OA.9.2 Principal's problem

Here, we show that even under a finer skill assessment, the principal prefers high levels of subtle discrimination in cases of low productivity and low levels of subtle discrimination in cases of high productivity. We assume that upon promoting a moderately skilled agent with $s_i = 0.5$, the principal's profit increases by H, as stated in the main text. Similarly, when promoting a highly skilled agent with $s_i = 1$, the profit increases by μH, where $\mu \geq 1$. Then, the principal's profit is:

$$\max_{w_l \geq 0, w_l + W \geq 0} l + \mu H \cdot P(s_{b \vee r} = 1) + H \cdot P(s_{b \vee r} = 0.5) - W - 2w_l \tag{OA.39}$$



(a) $\beta = 0.1$, $\alpha = 0.2$

(b) $\beta = 0.4$, $\alpha = 0.2$

(c) $\beta = 0.1$, $\alpha = 0.6$

(d) $\beta = 0.4$, $\alpha = 0.6$

Figure OA.13: Equilibrium investments of blue and red agents, e_b^* and e_r^*, as functions of the premium-cost ratio, σ, for different values of the subtle bias β and the parameter α.

When $w_l = 0$ and the solution is interior, that is $\sigma \leq \bar{\sigma}(\alpha,\beta)$, we can rewrite the principal's profit maximization problem as:

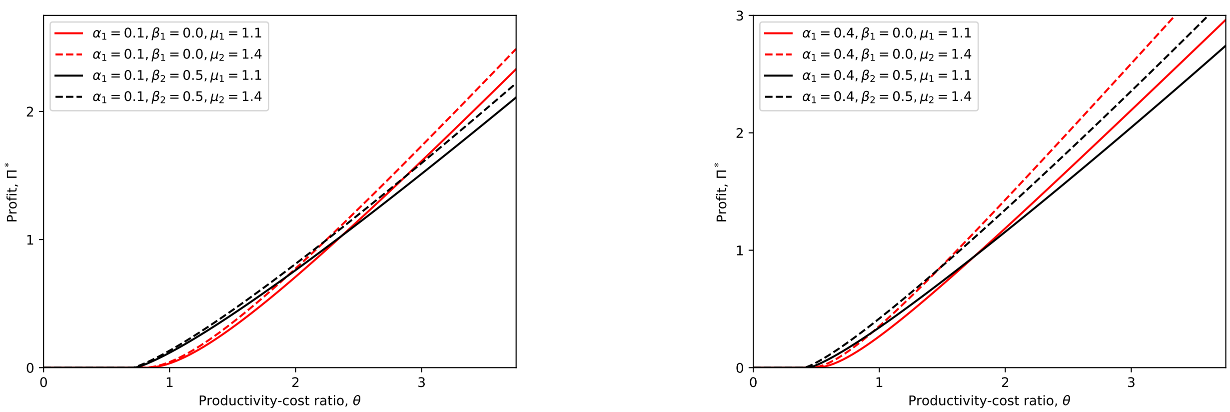
$$\pi(\alpha,\mu,\theta,k) = \max_{\sigma\in[0,\bar{\sigma}(\alpha,\beta)]} k\Big[\mu\theta P(s_{b\vee r}=1) + \theta P(s_{b\vee r}=0.5) - \sigma\Big], \tag{OA.40}$$

where $P(s_{b\vee r}=1) = \alpha e_b + \alpha e_r - \alpha^2 e_b e_r$ and $P(s_{b\vee r}=0.5) = e_b + e_r - e_b e_r(1+2\alpha)$, subject to

$$e_b(\sigma,\alpha,\beta) = \sigma(1+\alpha)\frac{0.5-\beta+2\beta\sigma(0.5+\beta)(1+\alpha+\alpha^2)}{1+4\sigma^2\beta^2(1+\alpha+\alpha^2)^2}; \tag{OA.41}$$

$$e_r(\sigma,\alpha,\beta) = \sigma(1+\alpha)\frac{0.5+\beta-2\beta\sigma(0.5-\beta)(1+\alpha+\alpha^2)}{1+4\sigma^2\beta^2(1+\alpha+\alpha^2)^2}. \tag{OA.42}$$

Figure OA.14 shows the optimal principal's profit as a function of the productivity-cost ratio θ, for different levels of subtle bias ($\beta_1 = 0$ and $\beta_2 = 0.5$) and different levels of the parameters α and μ. As we have shown in the main text, for low productivity (low-θ firms), the principal is better off under high levels of subtle discrimination ($\beta = 0.5$), while for high levels of productivity (high-θ firms), she is better off under no discrimination ($\beta = 0$). Note that the region where high subtle discrimination is optimal from the firm's point of view is decreasing in α. In other words, subtle discrimination becomes less profitable for firms as their workers obtain high skill with a greater chance, conditional on an effort level. Figure OA.15 illustrates this insight. It depicts the threshold productivity-cost ratio θ' such that for $\theta < \theta'$, the principal's profit is maximized under the highest level of subtle discrimination, $\beta^{pm} = 0.5$ and for $\theta > \theta'$ it is maximized under no subtle discrimination, $\beta^{pm} = 0$ as a function of parameter α and for two different values of parameter μ. Note that θ' decreases in both α and μ.



(a) $\alpha = 0.1$

(b) $\alpha = 0.4$

Figure OA.14: Equilibrium profit as a function of the premium-cost ratio, σ, for different values of the subtle bias β and the parameters α and μ.

OA.9.3 Principal's choice of skill assessment precision

In this subsection, we investigate under what conditions the principal prefers a more granular skill partition. For example, if an agent is highly skilled, $s_i = 1$, the principal might need to pay an extra cost to distinguish such agent from a skilled agent $s_i = 0.5$.

We assume that promoting a moderately skilled agent increases the principal's profit by H, while promoting a highly skilled agent increases his profit by μH, where $\mu > 1$. Below we compare the principal's profit when agents' skill can take three levels: $s_i = \{0, 0.5, 1\}$, when the principal distinguishes between all the three levels and when he distinguishes only skilled ($s_i = 0.5$ and $s_i = 1$) and non-skilled agents ($s_i = 0$).

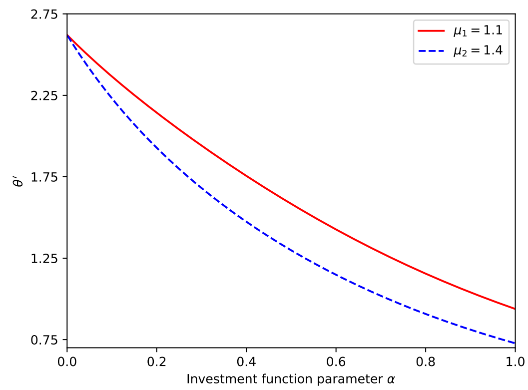


Figure OA.15: Threshold θ' as a function of α for different values of the parameter μ.

We solve the principal's problem under no subtle discrimination. First, the principal's profit when he recognizes three levels of skill is given by Eq. (OA.40 - OA.42), where $\beta = 0$. Second, let us consider a scenario wherein the principal is aware of the existence of highly productive agents with $s_i = 1$, but he lacks the ability (or intentionally chooses not) to differentiate between individuals possessing high skill levels ($s_i = 1$) and those with medium skill levels ($s_i = 0.5$). In this case, agent i faces a problem similar to the one in Eq. (4) of the main text:

$$\max_{e_i} \sigma \left[(1+\alpha)e_i(1-(1+\alpha)e_{-i}) + \frac{1}{2}\left(1-(1+\alpha)e_i-(1+\alpha)e_{-i}+2(1+\alpha)^2 e_i e_{-i}\right) \right] - \frac{e_i^2}{2}, \tag{OA.43}$$

because when an agent chooses effort level e_i, she becomes skilled (either moderately skilled or highly skilled) with probability $e_i + \alpha e_i$. With no subtle discrimination, the reaction functions are the same as when the principal can distinguish between all three skill levels:$e_b^* = e_r^* = \frac{\sigma}{2}(1+\alpha)$.

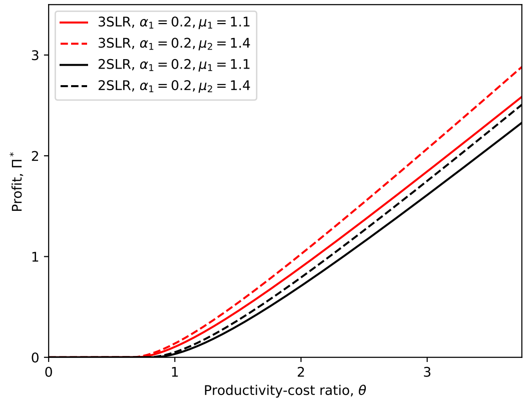
The principal's problem in this case is

$$\pi(k,\theta,\mu,\alpha) = \max_{\sigma \in [0,\bar{\sigma(\alpha)}]} k\left[\theta\left(e_b + e_r - e_b e_r\right)\left(\mu \frac{\alpha}{1+\alpha} + \frac{1}{1+\alpha}\right) - \sigma \right], \tag{OA.44}$$

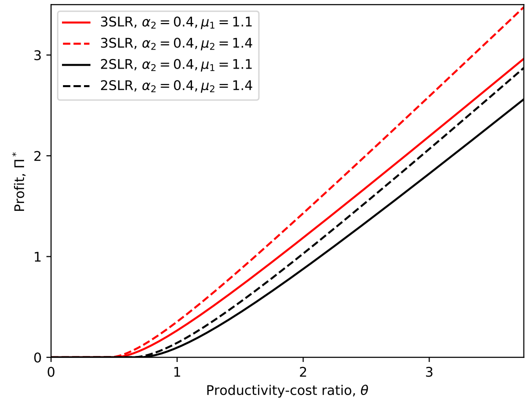
subject to $e_i = \frac{\sigma}{2}(1+\alpha)$.

Figure OA.16 shows the principal's optimal profit for cases when she can distinguish between three skill levels (red lines) and for cases when she can only distinguish between two skill levels (black lines). We plot the principal's profit for several values of α and μ.

The results in Figure OA.16 show that the principal's profit is higher when he possesses an ability to distinguish between all three skill levels. As a result, the principal would be willing to pay a fixed cost in order to obtain this ability. Importantly, the principal's willingness to pay is increasing in the productivity-cost ratio θ. That is, in low-productivity firms, the principal may decide not to separate between moderately and highly skilled agents if such separation is



(a) $\alpha = 0.2$



(b) $\alpha = 0.4$

Figure OA.16: Equilibrium profit as a function of the productivity-cost ratio, θ, for different values of μ and α. Red lines correspond to cases when the principal distinguishes between three skill levels, while black lines correspond to cases when the principal cannot distinguish between medium and highly skilled agents at the promotion stage.

costly. For example, one might need to conduct additional rounds of interviews to identify highly skilled candidates. If promotion of highly skilled candidates is not profitable enough, the firm may choose to conduct fewer rounds of interviews and to "pool" moderately and highly skilled agents in promotions.

OA.10 Welfare analysis

OA.10.1 Summary

Subtle discrimination simplifies welfare and policy analyses because welfare comparisons are not confounded by the direct effects of biases on the principal's utility. This property allows our model to produce sharper welfare and policy implications. Here we show that, for moderate to high stakes, subtle discrimination may harm everyone: the favored agent, the unfavored agent, and the firm. However, and perhaps surprisingly, for sufficiently low stakes, everyone may benefit from subtle discrimination. This result arises because low-productivity firms use biased contests as a means to incentivize agents.

In this section, we address a number of normative questions: How does the equilibrium compare with the first-best? What are the welfare implications of subtle discrimination? When is subtle discrimination inefficient?

OA.10.1.1 Comparison with the first best

It is instructive to compare the equilibrium effort levels for a given W to their first-best counterparts. For $\beta > 0$, there is typically no contract that implements the first-best investment levels. If $H > k$, the first-best outcome is $e_b^{FB} = 1$ and $e_r^{FB} = 0$. This outcome is unachievable under subtle discrimination: From Proposition 1, to have $e_b^* = 1$ we need $\sigma \geq \overline{\sigma}$, in which case we have $e_r^* = \min\left\{\frac{\sigma(1-2\beta)}{2}, 1\right\} > 0$ (because $\beta < 0.5$ if $\overline{\sigma}$ is finite). If $H \leq k$, the first-best requires both agents to invest $\tilde{e} = \frac{H}{H+k}$. But agents' investments are the same if and only if $\sigma = 1$, in which case we have $e_r^* = e_b^* = 0.5 \geq \tilde{e}$. Thus, except for the case in which $H = k$, there is no σ that implements the first-best investment levels in the presence of subtle bias ($\beta > 0$).

Things are different under no subtle bias ($\beta = 0$). If $H \leq k$, the first-best can be achieved by choosing $\sigma^{FB} = \frac{2H}{H+k}$ (i.e., $W^{FB} = \frac{2kH}{H+k}$). If $H > k$, the first-best cannot be achieved.

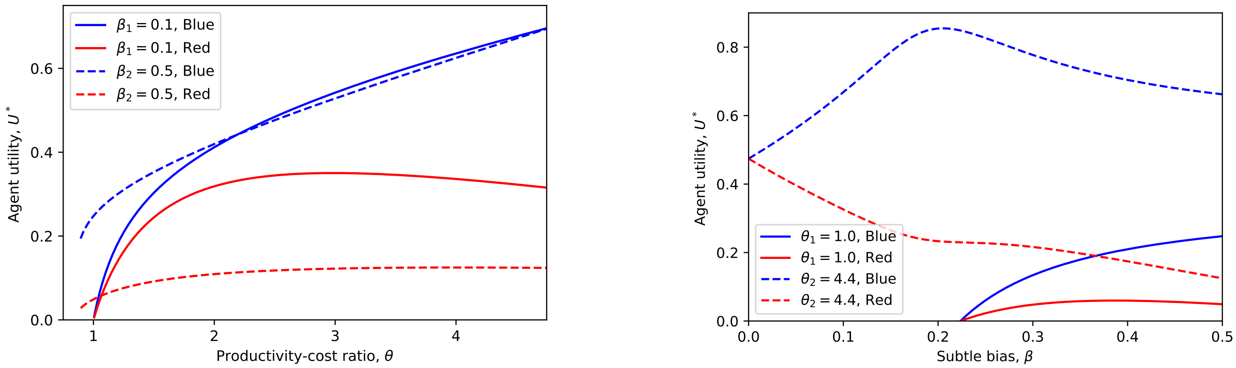
To summarize: (i) if the principal is subtly biased, there is no contract that implements the first-best outcome, except for the (measure-zero) case in which $H = k$; (ii) if the principal is unbiased, the first-best outcome can be implemented by a suitably-designed promotion contest if and only if $H \leq k$. The comparison with the first-best shows that subtle discrimination is a friction. Without a subtle bias, the first-best can sometimes be achieved. If there are additional contractual frictions, subtle discrimination can nevertheless be welfare-enhancing in some cases, as we show next.

OA.10.2 When does discrimination harm workers?

Figures OA.17a and OA.17b show the utilities of blue and red agents as functions of the productivity-cost ratio, θ, and the subtle bias, β, when the contract $\sigma(\theta, \beta)$ is optimally chosen by the principal. Two features are worth highlighting.

First, a stronger bias is not always beneficial to Blue. For high θ, increasing the bias may decrease Blue's utility. How could a bias in favor of blue agents harm these exact agents? A more biased principal offers lower stakes, reducing the benefits of promotion. As the figure shows, this dampening of incentives can offset Blue's gains from a higher bias. Thus, since profits may decrease with the subtle bias, there exist regions in which reducing the bias is a strict Pareto improvement, even in the absence of side transfers.

Second, there exists a region (for small values of θ) where the red agent prefers more discrimination to less. Therefore, for low levels of the productivity-cost ratio, all (the principal and both agents) prefer more discrimination to less. This result highlights that players at different layers of the corporate hierarchy, as well as in different industries, are heterogeneous in their preferences with respect to anti-discriminatory policies. While in positions or industries where productivity gains upon promotion are high everyone may benefit from decreased discrimination, this is not always the case in positions or industries with low productivity gains.



(a) $U^*(\theta)$, for $\beta_1 = 0.1$ and $\beta_2 = 0.5$.

(b) $U^*(\beta)$, for $\theta_1 = 1.0$ and $\theta_2 = 4.4$.

Figure OA.17: Agents' utilities, U^*, under optimal contract $\sigma(\theta, \beta)$.

OA.10.3 Social surplus

Figure OA.18 presents the level of subtle bias that maximizes the total social surplus, S, as a function of the productivity-cost ratio, θ. The relationship between subtle bias and social surplus is complex. There are three regions. In the first region, low-θ firms benefit from high subtle biases because the overcompensation effect helps to incentivize red agents. As we see from Figure OA.17b, for sufficiently low values of θ both Blue and Red benefit from increasing the bias.[7] In the second region, Red no longer benefits from the bias and, eventually, the discouragement effect becomes dominant, thus the firm also prefers a lower bias. Thus, for firms with intermediate levels of θ, the social-surplus-maximizing bias is $\beta = 0$. In the third region, Blue's utility is hump-shaped in the subtle bias (see Figure OA.17b), while the firm's profit is relatively flat in β. The optimal bias trades off the gains and losses to the agents. The socially-optimal bias is increasing in the productivity-cost ratio because discouraging Red is efficient when Blue is more likely to win, as it reduces the deadweight costs of effort duplication.

OA.11 Quotas

OA.11.1 Summary

We use our model to investigate the consequences of a hard quota aimed at protecting the unfavored group. We show that the quota has its desired effect only when the bias is sufficiently high. Even in that case, despite the fact that the quota implements equality of outcomes, unfavored agents still fare worse in terms of expected utility than favored agents. This result is explained by firms

[7]Note that the bias itself does not directly affect utilities. Thus, our welfare results fundamentally differ from those of models with non-subtle biases. For example, in Prendergast and Topel (1996), an increase in bias directly benefits supervisors.

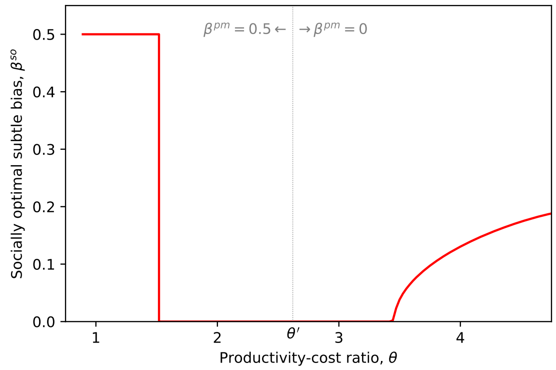


Figure OA.18: Socially optimal level of subtle bias as a function of θ.

reducing their promotion stakes under the quota, which leads to unfavored agents working harder than favored agents.

We have different results for soft (i.e., voluntary and non-binding) quotas. While potentially quite effective at curbing subtle discrimination, our results show that only high-productivity firms choose to implement soft quotas.

OA.11.2 Introducing a hard quota

The analysis in Section 3.5 of the main text reveals that not all firms would voluntarily take steps towards reducing subtle biases. At the same time, the welfare analysis shows that reducing subtle biases is sometimes socially desirable. Thus, it is instructive to consider possible interventions aimed at reducing or eliminating subtle discrimination.

Setting a (hard) quota is a popular policy tool to tackle a lack of diversity at top positions. Quotas are unlikely to deliver efficiency gains in our model, for two reasons: they constrain the principal's maximization problem and directly interfere with the agents' incentives to invest. Nevertheless, quotas may be a policy option for reasons other than efficiency, such as equity and fairness.

To consider quotas at the firm level, we extend the model as follows. At Date 0, the firm has a continuum of vacancies for job 1, with mass 2μ, and for job 2, with mass μ. Each worker in job 1 competes with exactly one worker for promotion and all pairs of workers are mixed (one red and one blue). In equilibrium, the probability that an agent of type i is promoted, p_i, is also the proportion of agents of type i found in job 2 at the end of the game. A quota is a target for p_i or, equivalently, a target for the promotion gap, Δp. For convenience we use the latter, thus a quota is fully described by a number $q \in [-1, 1]$.

Without loss of generality, we assume that the quota's goal is to reduce the promotion gap, that is, to promote more red agents: $q < \Delta p_0$ (the *pre-quota promotion gap*). Our interpretation is that the principal designs a firm-wide promotion policy, which is then implemented by a mass μ of supervisors, one for each pair of workers in job 1. We assume that supervisors have incentives aligned with the firm but are subtly biased. Here, unlike in Subsection 3.5, the firm cannot choose the bias of its supervisors. Because only they observe the skill s_i of their pairs of subordinates, any rule that allows supervisors discretion can be abused. Thus, the only way to comply with the quota is to force some supervisors to promote red agents regardless of skill. To do so, the principal offers a proportion δ of supervisors discretion over promotion decisions and forces a proportion $1-\delta$ of supervisors to promote only red agents.

The principal chooses δ to maximize profit subject to the quota constraint, $\Delta p = q$. The principal has two options: he can reveal the identities of the "constrained" and "unconstrained" supervisors to their subordinates, or he can keep them secret. For brevity, we only consider the full disclosure case.[8] The principal's problem is

$$\Pi(\beta,\theta,q) = \max_{\sigma\in[0,\overline{\sigma}(\beta)],\delta\in[0,1]} \delta\theta\left(e_b+e_r-e_b e_r\right)-\sigma, \tag{OA.45}$$

subject to

$$e_b = \frac{\sigma(0.5-\beta)+2\beta\sigma^2(0.5+\beta)}{1+4\beta^2\sigma^2};\ e_r = \frac{\sigma(0.5+\beta)-2\beta\sigma^2(0.5-\beta)}{1+4\beta^2\sigma^2}; \tag{OA.46}$$

$$\Delta p \equiv \delta\{e_b-e_r+[e_b e_r+(1-e_b)(1-e_r)]\,2\beta\}-(1-\delta)=q, \tag{OA.47}$$

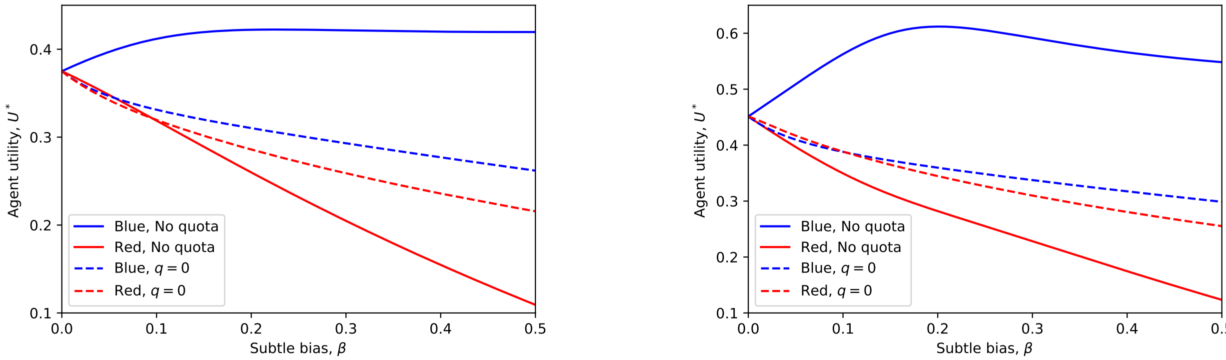
where the last equation is the quota constraint: the promotion gap must be q.

Firm profit is always higher when there is no quota or if the quota is not binding (i.e., $q=\Delta p_0$) because the quota constrains the principal's maximization problem. Still, there might be reasons to support quotas on grounds of redistributive equity. The key question is then: when do discriminated agents benefit from quotas?

Figure OA.19 shows the utilities of Blue and Red under a 50% quota (i.e., $q=0$). As expected, the quota typically reduces Blue's utility and increases Red's utility. However, for low biases, the quota may reduce Red's utility. This counter-intuitive result occurs because, under a quota, the firm offers smaller stakes. The negative effect dominates when the bias is small because, in this case, Red's probability of promotion increases by only a small amount after the quota.

We also see that the favored agent is typically better off than the discriminated agent, even when the quota imposes full parity. For Red to do better than Blue under a quota, the bias must be small and productivity θ must be high.

[8] The no disclosure case yields similar results.



(a) For $\theta_1 = 2.0$, $\beta = 0$ is socially optimal and $\beta = 0.5$ is profit maximizing

(b) For $\theta_2 = 3.2$, $\beta = 0$ is both socially optimal and profit maximizing

Figure OA.19: Agents' utilities as functions of subtle bias under no quota and under a fully disclosed quota, $\Delta p = q = 0$, for $\theta_1 = 2.0$ and $\theta_2 = 3.2$.

OA.11.3 Soft quotas

Firms may be able to achieve their diversity goals through voluntary actions, such as the adoption of a *soft quota* (or "soft affirmative action," as in Fershtman and Pavan (2021)). Rather than setting a strict numeric target, we can think of soft quotas as a recommendation to promote more red agents whenever possible. Suppose that, to implement a soft quota, the firm adopts a policy in which a supervisor pays a (vanishingly) small cost κ every time they promote a blue agent. For example, the supervisor needs to write a report explaining why the blue agent was more qualified than the red agent. As long as κ is sufficiently small and supervisors have strong incentives to maximize firm profit, the soft quota would only affect supervisors' behavior in tie-breaking situations. What types of firms adopt soft quotas? The answer follows from Proposition 7:

Corollary 1. *The firm adopts a soft quota that incentivizes the promotion of red agents if and only if* $\theta \geq \theta'$.

References

FERREIRA, D. AND R. NIKOLOWA (2023): "Talent discovery and poaching under asymmetric information," *The Economic Journal*, 133, 201–234.

FERSHTMAN, D. AND A. PAVAN (2021): ""Soft" affirmative action and minority recruitment," *American Economic Review: Insights*, 3, 1–18.

GREENWALD, B. (1986): "Adverse Selection in the Labour Market," *Review of Economic Studies*, 53, 325–347.

PRENDERGAST, C. AND R. H. TOPEL (1996): "Favoritism in organizations," *Journal of Political Economy*, 104, 958–978.

WALDMAN, M. (1984): “Job assignments, signaling, and efficiency,” *Rand Journal of Economics*, 15, 255–267.

About ECGI

ECGI is an international non-profit membership organisation that provides a platform for debate and dialogue among academics, policymakers, and business leaders, with a focus on major corporate governance, ESG, and stewardship issues.

The views expressed in this working paper are those of the authors, not those of the ECGI or its members.