Subtle Discrimination*

Elena S. Pikulina Daniel Ferreira

March 2023

Abstract

We propose a theory of *subtle discrimination*, defined as biased acts that cannot be objectively ascertained as discriminatory. We present a model in which candidates compete for a promotion. When choosing among equally qualified candidates, the principal subtly discriminates by breaking ties in favor of candidates from a particular group. Subtle discrimination matters because it affects decisions to invest in human capital. The model predicts that discriminated agents perform better in low-stakes careers while favored agents perform better in high-stakes careers. In equilibrium, firms are polarized: high-productivity firms strive to be "progressive" and have diverse top management teams, while low-productivity firms prefer to be "conservative" and have little diversity at the top.

Keywords: Discrimination, human capital, firm-specific skills, promotion

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 Murray Carlson, Jack Favilukis, Lorenzo Garlappi, Will Gornall, Sebastian Gryglewicz, Alex Imas, Alesio Piccolo, and Maddalena Ronchi 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, 2022 Annual Pacific Northwest Finance Conference, 2022 North-American Economic Science Association Conference, and 2023 American Finance Association Meeting for valuable comments.

1 Introduction

Often today the bias is just subtler, the attitudes more hidden, the rationalization more nuanced. Exclusions show up in forms that are harder to prove but continue to keep workplaces homogeneous. It's often so subtle that those in power find it hard to see, harder to acknowledge, and impossible to fix, in spite of all the stories, the data, and the research making it clear that the problem is very real. (Pao 2017, p. 9).

Social and organizational psychologists describe 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, not praising the performance of minority employees, and overpromoting men to managerial positions when choosing among equally-qualified candidates. It is hard to substantiate claims of subtle discrimination. For a discriminatory act to be prosecuted in the United States, the discriminated party must provide direct evidence of intent to harm or deny rights, or prove a clear pattern of adverse events unexplainable on grounds other than discrimination. Partly due to the threat of legal action, acts of overt discrimination have become relatively rare. By contrast, nondiscriminatory narratives and plausible deniability may cloak subtle discrimination and make it a common but invisible phenomenon.[2]

Despite its prevalence, the impact of subtle discrimination on workers and firms has received little attention in the economics literature. Our paper is a first attempt at formalizing the notion of subtle discrimination. We define subtle discrimination as biased

[1] See, for example, Dovidio and Gaertner (1986, 2000), Essed (1991), Deitch et al. (2003), Dipboye and Halverson (2004), Noh et al. (2007), Hebl et al. (2008), Van Laer and Janssens (2011), Jones et al. (2017), Dhanani et al. (2018), and Hebl et al. (2020). While studies often use alternative terms, such as "modern discrimination," "aversive discrimination," "everyday discrimination," "ambivalent discrimination" and "covert discrimination," 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 favourably for plaintiffs than other kinds of civil cases, including *habeas corpus* cases (Selmi (2000); Clermont and Schwab (2009)).

acts that cannot be objectively ascertained as discriminatory. That is, a single act of subtle discrimination leaves no hard evidence, allowing agents who subtly discriminate to rationalize their actions. 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. Finally, to conceal a small bias towards a specific group, a manager may say that he uses subjective criteria (e.g., "potential") to select between two candidates with the same objective qualifications and performance records.

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 firm-specific human capital. Labels are payoff-irrelevant; profit depends only on the promoted agent's acquired skills. The principal has a small bias in favor of the blue agent. Because the bias is small, when the red agent is objectively more qualified than the blue candidate, the principal prefers to promote the former. That is, the principal does not *overtly discriminate*. However, when both candidates are equally qualified, the principal is more likely to promote the blue candidate. Promoting the blue agent with a higher probability is an act of discrimination. This discriminatory act is subtle because one cannot observe the tie-breaking rule that the principal uses. Thus, in the context of the model, subtle discrimination is an inability or unwillingness to break ties fairly.[3]

If the principal's bias is such that discrimination occurs *only* in tie-breaking situations, we call it a *subtle bias*. While a non-subtle bias can cause both subtle and overt discrimination, a subtle bias can cause only subtle discrimination. Our model shows that subtle biases can have significant consequences for firms and workers.

Our notion of subtle bias accords well with the social psychology literature on dou-

[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. At times, individuals behave in discriminatory ways, and at other times they demonstrate their egalitarian views.

ble standards in competence assessment, which shows evidence of biased tie-breaking in competitive settings (Foschi et al. (1994)). A subtle bias can be caused by an arbitrarily small preference bias or by biased beliefs and stereotypes (see, e.g., Reuben et al. (2014) and Bordalo et al. (2016)). Alternatively, a subtle bias may be a manifestation of an implicit (i.e., unconscious) bias. Thus, decision-makers may not be aware of their own bias; subtle discrimination might not be intentional or controllable.[4]

Our analysis relies on ties being unexceptional. In practice, candidates' ties in qualifications are ubiquitous because evaluation scales are often discrete.[5] Frederiksen et al. (2017) show empirically that performance scales tend to be restricted, with five- or six-point scales being the norm. Ties are also likely when candidates' qualifications are assessed across several domains and candidates excel in different areas, i.e., there is no clear winner across all relevant qualifications. Similarly, ties are likely when candidates' scores are aggregated across several decision-makers (such as the members of a hiring committee), even if individual members avoid ties when ranking candidates. When ties occur, they are often broken based on subjective criteria, allowing subtle biases to affect decisions.[6]

Our paper makes three primary contributions. First, we propose a classification of discriminatory acts into two categories: overt and subtle. While our particular formalization necessarily leaves out important details, it is simple, intuitive and tractable. Second, in a model of promotions, we show that subtle discrimination and overt discrimination have different empirical predictions. Third, our model of subtle discrimination in pro-

[4]Psychologists define implicit (biased) attitudes as associations between an individual (or a social group) and specific attributes when those associations are partially or entirely outside of a person's awareness (Greenwald et al. (2002); Gawronski (2019)).

[5]In models of lexicographic decision making (Tversky (1969); Manzini and Mariotti (2012)), ties can arise with positive probability even when the assessment criteria are continuous; see the discussion in Section 2.

[6]A leading example of the relevance of "tie-breaking" is academic co-authorship. 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 trying to attribute credit for joint work. See Heilman and Haynes (2005) for further evidence of gender bias in team credit attribution.

motions 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. For example, denying someone a job *because* of their race or making derogatory comments about a person's identity are overtly discriminatory acts because they are clearly based on a person's group identity. In contrast, when two candidates are indistinguishable in their qualifications, passing over an agent for a promotion is not clear evidence of discrimination. In our model, when blue and red agents are equally skilled, the principal can rationalize (to others or to himself) the act of promoting the blue agent as an unbiased random choice, akin to flipping a coin. One would need to observe a (potentially infinitely) long series of discriminatory acts to establish subtle discrimination with certainty in such situations. In contrast, promoting an unskilled agent from a favored group over a skilled agent from an unfavored group is clear evidence of discrimination; we classify such acts as overt discrimination.

In our model, agents respond to the principal's subtle bias and to each others' decisions when investing in their human capital. In equilibrium, agents differ in their investment decisions, which creates an *achievement gap*, i.e., a difference in accumulated human capital and obtained 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 being promoted.[7] 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*. Unlike the discouragement effect, the overcompensation effect can dominate only when discrimination is subtle rather than overt.

We show that the sign and the magnitude of the achievement gap depend on the stakes

[7] See Coate and Loury (1993) for an early model of underinvestment in human capital in the context of statistical discrimination.

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 *more* 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 *less* 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. 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, and Azmat et al. (2020) for lawyers. Promotions in such careers are typically associated with large increases in pay and non-pecuniary benefits, such as prestige and status. 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. By contrast, Benson et al. (2021) find that women in management-track careers in retail have better (pre-promotion) performance than men.[8] 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 increasing promotion gaps at the top of hierarchies, a fact that is known as the

[8] Despite women's better performance, supervisors still consider men to have higher "potential" on average, which leads to higher promotion rates for men.

"leaky pipe" phenomenon (Lundberg and Stearns (2019); Sherman and Tookes (2022)).

To shape their subtle biases, we allow firms to adopt different anti-discrimination (i.e., diversity and equity) policies. Firms can choose to become more progressive (i.e., less biased) or conservative (i.e., more biased). 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. For example, firm polarization in diversity preferences is a potential explanation for the evidence that large and well-performing firms have more women on their boards (Adams and Ferreira (2009)). In the context of the model, causality runs from firm quality to board diversity.

Our model offers a novel perspective on the costs and benefits of companies focusing on social issues, especially regarding the diversity of their workforces. While some argue that progressive values typically do not conflict with the pursuit of profit (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 promotable tasks. Moreover, the model shows that such benefits accrue primarily to firms with high-stakes career paths. By 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.

Economists traditionally classify discriminatory acts based on their source rather than transparency. Some view discriminatory acts as a consequence of rational statistical discrimination (Phelps (1972); Arrow (1973)). A second view is that discrimination is caused by biases, such as biases in preferences or tastes (Becker (1957)), beliefs (Bordalo et al. (2016); Bohren et al. (2019a)), or incentives (Dobbie et al. (2021)). Empirically, the gold standard for separating these two views is the "Becker marginal outcome test" (Becker (1957, 1993)).[9] Consider a firm where women consistently have lower promotion rates than men. If rational statistical discrimination causes a promotion gap, all else constant, marginally promoted men and women should have similar performances after promotion. Thus, if we observe marginally-promoted women performing better than marginally-promoted men, we can conclude that biases cause the promotion gap.[10] Unlike the biases in taste-based or stereotype models, subtle biases cannot be detected by the Becker outcome test. If workers with similar qualifications are close substitutes, a slight bias towards one group does not negatively impact a firm's profit. Thus, our model implies that subtle discrimination should feature alongside statistical discrimination as the null hypothesis in Becker outcome tests in promotion contexts.

2 Related Literature

Our model setup is similar (in spirit, but not in its details) to that of Prendergast (1993). Prendergast (1993) proposes a model of promotions in which the firm cannot contractually commit to compensating workers for acquiring firm-specific human capital. Our model differs from his in two significant aspects. First, in our model, promotions are competitive, i.e., candidates compete for a limited number of positions. Second, in our model, the principal is subtly biased in favor of candidates from a particular group.

Our model relates to the literature on favoritism and other biases in subjective perfor-

[9] Alternatively, Bohren et al. (2019b) show how to test for the source of discrimination by analyzing the implications of a dynamic model of discrimination.

[10] For empirical applications of the Becker outcome test in the context of promotions, see Benson et al. (2021) and Huang et al. (2022).

mance 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. By contrast, in our model, favoritism matters only because it affects ex-ante incentives.

More broadly, our study is related to the theoretical literature on discrimination (see Arrow (1998), Fang and Moro (2011), and Lang and Lehmann (2012) for reviews). In their seminal work on affirmative action, Coate and Loury (1993) show that negative stereotypes can be self-fulfilling because discriminated agents may not undertake investments that make them more productive. Similarly, in our model, discrimination may discourage some agents from investing. However, because workers compete for the same position, their investment decisions are interdependent. We show that 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. We show that this result obtains only when discrimination is subtle. In addition, the strategic interactions between agents imply a unique equilibrium, which is rare in the literature on self-fulfilling discrimination.[11]

In our model, agents impose externalities on each other. In this sense, our model is similar to those by Mailath et al. (2000) and Moro and Norman (2004), who study integrated labor markets where workers from one group impose externalities on another group. In both models, asymmetric equilibria exist in which agents with identical qualifications receive different wages. That is, discrimination is ex-post observable. By contrast, in our model, wages are not conditional on agents' labels, and therefore discrimination cannot be verified ex-post.

[11]Glover et al. (2017) provide evidence of a different form of self-fulfilling discrimination: workers perform worse when under the supervision of a biased manager.

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 is also related to a strand of the discrimination literature that focuses on bias amplification. Lang et al. (2005) show that in markets where firms post wages, weak discriminatory preferences can cause large wage differentials. 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 differences between groups. Differently from these models, in our model the source of bias amplification is the competitive nature of promotion tournaments. While agents can take actions that amplify the consequences of small biases, we show that these actions can also lead to the attenuation of such biases.[12]

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)). 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.

[12] Recently, Kline et al. (2022) show that only a small fraction of U.S. employers discriminate. However, these models suggest that even a small bias can have substantial consequences for the economy.

Although we do not model the preferences and beliefs at the root of subtle discrimination, we note that our notion of subtle discrimination is compatible with models of lexicographic preferences. In particular, our decision-making heuristic can be mapped into Tversky's (1969) notion of lexicographic semiorder; see also Manzini and Mariotti (2012) for a generalization. Consider a decision-maker that chooses between candidates (call them b and r) based on two criteria, s_1 and s_2. The decision-maker uses s_2 to separate the candidates if and only if s_1 cannot separate them. Crucially, the candidates may tie on the first criterion even when it is a continuous variable; a tie is declared when the difference between the two candidates is less than $\epsilon > 0$. In our model, the tie-breaking criterion (the candidate's label) is payoff-irrelevant. Thus, it can also be interpreted as a rationalization for the decision (e.g., b has higher "potential" than r), as in Cherepanov et al.'s (2013) theory of rationalization. That is, while the principal prefers b to r, choosing b is not rationalizable when r is clearly more qualified.

Our notion of subtle discrimination is closely related to (but also different from) Cunningham and de Quidt's (2022) concept of implicit preferences. They consider a setup in which a decision maker selects a woman over a man whenever both have identical qualifications but selects a man over a woman when their qualifications are mixed, i.e., they cannot be objectively ranked. Cunningham and de Quidt (2022) equate such choices to an explicit preference for women and an implicit preference for men. Applying our terminology to their example, we say that the decision-maker has a subtle bias towards women in the first case and a subtle bias towards men in the second. Thus, in their model, subtle biases depend on the nature of the tie (i.e., unambiguous versus ambiguous ties).

In the empirical literature, Hospido et al. (2019), Bosquet et al. (2019) and Azmat et al. (2020) provide evidence that, in high-stakes environments, women have lower promotion rates, partially because they are less likely to seek promotion in the first place. Our model suggests that such a discouragement effect can be a result of subtle discrimination, whose consequences are amplified in high-stakes careers. Moreover, several recent papers provide suggestive evidence of subtle discrimination. Benson et al. (2021) find that, despite having the same ratings on performance both before and after promotions, women consistently receive lower ratings on "potential" than men. When it comes to demotions, women are more likely to get fired than men for professional misconduct (Egan et al. (2022)). Finally, women also receive less credit for innovative behavior in the workplace (Luksyte et al. (2018)) and for work-related experience (Cziraki and Robertson (2021)).

Our results also speak to the literature on the gender gap in willingness to compete (Niederle and Vesterlund (2007); see also Niederle and Vesterlund (2011) for a review). Our model predicts that women are less willing to compete against men than against other women. Using a lab experiment, Geraldes (2020) shows that when given an opportunity to choose a competitor's gender, women are as likely to enter a competition as men are. According to our model, female unwillingness to compete with men should become stronger as the stakes increase. Using a high-stakes TV game show, Buser et al. (2023) show that women are less willing to compete against men.

3 A Model of Subtle Discrimination in Promotions

After presenting the setup in Subsection 3.1, in Subsection 3.2, we describe the first-best solution to serve as a benchmark. We then solve the model for an exogenously given compensation contract in Subsection 3.3. In Subsection 3.4, we let firms choose compensation contracts optimally. In Subsection 3.5, we endogenize the subtle bias.

3.1 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*). Both vacancies need to be filled. Red and Blue are payoff-irrelevant labels. To save on notation, we normalize the revenue generated by the agents on their entry-level jobs to zero.

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 in the sector. Explicitly modeling the market for workers would not change the main implications of the model, as long as candidate pools do not become fully segregated. That is, we implicitly assume that some frictions prevent firms from selecting only applicants of a single type. We leave extensions incorporating endogenous labor supply to future work.

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, which we call "skill." Skill is an observable but not verifiable binary variable: $s_i \in \{0,1\}$. Agent i's probability of acquiring the skill is e_i. Both agents are risk-neutral and have the same skill-acquisition cost function, $c(e_i)$, which we assume is strictly increasing and convex. That is, agent i's utility is $u_i = w_i - c(e_i)$, where w_i is the agent's monetary compensation. Without loss of generality, we set $c(0) = 0$.

We interpret skill broadly as any kind of observable evidence that predicts an agent's future performance. For example, in the legal profession, hours billed to clients and new client revenue raised are the main tools used to assess the performance of associates (Cotterman (2004), Heinz et al. (2005)). We assume that the skill is firm-specific in the sense that it is less valuable to agents who leave the firm. For example, a lawyer who raises significant revenue for her firm may not be able to credibly show her record to other firms. Firm-specific skill can also be interpreted as a unique weighted combination of general skills that is valuable for a particular firm or narrow industry (Lazear (2009)). For example, a manager who works for a firm that develops payroll software for businesses must know something about accounting, labor laws, tax laws, software and computer programming. While none of these skills is firm-specific, their unique combination is.

At Date 2, a decision-maker – whom we call the *principal* – chooses one of the agents to fill a top position (*job 2*) in the firm. The agent who is not promoted remains at the entry-level job. We assume that once the principal observes an agent's skill level, no further information is useful for predicting the agent's performance in job 2. In other words, an agent's skill is a sufficient statistic for the agent's expected productivity. Promoting an unskilled agent ($s_i = 0$) increases the principal's expected payoff by $l > 0$ while promoting a skilled agent ($s_i = 1$) 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. We interpret H as a firm characteristic. Larger H means that human capital is more important at higher hierarchical levels.

We interpret the assumption that $l > 0$ as indicating a preference for internal candidates for job 2, perhaps because of on-the-job learning. This assumption makes the model richer: the principal needs to break a tie between the two agents when they are both skilled or unskilled. If, instead, we were to assume $l < 0$, the principal would prefer to leave job 2 vacant when both agents are unskilled. This case would require the principal to break a tie only when both agents are skilled.

Although the principal cannot offer wages contingent on skill acquisition (because skill is not verifiable), 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 $\boldsymbol{w} = (w_1, W)$ 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 $\boldsymbol{w}$ to both agents. The principal uses the promotion contest to provide incentives for skill acquisition. Because $H > 0$, 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 agents, even when both agents are unskilled. As l is a free parameter in the model, we assume that it is sufficiently high so that the principal can credibly commit not to leave job 2 vacant.[13] Thus, the firm's (expected) profit under a profit-maximizing principal is $\Pi = l + H(e_b + e_r - e_b e_r) - 2w_1 - W$.

3.1.1 Subtle Discrimination: Definition and Discussion

We model subtle discrimination as a decision-making heuristic. When both agents have the same skill level ($s_i = s_{-i}$), both are equally productive at job 2 and, thus, the principal is indifferent between the two. Because only one agent can be promoted, the principal uses a tie-breaking rule. When ties occur, we assume that the principal promotes agent i with probability $\frac{1}{2} + \beta_i$, $i \in \{b, r\}$, where $\beta_b = -\beta_r$. We say that the principal is *subtly biased* in favor of Blue if $\beta_b = \beta > 0$. The principal's decision-making heuristic is thus equivalent to a lexicographic criterion: The principal always prefers the agent with the highest expected productivity; when there is a tie, the principal then relies on his (biased) "gut feeling." One interpretation is that the principal always rationalizes his choice (as in Cherepanov et al. (2013)) and, thus, a tie is not perceived as such.

Our notion of subtle discrimination is compatible with the principal being unconsciously biased. This interpretation is valid under the additional assumption that the principal does not benefit from promoting a particular candidate. If we assume that the bias does not enter the principal's utility, the bias would be imperceptible to a principal who is initially unaware of it. In practice, the principal might find it difficult to correct the bias (at least in a finite series of decisions) if he believes that his choices are unbiased.[14] Such unconscious biases are most likely to pertain to System 1 thinking, i.e., fast, automatic, and effortless associations (Kahneman (2011)).

Alternatively, our definition of subtle discrimination can be seen as a limiting case of (explicit) taste-based discrimination (Becker (1957)). Suppose that there is a continuum of

[13] Although we assume that w does not depend on labels, one could think of a situation in which different agents are assigned to different career tracks. If only one agent has a path to job 2, then no agent invests in skill acquisition. Thus, segregating agents into different career paths is never optimal.

[14] For example, Begeny et al. (2020) and Régner et al. (2019) find that decision-makers are more likely to favor men in their evaluations and promotion decisions if they do not explicitly believe in external barriers and biases faced by women in their professional fields.

principals. All of them maximize expected profit, but there is a fraction 2β who also derive incremental utility $\epsilon > 0$ from promoting blue agents. As $\epsilon \to 0$, the bias only affects decisions when agents have the same skill level. Thus, when workers are matched with a principal of unknown type, conditional on a tie, the blue agent expects to be promoted with probability $\frac{1-2\beta}{2} + 2\beta = \frac{1}{2} + \beta$.

In what follows, we do not take a stand on whether subtle biases are implicit or explicit; our model can accommodate either interpretation.

3.1.2 Subtle Discrimination versus Overt Discrimination

Our notion of subtle discrimination excludes discriminatory actions that are relatively easy to identify as such. We call such actions *overt discrimination*. In the context of our model, the principal overtly discriminates if he promotes an unskilled blue agent ahead of a skilled red agent. Such an act would be relatively easy to identify as discriminatory because agents' skill levels are observable (although not verifiable). To model overt discrimination, we assume that when $s_b = 0$ and $s_r = 1$, the principal promotes Blue with probability $\delta > 0$, where $\delta \in [0, 1]$ is a measure of *overt bias* in our model.

A person who overtly discriminates is also likely to subtly discriminate. Since, in our model, overt discrimination is costly, we expect subtle discrimination to be more pervasive than overt discrimination. Accordingly, we assume that $\beta \geq \frac{\delta}{2}$. To understand this assumption, consider the case in which explicit preferences drive both types of discrimination, i.e., some principals derive utility from promoting blue agents. Suppose there is a continuum of principals and let δ denote the proportion of principals with discriminatory tastes. When agents are matched with a principal of unknown type, conditional on $s_b = 0$ and $s_r = 1$, the blue agent expects to be promoted with probability δ. Conditional on a tie ($s_b = s_r$), the blue agent expects to be promoted with probability $\frac{1+\delta}{2}$, which implies $\beta = \frac{\delta}{2}$. That is, a tasted-based overt bias of size δ implies a subtle bias of at least $\frac{\delta}{2}$. Due to the costless nature of subtle discrimination, we would expect the subtle bias to be greater than this lower bound.

We will show that overt discrimination has different empirical implications from those of subtle discrimination (see Subsection 3.3.3). Otherwise, in the rest of the paper, we assume $\delta = 0$.

3.2 Benchmark: First-best Investment Levels

As a benchmark, we consider the problem of an unbiased social planner who maximizes total surplus. Define (expected) social surplus as $S = \Pi + E[u_b] + E[u_r]$. The planner's problem is

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

The planner faces a trade-off between effort duplication and effort sharing. On the one hand, asking both agents to invest in skill acquisition implies that, with positive probability, some acquired skills goes to waste. This waste is the cost of effort duplication. On the other hand, if only one agent invests, her marginal cost of effort is much 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 proofs are provided in Appendix A):

Proposition 1. *The first-best investment levels can take one of two forms: (i) $e_b^{FB} = e_r^{FB} = \tilde{e} < 1$ or (ii) $e_b^{FB} > 0$ and $e_r^{FB} = 0$ (or, equivalently, $e_b^{FB} = 0$ and $e_r^{FB} > 0$).*

Proposition 1 says that the first-best outcome can be symmetric or asymmetric. If the benefits from effort sharing are greater than the costs of effort duplication, the social planner forces both agents to choose the same investment level (Case (i)). If the costs of duplication outweigh the benefits from effort sharing, the social planner asks only one agent to invest in skill acquisition (Case (ii)).

Consider the case in which $c(e_i) = \frac{ke_i^2}{2}$. Then, if $H \leq k$, the first-best solution is

$$e_b^{FB} = e_r^{FB} = \tilde{e} = \frac{H}{H+k}. \tag{2}$$

That is, treating both agents equally is socially optimal. 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 ask only one of them to invest in skill acquisition.

To simplify the exposition, for the rest of the paper, we assume that the cost function is $c(e_i) = \frac{ke_i^2}{2}$. We choose to sacrifice generality to obtain analytical proofs for most results, which help explain the economic forces at play. Nevertheless, none of the results we emphasize depends on the cost function being quadratic; in the Internet Appendix, we replicate our main results for other convex cost functions.

3.3 Equilibrium under Exogenous Compensation Contracts

Here we describe the agents' investment choices under a fixed contract $\boldsymbol{w}$. We assume that the contract is individually rational; both Blue and Red accept the contract at Date 0. At Date 1, the agents simultaneously choose their investment (i.e., effort) 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 will be biased in favor of Blue. That is, in case of a tie, he promotes Blue with probability $\frac{1}{2} + \beta$.

3.3.1 Equilibrium Characterization

We define agent i's expected utility as:

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

In the agent's expected utility function, the first term is the baseline reward, the second term is the promotion premium times the probability of promotion, and the third term is the skill-acquisition cost. The first term inside the square brackets corresponds to the probability of agent i acquiring the skill when agent $-i$ fails to acquire the skill. In this case, the principal promotes agent i. The second and third terms correspond to the probability of promotion via a tie-breaking decision. That is, when both agents are either skilled or unskilled, the principal breaks the tie by flipping a "mental coin," which is biased in favor of Blue.

An agent's problem at Date 1 is to maximize his/her expected utility $U_i(\boldsymbol{e}, \boldsymbol{w})$ by choosing an investment level e_i taking the contract, $\boldsymbol{w}$, and the effort of the other agent, e_{-i}, as given:

$$\max_{e_i \in [0,1]} U_i(\boldsymbol{e}, \boldsymbol{w}). \tag{4}$$

Assuming an interior solution,[15] the agents' reaction functions are

$$e_b = \frac{W}{k}\left(\frac{1}{2} - \beta + 2\beta e_r\right), \tag{5}$$

$$e_r = \frac{W}{k}\left(\frac{1}{2} + \beta - 2\beta e_b\right). \tag{6}$$

Define $\sigma \equiv \frac{W}{k}$, i.e., the ratio of the promotion premium to the cost parameter. We call σ the *premium-cost ratio*. The premium-cost ratio is a reaction function shifter (see Eq. (5) and (6)). Higher σ implies a higher net marginal benefit of investment for any given pair (e_b, e_r). Intuitively, a high premium-cost ratio 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., a case in which there is much to gain from investing in skill acquisition. By contrast, if σ is low, agents benefit little from investing. In this case, we say that the agents are on a low-stakes career path. Thus, we also informally refer to σ as the "stake" of a career path.

[15] In Subsection 3.4, we show that optimal contracts always imply interior solutions.

In the baseline case with no subtle bias ($\beta = 0$), the reaction functions (5)-(6) are flat, implying that $e_b^* = e_r^* = \frac{\sigma}{2}$ is the dominant strategy. That is, if there is no bias, the agents choose their optimal investment levels without any strategic considerations. If $\beta > 0$, Blue's reaction function is positively sloped and Red's reaction function is negatively sloped. Intuitively, with a subtle bias in favor of blue agents, 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. By contrast, Red adopts the opposite strategy in an attempt to avoid ties.[16]

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:*

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

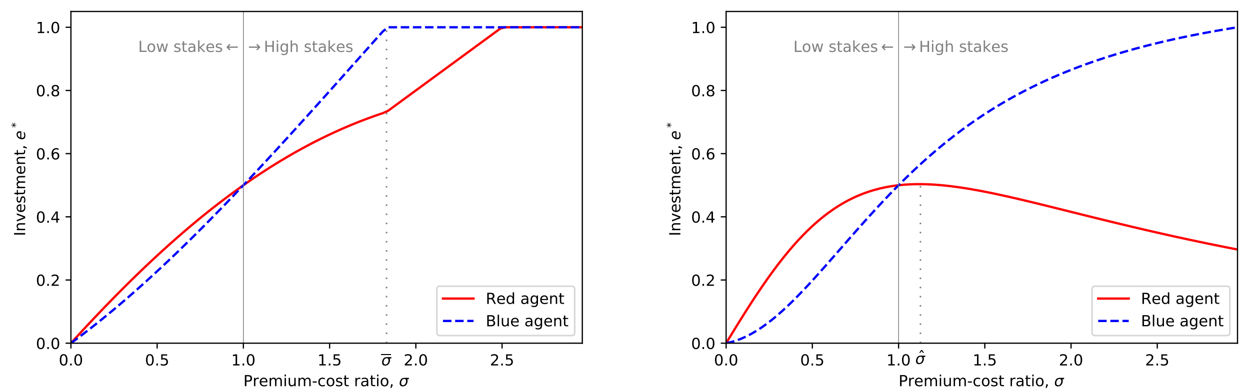
$$e_r^* = \frac{\sigma(0.5 + \beta) - 2\beta\sigma^2(0.5 - \beta)}{1 + 4\beta^2\sigma^2}. \tag{8}$$

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

3.3.2 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.

[16] If $l < 0$, i.e., the principal prefers not to promote unskilled agents, both reaction functions are negatively sloped, implying that investments are strategic substitutes.



(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 $\beta_2 = 0.4$.

The following corollary formalizes the comparative statics illustrated in Figure 1. For simplicity 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 skill acquisition in an attempt 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.[17] On the other hand, Red is discouraged from investing because her chances of promotion are slim even if she acquires the necessary skills. We call this force the *discouragement effect*.[18] Parameter σ determines what 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 in-

[17] Hengel (2022) provides evidence of the overcompensation effect in academic writing: papers written by women are better written than equivalent papers written by men.

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

vesting. At high investment levels, the probability of separation is low while the marginal cost of investing is high.[19]

Corollary 1 is empirically testable. 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 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. (2021) 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.

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

Remark 2. A potential consequence of the discouragement effect is that a principal who is unaware of his bias (and the strategic interaction it creates between the two agents) might interpret Red's behavior as a lack of interest in high-paying positions. In other words, he might incorrectly "learn" that red and blue agents have different preferences with respect to earned income. Such learning might further reinforce the principal's subtle bias or even result in an explicit bias in favor of blue agents.[20]

[19]If $l < 0$, the discouragement effect always dominates the overcompensation effect, and Red invests less than Blue in equilibrium. We focus on the $l > 0$ case because either effect may dominate in equilibrium.

[20]In order to do so, the principal could construct a narrative that is not contradicted by his observed data (see, for example, Eliaz and Spiegler (2020)) rather than engage in rational Bayesian learning.

3.3.3 Subtle Discrimination versus Overt Discrimination

To understand how the model's predictions relate to the type of discrimination, here we consider how they would change in the presence of overt discrimination, i.e., $\delta > 0$. The reaction functions become

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

$$e_r = \sigma \left[\frac{1}{2} + \beta - \delta - (2\beta - \delta) e_b \right]. \tag{10}$$

As discussed earlier, we assume $2\beta \geq \delta$, i.e., subtle discrimination is at least as strong as overt discrimination. Let $\epsilon \equiv \beta - \frac{\delta}{2} \geq 0$ denote the *excess subtle bias*. If $\epsilon = 0$ (no excess subtle bias), the reaction functions in (9) and (10) are 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 2. *If $\delta \geq 0$, $e_r^* \geq e_b^*$ if and only if $\sigma \leq \frac{1}{1-\delta}$.*

Corollary 2 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 in situations where overt discrimination is mild or nonexistent, but subtle discrimination remains.

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

3.3.4 Stakes and Investment in Skills

The next corollary presents further comparative statics results.

Corollary 3. *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.*

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

3.3.5 Comparison with the First Best

It is instructive to compare the equilibrium effort levels 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 2, 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 (from (7)) $e_r^* = e_b^* = 0.5 \geq \tilde{e}$. 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 if there is no subtle discrimination ($\beta = 0$). 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 in Section 4.

3.3.6 The Promotion Gap

We now define the promotion gap between blue and red agents:

Definition 1. *Let p_i denote agent i's promotion probability, $i \in \{b, r\}$. The promotion gap is*

$$\Delta p \equiv p_b - p_r = e_b - e_r + \left[e_b e_r + (1-e_b)(1-e_r)\right] 2\beta. \tag{11}$$

That is, the promotion gap is the difference between the promotion probabilities of blue and red agents.

Note that the promotion gap in Eq. (11) has two terms. The first term, $e_b - e_r$, is the difference in the probabilities of skill acquisition. Given our broad interpretation of what skills are, we call this difference the *achievement gap*. All else constant, a larger achievement gap increases the promotion gap. The second term is the difference in promotion probabilities between Blue and Red that arises as a direct consequence of the subtle bias. That is, this term is the promotion gap conditional on a tie times the probability of a tie. We call this term the *favoritism gap*. Note that the subtle bias affects the equilibrium investment levels, thus β affects both the achievement gap and the favoritism gap.

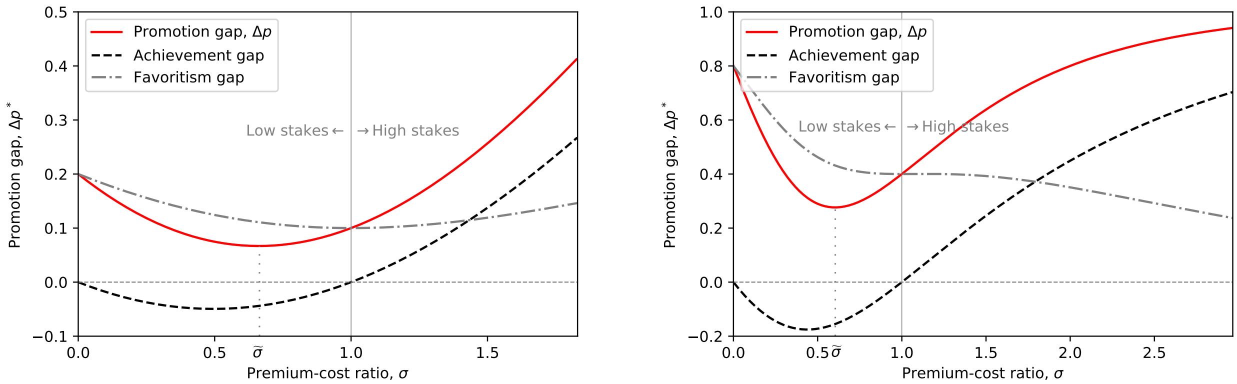
The next proposition shows how the equilibrium promotion gap varies with the premium-cost ratio.

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. Note that, for large values of the premium-cost ratio, even a small subtle bias can be significantly amplified through the strategic interactions between the agents.

Note also that, 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 often as the promotion premium increases, the principal is less likely to make biased promotion decisions as the stakes increase. In such scenarios, we would expect to find little direct evidence of discrimination.

The fact that the promotion gap eventually increases with the promotion premium can explain the "leaky pipe" phenomenon, i.e., increasing promotion gaps at higher hierarchical levels. In hierarchies with convex wage profiles, the net benefit from promotion increases with rank. In such hierarchies, we would expect the promotion gap to increase with rank.



(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 $\beta_2 = 0.4$.

3.4 Optimal Compensation Contracts

We now allow the principal to design the compensation contract. The principal is not allowed to explicitly discriminate through contracts, thus he must offer the same contract $\boldsymbol{w} = (w_1, W)$ 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, but finds it impossible to commit to flipping an unbiased mental coin, i.e., to behave as if $\beta = 0$.[21] 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 there is no commitment technology 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 – in-

[21]One limitation of System 1 is that it cannot be turned off. Kahneman (2011) illustrates this point with the famous Müller-Lyer optical illusion, where two horizontal lines of the same length appear to have different lengths because they end with fins pointing in different directions. One cannot decide to see the lines as equal even if one knows that they are.

terpretation 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 supervisor's bias (see Prendergast and Topel (1996) for a model along these lines).

Agents' outside utilities are normalized to zero. 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. For simplicity, we assume that the principal is risk-neutral and derives no utility from discrimination (that is, the subtle bias is either unconscious or caused by an infinitesimally small preference parameter). His profit-maximization problem (after entry, i.e., gross of entry costs) is as follows:

$$\max_{w_1 \geq 0, w_1 + W \geq 0} l + H\left(e_b + e_r - e_b e_r\right) - 2w_1 - W, \tag{12}$$

subject to

$$e_b = \arg\max_{e \in [0,1]} eW \left[\left(\frac{1}{2} - \beta\right) + 2\beta e_r\right] - \frac{ke^2}{2}, \tag{13}$$

$$e_r = \arg\max_{e \in [0,1]} eW \left[\left(\frac{1}{2} + \beta\right) - 2\beta e_b\right] - \frac{ke^2}{2}, \tag{14}$$

$$U_b(\boldsymbol{e}, \boldsymbol{w}) \geq 0, \tag{15}$$

$$U_r(\boldsymbol{e}, \boldsymbol{w}) \geq 0. \tag{16}$$

The principal faces two incentive compatibility (IC) constraints (Eq. 13 and Eq. 14) and two individual rationality (i.e., participation) constraints (Eq. 15 and Eq. 16). Because $w_1 \geq 0$ and $w_1 + W \geq 0$, agent i can guarantee a non-negative payoff by choosing $e_i = 0$. Thus, the participation constraints (Eq. 15 and Eq. 16) do not bind. 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," i.e., σ. Parameter σ denotes different contracts with different stakes involved; a high-stakes career path is a contract in which the prize from winning the promotion is high relative to the cost of investment. For both convenience and interpretation, from now on, we think of the principal's problem as that of choosing 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 simplified as follows:

$$\Pi(k, \beta, \theta) = \max_{\sigma \in [0, \overline{\sigma}(\beta)]} k\theta \left(e_b + e_r - e_b e_r\right) - k\sigma, \tag{17}$$

subject to

$$e_b = \frac{\sigma(0.5 - \beta) + 2\beta\sigma^2(0.5 + \beta)}{1 + 4\beta^2\sigma^2}, \tag{18}$$

$$e_r = \frac{\sigma(0.5 + \beta) - 2\beta\sigma^2(0.5 - \beta)}{1 + 4\beta^2\sigma^2}, \tag{19}$$

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 $\varepsilon \to 0$).

We first solve a baseline case of the above problem with no subtle discrimination ($\beta = 0$). In this case, we can explicitly solve for the optimal contract.

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. Note that 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 the marginal cost from investing when $e_b = e_r = 0$ is zero (i.e., $c'(0) = 0$), while the marginal social benefit from investing when $e_b = e_r = 0$ is positive and equal to $H > 0$. Thus, when $\theta \leq 1$, the firm inefficiently stays out of the sector. Such inefficiency occurs because the non-negative wage constraint prevents the firm from extracting all the surplus from the agents.[22]

Consider now the general case in which $\beta \geq 0$. Although it is not possible to solve analytically for the optimal contract in all cases, the existence and uniqueness of the optimal contract are easily established:

Proposition 5. *For every set of parameters* $(k > 0, \beta \in [0, 0.5], \theta > 0)$, *there exists a unique*[23] *solution* $\sigma(k, \beta, \theta)$ *to the principal's problem (if the firm chooses not to operate, we set* $\sigma = 0$*).*

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

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* $\sigma(\beta, \theta) = \overline{\sigma}(\beta)$.

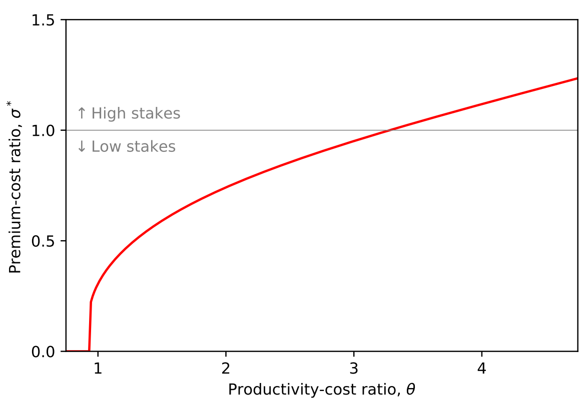
2. *The optimal stake,* $\sigma(\beta, \theta)$, *is strictly increasing in* $\theta \in \left[\underline{\theta}(\beta), \overline{\theta}(\beta)\right]$.

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

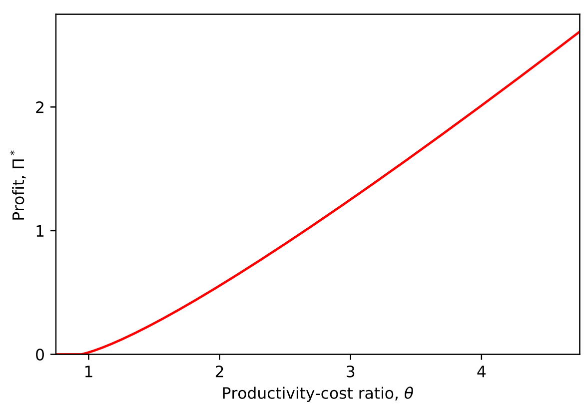
[22] The result that, with no bias, the firm operates only when $\theta > 1$ (i.e., $H > k$; the first-best requires only one agent to be employed) is special to the quadratic cost function. As we show in the Internet Appendix, under different cost functions, the firm may operate when the first-best outcome requires both agents to be employed. What remains true under any cost function is that limited liability generally makes low-productivity firms unprofitable and thus not viable.

[23] Uniqueness here is in the generic sense; multiple solutions may arise for measure-zero combinations of parameters (k, β, θ).

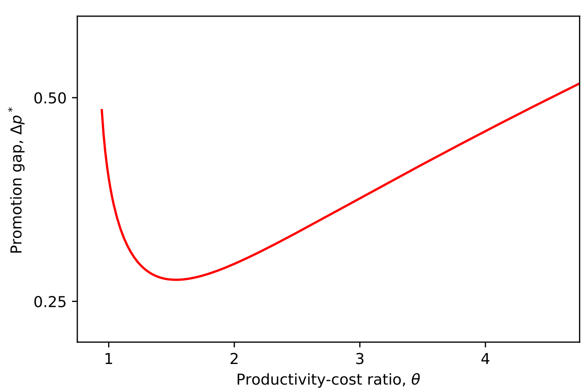
Part 2 of Proposition 6 implies that human-capital-intensive firms (high-θ firms) offer career paths involving higher stakes. Because the optimal stake is increasing in θ, all the comparative statics in the previous subsection are unchanged once we replace σ with θ. In particular, if we define $\widetilde{\theta}(\beta)$ 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 ($\beta = 0.4$).

3.5 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.

Formally, to avoid a "multi-selves" interpretation of the problem, here we assume that the principal delegates the promotion decision to a supervisor. Suppose the principal can select a supervisor with known bias β at no cost. Which supervisor would the principal choose? The principal's problem is

$$\Pi(\theta) = \max_{(\sigma,\beta)\in[0,\overline{\sigma}(\beta)]\times[0,0.5]} \theta\left(e_b + e_r - e_b e_r\right) - \sigma, \tag{20}$$

subject to

$$e_b = \frac{\sigma(0.5-\beta) + 2\beta\sigma^2(0.5+\beta)}{1+4\beta^2\sigma^2}, \tag{21}$$

$$e_r = \frac{\sigma(0.5+\beta) - 2\beta\sigma^2(0.5-\beta)}{1+4\beta^2\sigma^2}. \tag{22}$$

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.

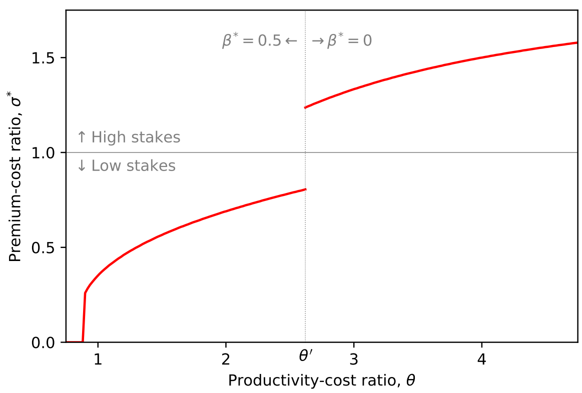
Proposition 7. *There exists $\theta' < \overline{\theta}$ such that*

$$\beta(\theta) = \begin{cases} 0.5 & \text{if } \theta \in (0, \theta'] \\ 0 & \text{if } \theta \in [\theta', \overline{\theta}] \end{cases}. \tag{23}$$

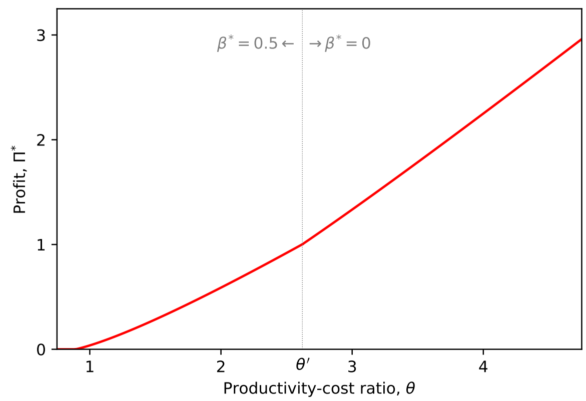
Furthermore, $\sigma(\theta) < 1$ if $\theta \in (0, \theta']$ and $\sigma(\theta) > 1$ if $\theta \in [\theta', \overline{\theta}]$.

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. 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 choose to offer low-stake careers (i.e., $\sigma(\theta) < 1$). 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.

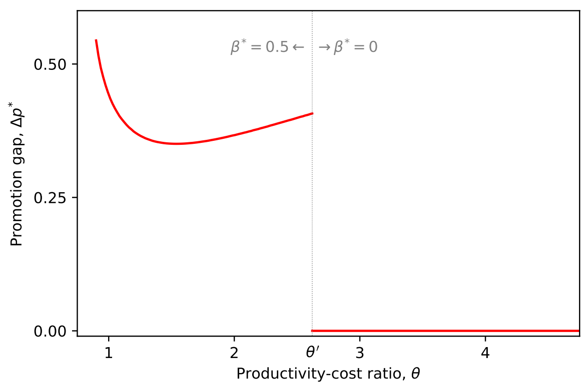
By contrast, for high-θ firms, the profit is maximized when the subtle bias is zero. That is, firms that offer high-stakes careers prefer not to discriminate. Intuitively, in firms with high-stake careers, the discouragement effect is strong, hindering the performance of discriminated agents. In such sectors, discriminating firms are less profitable than non-discriminating firms, and thus more likely to be driven out by competition.



(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.

3.5.1 Discussion: Firm Polarization

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. By 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 with respect to discrimination and diversity.[24]

High-productivity firms prefer to promote a work environment free of discrimination. These firms strive to be perceived as "progressive" and "activist." They have more diversity at the top (i.e., a smaller promotion gap). These firms also offer careers with higher

[24]The firm polarization result is robust to using a more general cost function (see the Internet Appendix).

stakes and are likely to be large, profitable, and human capital-intensive. By contrast, low-productivity firms do not take actions to counter subtle discrimination. These firms do not mind being perceived as "conservative" and are less diverse at the top. They offer careers with low stakes and are smaller, less profitable, and less human capital-intensive than "progressive" firms. Note also 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 $\eta \to 0$).

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)). We are unaware of theoretical work that explains these cross-sectional correlations. With respect to firm size, Adams (2016) writes that "more research needs to be done on the reasons why women are less represented on the boards of small firms, but the evidence that this is the case is clear." Subtle discrimination can explain these findings. High-productivity (i.e., large and profitable) firms may choose to take actions that incentivize the recruitment of women to their boards. These actions would reduce their promotion gaps and thus increase the proportion of women in top jobs.

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?[25] The answer follows from Proposition 7:

[25]See the Internet Appendix for an analysis of hard quotas.

Corollary 4. *The firm adopts a soft quota that incentivizes the promotion of red agents if and only if* $\theta \geq \theta'$.

4 Welfare Analysis and Empirical Testing

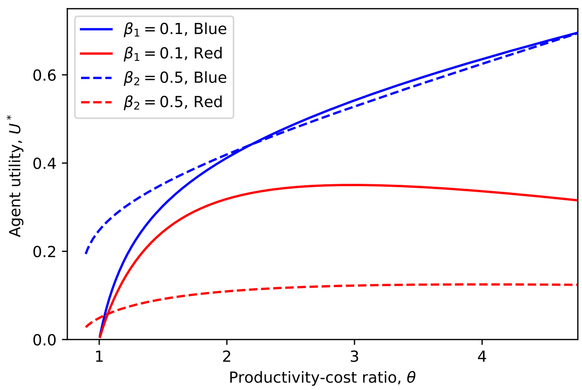
Unlike traditional models of taste-based discrimination, in our model, changing the bias does not mechanically affect utilities. This property allows our model to produce sharper welfare and policy implications. In this section, we first address a number of normative questions: What are the welfare implications of subtle discrimination? When is subtle discrimination inefficient? Then, we discuss the implications of subtle discrimination for traditional tests for discrimination.

4.1 When Does Discrimination Harm Workers?

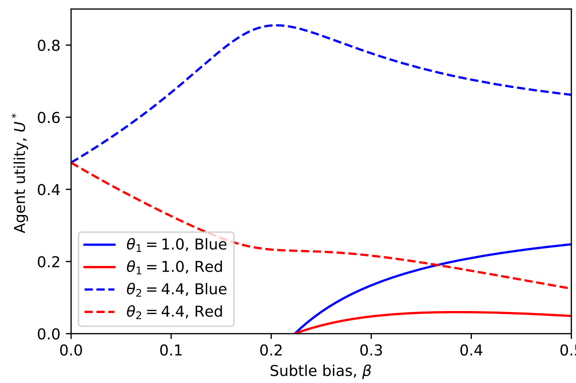
Figures 5a and 5b show the utilities of blue and red agents as functions of the productivity-cost ratio, θ, and the subtle bias, β, when the contract $\sigma(\theta, \beta)$ is optimally chosen by the principal. Two features are worth highlighting.

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) As a function of θ for small and large values of the subtle bias, $\beta_1 = 0.1$ and $\beta_2 = 0.5$.



(b) As a function of β for low and high levels of the productivity-cost ratio, $\theta_1 = 1.0$ and $\theta_2 = 4.4$.

Figure 5: Agents' utilities, U^*, under optimal contract $\sigma(\theta, \beta)$.

4.2 Social Surplus

Figure 6 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 5b, for sufficiently low values of θ both Blue and Red benefit from increasing the bias.[26] 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 5b), while the firm's profit is relatively flat in β. The optimal bias trades off the

[26]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.

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.

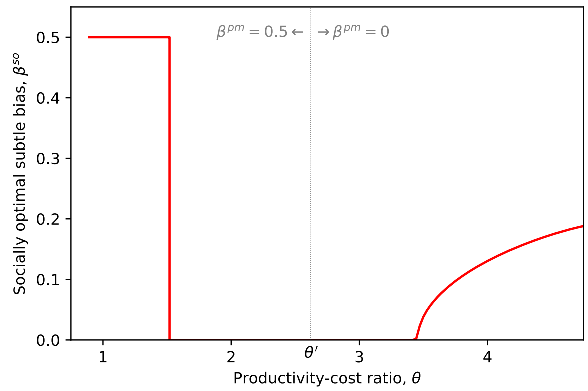


Figure 6: Socially optimal level of subtle bias as a function of the productivity-cost ratio, $\beta^{so}(\theta)$.

The overcompensation effect is the reason why subtle discrimination is not always economically inefficient. Because firms do not internalize the effect of their contractual choices on their workers' welfare, a larger bias can lead to a Pareto superior equilibrium. This is yet another example of the well-known result that two frictions may be better than one. The social optimality of moderate to high subtle biases is a consequence of the assumption of non-negative wages, which is a contractual friction. Subtle discrimination is also a friction; in the absence of other frictions, the socially optimal β is always zero.

4.3 Testing for Subtle Discrimination

Our notion of subtle discrimination is relevant in competitive settings, i.e., in situations where agents compete for a fixed prize. The typical "outcome test" for discrimination is based on comparing the ex-post performances of the marginally-treated agents. The idea is that if one group is held to higher standards than the other group, the marginally-treated agent from the unfavored group performs better than the marginally-treated agent 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.

In our model, marginally-promoted blue and red agents are equally productive. Thus, a well-designed outcome test cannot reject its null hypothesis. A key implication is that subtle discrimination should feature alongside statistical discrimination and no discrimination as the null hypothesis in outcome tests in competitive situations.

Although standard outcome tests cannot detect subtle discrimination, there are several ways in which one can test for subtle discrimination in competitive situations. One is to identify a direct shock to the bias, i.e., a shock to β. According to our model, an ex-post, unanticipated small shock to β would change the observed promotion gaps between the groups but would have no impact on 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. That is, the shock to gender preferences changes the observed hiring gaps. However, they also find that the shock has no effect on 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).[27]

Another approach to testing for subtle discrimination is to consider the impact of discrimination on those who are subjected to discrimination (for an example of this approach, see Hengel (2022)). In our context, this requires testing the predictions of our model for the investment choices made by the agents, in particular, how they relate to the stakes faced by the agents. An instructive example – although in a somewhat different context – is the work of Filippin and Guala (2013), who ran a lab experiment where individuals assigned to different groups submit bids in an all-pay auction. The winner is

[27] Even explicit preference-based discrimination may sometimes be payoff-irrelevant. Huber et al. (2021) find that the expulsion of Jewish managers in Nazi Germany affected firms' stock prices, but only for those firms with managers with characteristics that were hard to replace.

selected by an auctioneer who has strong incentives to reward the highest bidder. Nevertheless, the auctioneers more frequently assign the prize to a member of their own group when bids of two or more players are tied. In response, out-group bidders reduce their bids, leading to a decrease in their earnings and a substantial gap in outcomes between groups.

Finally, direct tests of subtle discrimination can also be designed in the lab. Foschi et al. (1994) designed an experiment where subjects must promote at most one of two candidates. They can also choose to promote no one. When subjects choose between a pair of candidates of the same sex, sometimes no one is promoted. Thus, the authors can infer the minimum threshold of qualifications for promotions for each sex. They find that subjects use similar thresholds for pairs of male and female candidates. That is, they show that men and women are held to the same standards when competing against someone of the same sex. By contrast, when men and women with identical qualifications compete for the same position, they find that subjects are more likely to promote men. Thus, the evidence is consistent with our definition of subtle discrimination but inconsistent with statistical discrimination.[28]

5 Conclusion

Most cases of discrimination we witness in day-to-day life are subtle. Subtle discrimination leaves no trace and is subject to plausible deniability. Although subtle discrimination may harm those at the receiving end of discriminatory actions, it may not have many immediate consequences for the perpetrating parties.

Our leading example of subtle discrimination is the use of biased tie-breaking rules in promotion contests. When candidates are indistinguishable in terms of their future

[28] In related work, Reuben et al. (2014) and Moss-Racusin et al. (2012) show experimental evidence that subjects' preexisting subtle biases explain their propensity to hire male candidates when choosing between candidates with similar qualifications.

productivity, the firm is indifferent between biased and unbiased tie-breaking rules. This indifference opens the door for decision rules that favor characteristics unrelated to future productivity. Thus, subtle discrimination may result from a small bias, which manifests itself only when decisions are ex-post inconsequential. However, despite the small size of subtle biases, our model shows that they may have significant implications. First, subtle biases in promotion decisions distort candidates' incentives to take actions that increase their promotability. Second, the competitive nature of promotion contests can amplify the ex-ante effects of subtle biases.

Our model generates several novel predictions. In particular, it can explain 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. By contrast, low-productivity firms benefit from holding discriminated workers to higher standards, as these employees overcompensate by working harder.

References

ADAMS, R. B. (2016): "Women on boards: The superheroes of tomorrow?" *The Leadership Quarterly*, 27, 371–386.

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.

——— (1998): "What has economics to say about racial discrimination?" *Journal of Economic Perspectives*, 12, 91–100.

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

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.

BEGENY, C., M. RYAN, C. MOSS-RACUSIN, AND G. RAVETZ (2020): "In some professions, women have become well represented, yet gender bias persists—Perpetuated by those who think it is not happening," *Science Advances*, 6, eaba7814.

BENSON, A., D. LI, AND K. SHUE (2021): ""Potential" and the gender promotion gap," Working paper, University of Minnesota.

BOHREN, J. A., K. HAGGAG, A. IMAS, AND D. G. POPE (2019a): "Inaccurate statistical discrimination: An identification problem," Working paper, National Bureau of Economic Research.

BOHREN, J. A., A. IMAS, AND M. ROSENBERG (2019b): "The dynamics of discrimination: Theory and evidence," *American Economic Review*, 109, 3395–3436.

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.

BUSER, T., M. J. VAN DEN ASSEM, AND D. VAN DOLDER (2023): "Gender and willingness to compete for high stakes," *Journal of Economic Behavior & Organization*, 206, 350–370.

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.

COTTERMAN, J. D. (2004): *Compensation plans for law firms*, American Bar Association.

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

CZIRAKI, P. AND A. ROBERTSON (2021): "Credentials matter, but only for men: Evidence from the S&P 500," Working paper, University of Toronto.

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.

DEITCH, E. A., A. BARSKY, R. M. BUTZ, S. CHAN, A. P. BRIEF, AND J. C. BRADLEY (2003): "Subtle yet significant: The existence and impact of everyday racial discrimination in the workplace," *Human Relations*, 56, 1299–1324.

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.

——— (2000): "Aversive racism and selection decisions: 1989 and 1999," *Psychological science*, 11, 315–319.

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.

EGAN, M., G. MATVOS, AND A. SERU (2022): "When Harry fired Sally: The double standard in punishing misconduct," *Journal of Political Economy*, 130, 1184–1248.

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.

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.

FILIPPIN, A. AND F. GUALA (2013): "Costless discrimination and unequal achievements in an experimental tournament," *Experimental Economics*, 16, 285–305.

FOSCHI, M., L. LAI, AND K. SIGERSON (1994): "Gender and double standards in the assessment of job applicants," *Social Psychology Quarterly*, 326–339.

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.

FREDERIKSEN, A., F. LANGE, AND B. KRIECHEL (2017): "Subjective performance evaluations and employee careers," *Journal of Economic Behavior & Organization*, 134, 408–429.

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

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.

GAWRONSKI, B. (2019): "Six lessons for a cogent science of implicit bias and its criticism," *Perspectives on Psychological Science*, 14, 574–595.

GERALDES, D. (2020): "Women dislike competing against men," Working paper, University College Dublin.

GLOVER, D., A. PALLAIS, AND W. PARIENTE (2017): "Discrimination as a self-fulfilling prophecy: Evidence from French grocery stores," *The Quarterly Journal of Economics*, 132, 1219–1260.

GREENWALD, A. G., M. R. BANAJI, L. A. RUDMAN, S. D. FARNHAM, B. A. NOSEK, AND D. S. MELLOTT (2002): "A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept." *Psychological Review*, 109, 3.

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.

HEBL, M., J. M. MADERA, AND E. KING (2008): *Exclusion, Avoidance, and Social Distancing*, Psychology Press, 157–180, Diversity Resistance in Organizations.

HEILMAN, M. E. AND M. C. HAYNES (2005): "No credit where credit is due: attributional rationalization of women's success in male-female teams." *Journal of Applied Psychology*, 90, 905.

HEINZ, J. P., R. L. NELSON, R. L. SANDEFUR, AND E. O. LAUMANN (2005): *Urban lawyers: The new social structure of the bar*, University of Chicago Press.

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.

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., E. J. MAYER, AND D. P. MILLER (2022): "Gender Bias in Promotions: Evidence from Financial Institutions," Working paper, Southern Methodist University.

HUBER, K., V. LINDENTHAL, AND F. WALDINGER (2021): "Discrimination, managers, and firm performance: Evidence from "aryanizations" in nazi germany," *Journal of Political Economy*, 129, 2455–2503.

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.

KLINE, P., E. K. ROSE, AND C. R. WALTERS (2022): "Systemic discrimination among large US employers," *The Quarterly Journal of Economics*, 137, 1963–2036.

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. P. (2009): "Firm-specific human capital: A skill-weights approach," *Journal of Political Economy*, 117, 914–940.

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

LUKSYTE, A., K. L. UNSWORTH, AND D. R. AVERY (2018): "Innovative work behavior and sex-based stereotypes: Examining sex differences in perceptions and evaluations of innovative work behavior," *Journal of Organizational Behavior*, 39, 292–305.

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.

MOSS-RACUSIN, C. A., J. F. DOVIDIO, V. L. BRESCOLL, M. J. GRAHAM, AND J. HANDELSMAN (2012): "Science faculty's subtle gender biases favor male students," *Proceedings of the National Academy of Sciences*, 109, 16474–16479.

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

NIEDERLE, M. AND L. VESTERLUND (2007): "Do women shy away from competition? Do men compete too much?" *The Quarterly Journal of Economics*, 122, 1067–1101.

——— (2011): "Gender and competition," *Annual Review of Economics*, 3, 601–630.

NOH, S., V. KASPAR, AND K. A. WICKRAMA (2007): "Overt and subtle racial discrimination and mental health: Preliminary findings for Korean immigrants," *American Journal of Public Health*, 97, 1269–1274.

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

PAGANO, M. AND L. PICARIELLO (2022): "Corporate Governance, Favoritism and Careers," Working paper, University of Naples Federico II.

PAO, E. K. (2017): *Reset: My fight for inclusion and lasting change*, Random House.

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.

RÉGNER, I., C. THINUS-BLANC, A. NETTER, T. SCHMADER, AND P. HUGUET (2019): "Committees with implicit biases promote fewer women when they do not believe gender bias exists," *Nature Human Behaviour*, 3, 1171–1179.

REUBEN, E., P. SAPIENZA, AND L. ZINGALES (2014): "How stereotypes impair women's careers in science," *Proceedings of the National Academy of Sciences*, 111, 4403–4408.

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.

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 for an interior solution are

$$H(1-e_j) - c'(e_i) = 0, \tag{24}$$

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

$$\tilde{e} = 1 - \frac{c'(\tilde{e})}{H}. \tag{25}$$

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 $e_r^{FB} = 0$.

□

Proof of Proposition 2. Equations (7) and (8) represent the unique solution to the system of equations given by (5) and (6). From (7), 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 $b = 2\beta(1-2\beta) > 0$. It suffices to show that $-a + \sqrt{a^2+2b} >$

b. Indeed, $-a+\sqrt{a^2+2b} > b \iff a^2+2b > a^2+2ab+b^2 \iff 2 > b+2a \iff 2 > 2\beta - 4\beta^2 + 1 - 2\beta \iff 1 > -4\beta^2$.

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:

$$\sigma'(\beta) = \frac{0.5+\beta-\sqrt{\frac{1}{4}-3\beta-7\beta^2}}{2\beta\,(2\beta+1)} > 0.$$

Note that $f_r(1) > 0$, and its derivative at $\sigma = 1$ is

$$\frac{\partial f}{\partial \sigma}(\sigma = 1) = 2\beta\,(2\beta+1) - (0.5+\beta)\,,$$

which is strictly negative for $\beta \in \left(0, \frac{1}{14}\right]$. Thus, it must be that $\sigma'(\beta) > 1$. Note also that $f_r(\sigma) - f_b(\sigma) = 2\beta\sigma(\sigma-1)$, which is positive if and only if $\sigma > 1$. Thus, at $\sigma'(\beta)$ we have $f_r(\sigma'(\beta)) > f_b(\sigma'(\beta))$, which implies $\overline{\sigma}(\beta) < \sigma'(\beta)$.

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. □

Proof of Corollary 1. Note that, from (7) and (8), $e_r^* = e_b^* \frac{(0.5+\beta)-2\beta\sigma(0.5-\beta)}{(0.5-\beta)+2\beta\sigma(0.5+\beta)}$. Straightforward manipulation of this equality implies that $e_r^* \geq e_b^*$ if and only if $\sigma \leq 1$. □

Proof of Corollary 2. Solving (9) and (10) yields

$$e_b^* = \frac{\sigma(0.5-\beta) + (2\beta-\delta)\sigma^2(0.5+\beta-\delta)}{1+(2\beta-\delta)^2\sigma^2}; \tag{26}$$

$$e_r^* = \frac{\sigma(0.5+\beta-\delta) - (2\beta-\delta)\sigma^2(0.5-\beta)}{1+(2\beta-\delta)^2\sigma^2}, \tag{27}$$

which implies $e_r^* = e_b^* \frac{(0.5+\beta-\delta)-(2\beta-\delta)\sigma(0.5-\beta)}{(0.5-\beta)+(2\beta-\delta)\sigma(0.5+\beta-\delta)}$. Straightforward manipulation of this equality implies that $e_r^* \geq e_b^*$ if and only if $\sigma(1-\delta) \leq 1$. □

Proof of Corollary 3. 1. Differentiating (7) with respect to σ yields

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

which is strictly positive 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 (8) 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.

□

Proof of Proposition 3. The equilibrium promotion gap is

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

Its derivative with respect to σ is

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

Define the function $A(\sigma)$ as the the numerator of the expression above:

$$A(\sigma) = -2 + 3\left(1 - 4\beta^2\right)\sigma + 24\beta^2\sigma^2 + 4\beta^2\left(4\beta^2 - 1\right)\sigma^3.$$

$A(\sigma)$ 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:

$$A'(\sigma) = 3(1 - 4\beta^2) + 48\beta^2\sigma + 12\beta^2\left(4\beta^2 - 1\right)\sigma^2.$$

To find the roots for $A'(\sigma) = 0$, apply the quadratic root formula to obtain

$$\sigma^m = \frac{4\beta - \sqrt{16\beta^2 + (1 - 4\beta^2)^2}}{2\beta\,(1 - 4\beta^2)}, \sigma^M = \frac{4\beta + \sqrt{16\beta^2 + (1 - 4\beta^2)^2}}{2\beta\,(1 - 4\beta^2)}.$$

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 condition for an interior solution is trivially satisfied:

$$f_b(1) = \beta\,(2\beta - 1) - (0.5 - \beta) + 1 = 2\beta^2 + 0.5 > 0.$$

At $\sigma = 1$, 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 $\min\{r_2, \overline{\sigma}\} \equiv \widetilde{\sigma}$. □

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 thus

$$\max_{\sigma \in [0, \overline{\sigma}(0)]} \theta\left[1 - \left(1 - \frac{\sigma}{2}\right)^2\right] - \sigma.$$

The first order condition for an interior solution is

$$\theta\left(1 - \frac{\sigma}{2}\right) - 1 = 0.$$

The second-order condition 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 $\theta \geq 1$, 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 exist. 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 $[0, \overline{\sigma}(\beta)]$. $\square$

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 (7) and (8), 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):

$$\frac{\partial \Pi}{\partial \theta} = f(\sigma(\beta, \theta), \beta) > 0. \tag{28}$$

To prove Part 1, notice first that at $\theta = 0$, trivially, $\sigma(\beta, 0) = 0$ and the profit is zero. For $\theta = \overline{\sigma}(\beta) + \varepsilon$, 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

$$\overline{\theta}(\beta) \equiv \frac{1}{f_\sigma(\overline{\sigma}(\beta), \beta)}, \tag{29}$$

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 first-order condition 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 first order condition):

$$\frac{\partial \sigma}{\partial \theta} = -\frac{f_\sigma(\sigma^*, \beta)}{\theta f_{\sigma\sigma}(\sigma^*, \beta)} = -\frac{1}{\theta^2 f_{\sigma\sigma}(\sigma^*, \beta)} > 0,$$

proving Part 2. Part 3 follows from (28).

□

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:

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

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

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

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

$$\Delta(\theta) = \Pi(\sigma(0,\theta),0,\theta) - \Pi(\sigma(0.5,\theta),0.5,\theta),$$

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 $\theta' \in (1,4)$[29] 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 (after simplification)

$$f(\sigma,\beta_{root}(\sigma)) = \frac{\sigma^2 + 2\sigma + 25}{32},$$

[29] Numerically, we obtain that $\theta' \approx 2.62054$.

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 $\sigma = 16 - \theta$. Thus, 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, $\sigma(\overline{\theta}) < \frac{5}{3}$.

$\square$

Subtle Discrimination
Internet Appendix

Elena S. Pikulina Daniel Ferreira

March 2023

1 Hard Quotas

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 (as shown in the second region in Figure 6). 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. In line with the basic model, 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*). Here we adopt the interpretation 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 we assume that the firm cannot choose the bias of its supervisors. Because only supervisors observe the skill s_i of their pairs of subordinates, any rule that allows supervisors some discretion can be abused. Thus, the only way to comply with the quota is for the principal 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.[1] The principal's problem is

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

subject to

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

$$e_r = \frac{\sigma(0.5 + \beta) - 2\beta\sigma^2(0.5 - \beta)}{1 + 4\beta^2\sigma^2}, \tag{3}$$

$$\Delta p \equiv \delta\{e_b - e_r + \left[e_b e_r + (1 - e_b)(1 - e_r)\right] 2\beta\} - (1 - \delta) = q, \tag{4}$$

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

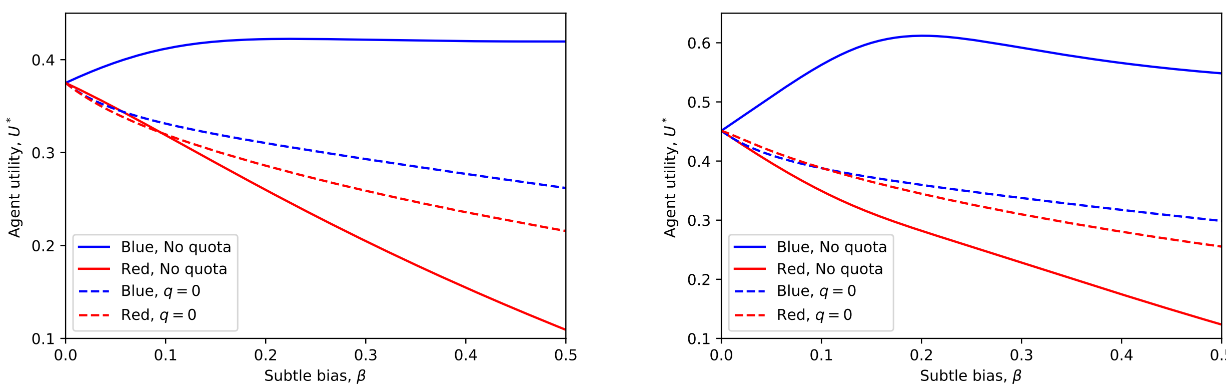
Firm profit is always higher when there is no quota or if the quota is not binding (i.e., $q = \Delta p_0$). This reduction in profit is expected; the quota constrains the principal's maximization problem. Still, there might be reasons to support quotas on grounds of

[1]The no disclosure case yields similar results.

redistributive equity. The key question is then: when do discriminated agents benefit from quotas?

Figure IA.1 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 counterintuitive result occurs because, under a quota, the firm chooses to offer a smaller promotion bonus. This negative effect dominates when the bias is low because, in such a 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 must be high.



(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 IA.1: 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 $\theta_2 = 3.2$.

2 General Cost Function

In this section 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:

$$c(e) = \frac{k}{\alpha} \frac{e^{\alpha}}{1 - e^{\gamma}}, \tag{5}$$

Figure IA.2: $\bar{\gamma}$ as a function of the cost function parameter α

The above 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:

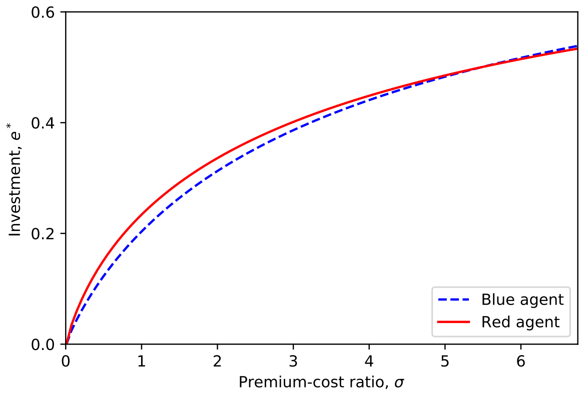
$$s_{2a}(\theta; \alpha, \gamma) = \max_{e} \theta e(2 - e) - 2 \times \frac{k}{\alpha} \frac{e^{\alpha}}{1 - e^{\gamma}}.$$

Then, for $\alpha \in [2.0, 10.0]$, we compute $\bar{\gamma}$ such that for all $\gamma < \bar{\gamma}$ (see Figure IA.2), the social welfare is maximized when both agents invest in their human capital for all values of the firm productivity parameter 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 $\gamma < \bar{\gamma}$. This way we confirm that our results in the main text are not driven by the fact that under a quadratic cost function, for $\theta > 1$ it is more socially efficient to treat agents asymmetrically, i.e., ask only one of them to invest. In the remainder of this appendix we assume $\alpha = 2.0$ ($\bar{\gamma} \approx 18.5013$ for $\alpha = 2.0$).

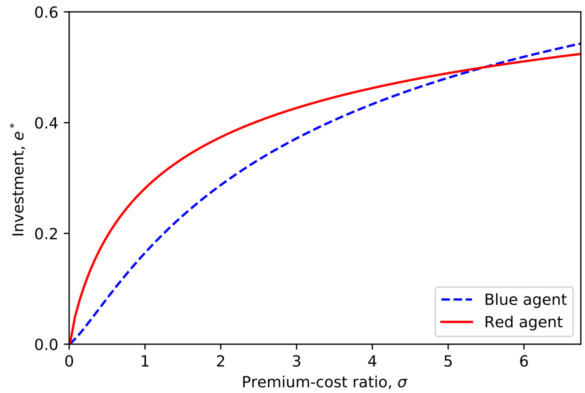
Figure IA.3 shows the equilibrium investment levels as a function of the premium-cost ratio, σ, for two levels of the subtle bias, $\beta \in \{0.1, 0.4\}$ and the following levels of the cost function parameters, $\gamma \in \{0.5, 9.0, 18.0\}$. Note that for $\alpha = 2.0$, all these values of γ are below $\bar{\gamma}$. Figure IA.3 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 IA.4 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 the premium-cost ratio σ, the contribution of the achievement gap to the promotion gap is lower than under low values of σ.

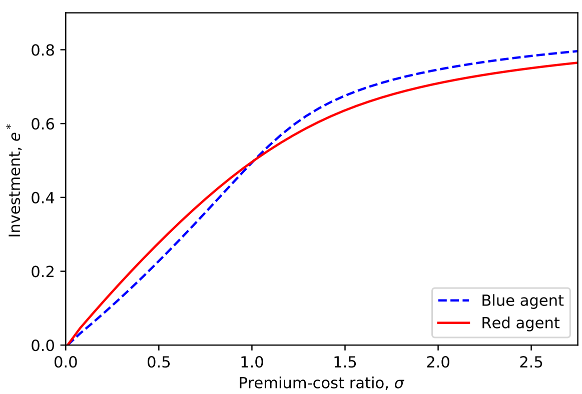
Figure IA.5 shows the solution for the principal's problem, the optimal premium-cost ratio, $\sigma^*(\theta; \beta, \gamma, \alpha)$, the optimal profit, $\Pi^*(\theta; \beta, \gamma, \alpha)$, and the resulting promotion gap, $\Delta p^*(\theta; \beta, \gamma, \alpha)$ as functions of the productivity-cost ratio θ and for given values of the subtle bias β and parameters γ and α. In particular, $\beta \in \{0.1, 0.4\}$, $\gamma \in \{0.5, 9.0, 18.0\}$ and $\alpha = 2.0$.



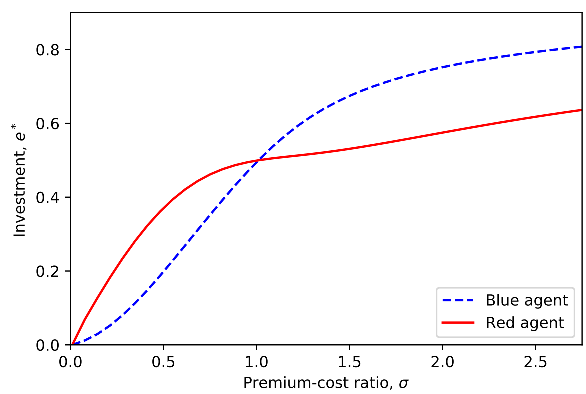
(a) $\beta = 0.1$, $\gamma = 0.5$



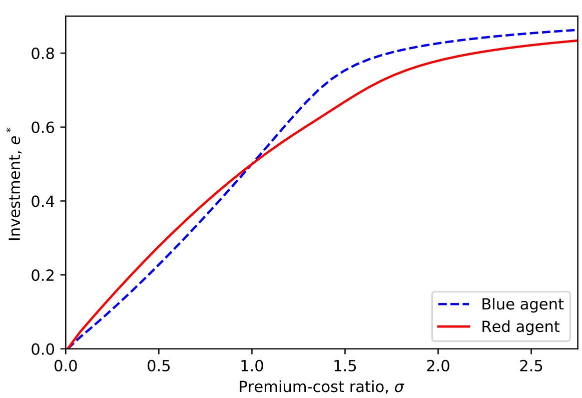
(b) $\beta = 0.4$, $\gamma = 0.5$



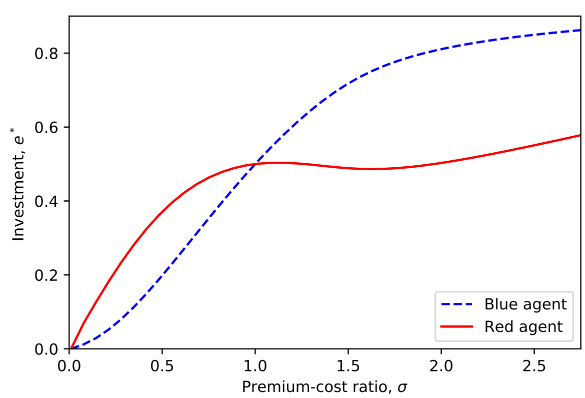
(c) $\beta = 0.1$, $\gamma = 9.0$



(d) $\beta = 0.4$, $\gamma = 9.0$

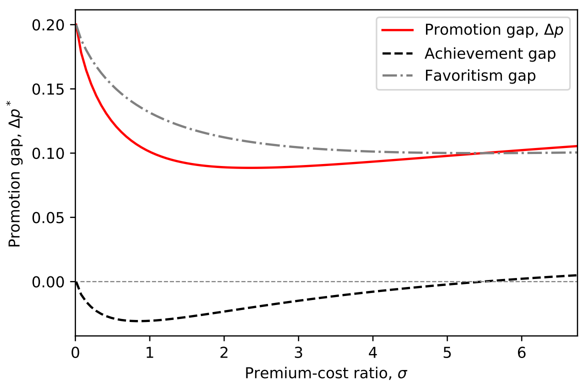


(e) $\beta = 0.1$, $\gamma = 18.0$

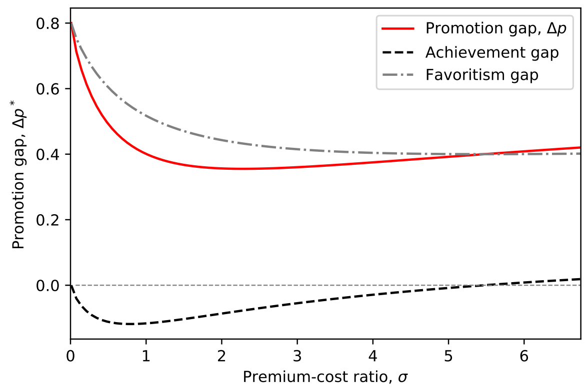


(f) $\beta = 0.4$, $\gamma = 18.0$

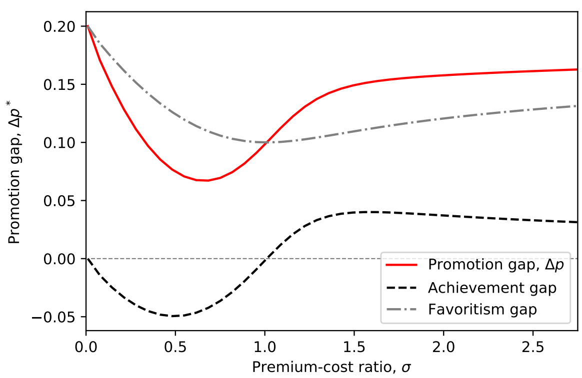
Figure IA.3: Equilibrium investments of blue and red agents, e_b^* and e_r^*, as functions of the premium-cost ratio, σ, for different values of the subtle bias β and the cost function parameter γ.



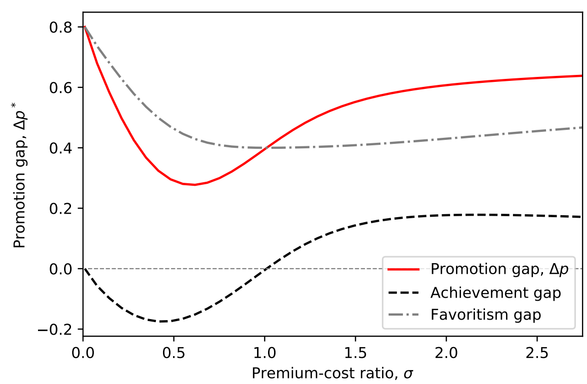
(a) $\beta = 0.1$, $\gamma = 0.5$



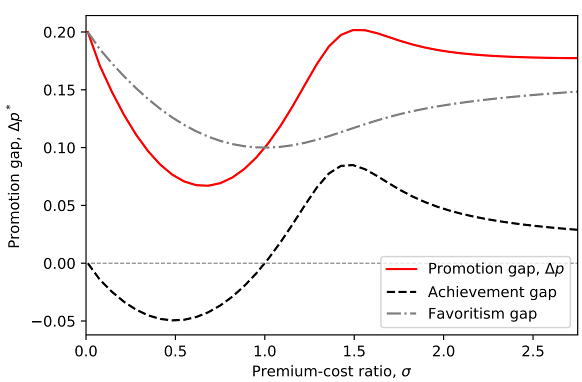
(b) $\beta = 0.4$, $\gamma = 0.5$



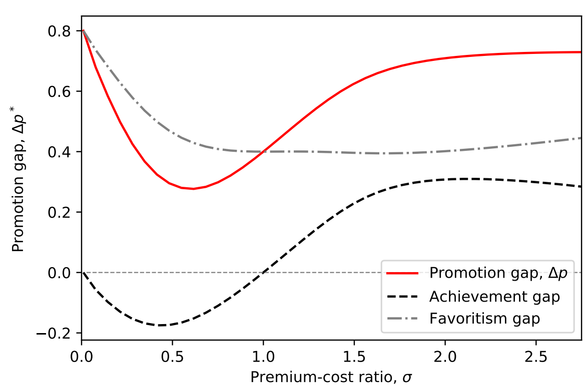
(c) $\beta = 0.1$, $\gamma = 9.0$



(d) $\beta = 0.4$, $\gamma = 9.0$

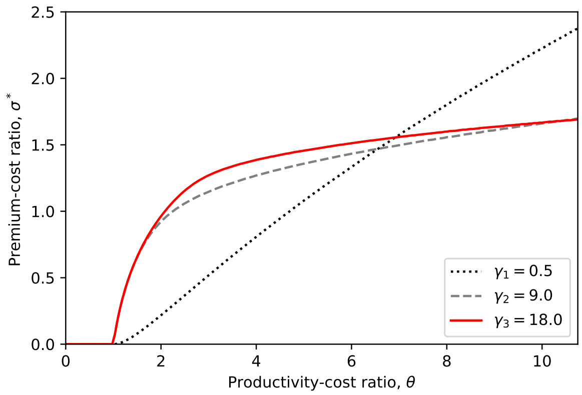


(e) $\beta = 0.1$, $\gamma = 18.0$

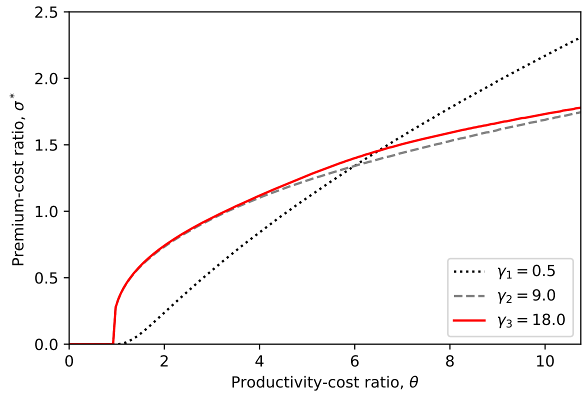


(f) $\beta = 0.4$, $\gamma = 18.0$

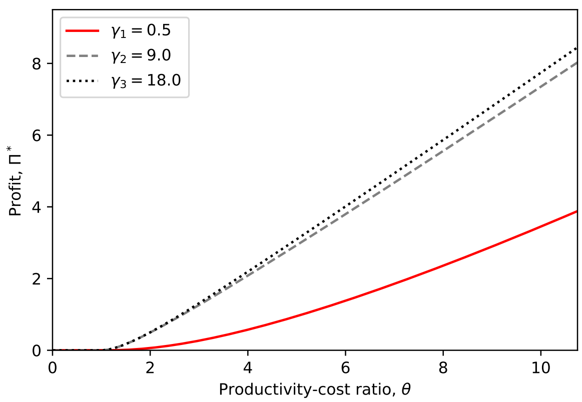
Figure IA.4: Equilibrium promotion gap, Δp^*, as a function of the premium-cost ratio, σ, for different values of the subtle bias β and the cost function parameter γ.



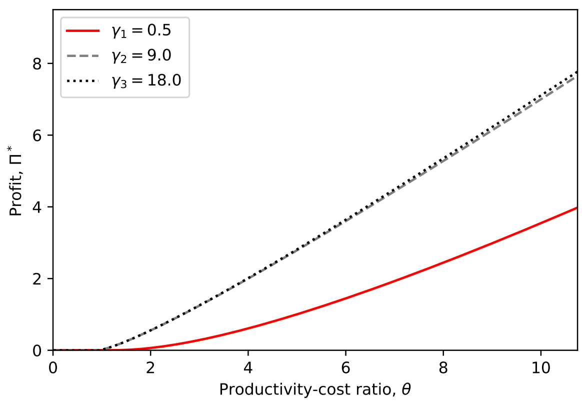
(a) Premium-cost ratio, σ^*, for $\beta = 0.1$



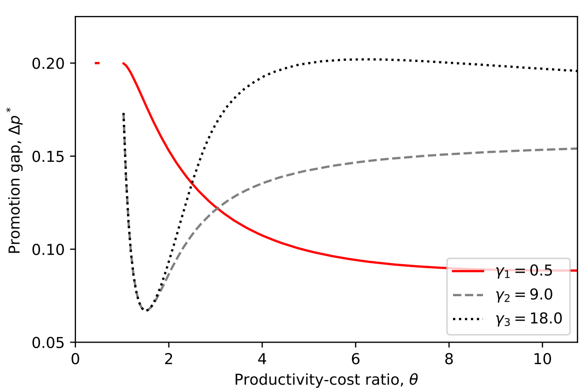
(b) Premium-cost ratio, σ^*, for $\beta = 0.4$



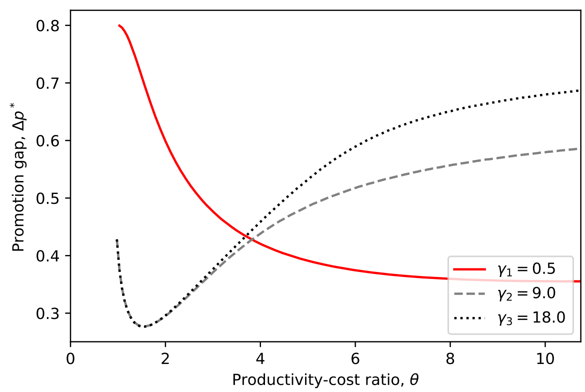
(c) Firm profit, Π^*, for $\beta = 0.1$



(d) Firm profit, Π^*, for $\beta = 0.4$



(e) Promotion gap, Δp^*, for $\beta = 0.1$



(f) Promotion gap, Δp^*, for $\beta = 0.4$

Figure IA.5: 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.

Now we consider a situation when a firm could optimally choose its own subtle bias. In particular, we confirm that there exists θ' such that for all $\theta < \theta'$, the firm prefers the

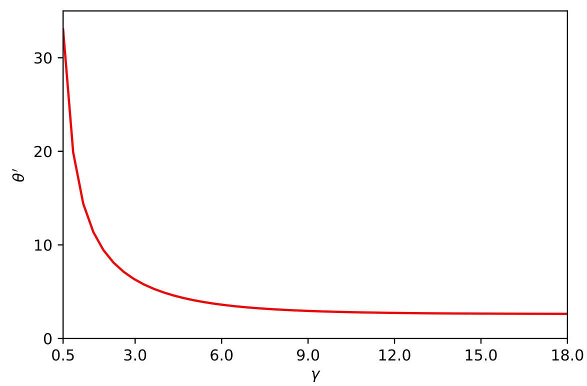
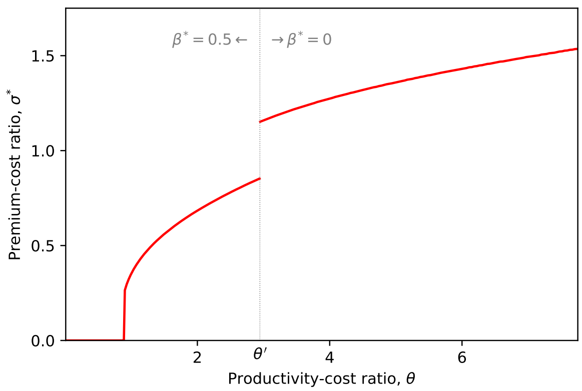


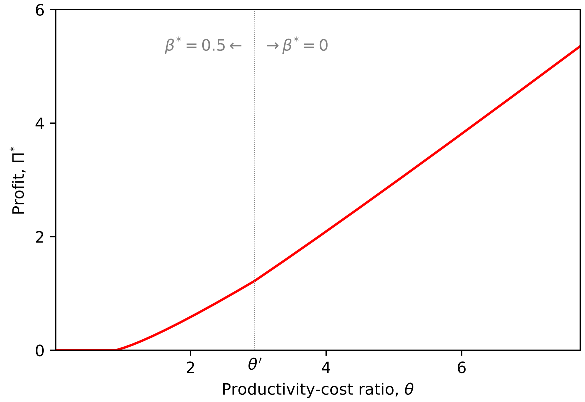
Figure IA.6: θ' as a function of the cost function parameter γ

maximum level of subtle bias, $\beta^* = 0.5$ and for all $\theta > \theta'$, the firm prefers the minimum level of subtle bias, $\beta^* = 0$. Figure IA.6 shows θ' as a function of γ, where $\gamma \in [0.5, 18.0]$ and $\alpha = 2$. Therefore, we confirm that firm polarization occurs even under a more general cost function.

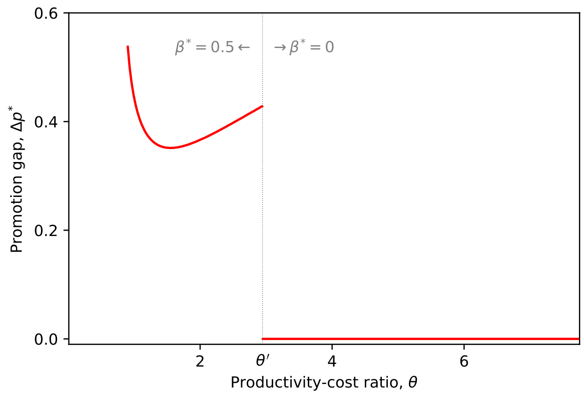
Finally, in Figure IA.7 we replicate the results in Figure 4 from the main text for $\gamma = 9.0$ and $\alpha = 2.0$. For other values of the cost function parameters, γ and α, the optimal stake, σ^*, the resulting profit, Π^*, and the promotion gap, Δp^* have similar shapes. These figures are available upon request.



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



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



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

Figure IA.7: Optimal premium-cost ratio, equilibrium profit, and promotion gap, as functions of the productivity-cost ratio, θ, for $\gamma = 9.0$ and $\alpha = 2.0$.