Discrimination in Promotion

Anja Prummer

Working Paper No. 905 June 2020 ISSN 1473-0278

School of Economics and Finance

Queen Mary University of London

Discrimination in Promotion*

Anja Prummer†

June 1, 2020

Abstract

Why do women hit the glass ceiling? Women are hired, but then fail to rise through the ranks. We propose a novel explanation for this pattern, namely preference- and belief-free discrimination. In our setting, an employer can increase effort by inducing differential value distributions for a promotion across workers, who compete for the promotion by exerting effort. Initially, workers possess the same distribution of valuations. Introducing inequality between workers makes them more recognisable, reducing their information rent, which in turn increases effort. However, higher inequality reduces competition. If value is redistributed, the reduction in information rent outweighs the loss in competitiveness, making discrimination between workers optimal.

Keywords: Discrimination, Mechanism Design, Information Design
JEL codes: D82, J16.

*We are grateful to seminar participants at Duke, UNC and Southern Economic Association. All remaining errors are ours. First draft: February 2020.

†Contact: School of Economics and Finance, Queen Mary University of London, Mile End Road, London, E1 4NS, UK, a.prummer@qmul.ac.uk

1 Introduction

The gender gap in promotions remains remarkable, as evidenced by the ubiquitous discussion on the "glass ceiling effect" (Cotter, Hermsen, Ovadia, and Vanneman (2001)). This glass ceiling is particularly pronounced among lawyers, in finance as well as in academia.[1] These men and women originate from top graduate programs, leaving little doubt about their high ability, and still display very different career trajectories: Women remain stuck at lower tier jobs, unable to rise through the ranks, hitting the invisible glass ceiling.

We propose a novel explanation for the glass ceiling effect, namely employers, who find it optimal to influence how much workers value a promotion, leading to differential effort and ultimately a disparity in promotion attainment. Even though employers do not display differences in beliefs or preferences, they still treat or reward workers distinctly, leading to discrimination.

In our model, workers compete for a promotion by exerting effort. Their effort choice depends on the valuation for the promotion, which is private information. The employer only knows the distribution of the valuation, which is a priori identical for the workers. While workers are identical in their payoff-relevant characteristics, they differ in their label. The employer maximises the sum of workers' effort, their total effort, by designing an optimal mechanism and additionally, by influencing workers' distribution of values for the promotion. If the firm is restricted to *dispersing value*, which entails a weakly lower average value for the promotion, then total effort can increase or decrease. Total effort decreases if we force adjustments to be to distributions which are first order stochastically dominated by the original value distribution. However, if we allow for any distribution that leads to a weakly lower average value, a setting that includes second order stochastic dominance, then the employer increases total effort by adjusting valuations. If the employer has additionally the possibility to *reallocate value* across workers by offering a differential wage schedule, then he finds it always optimal to do so: he will support one worker, at the expense of the other, leading to inequality in promotion of initially perfectly identical workers.

As adjusting workers' distributions is optimal, any differences in attitudes towards or beliefs about workers, or marginal differences in initial distributions are compounded– in light of the numerous well-documented disparities across gender it is therefore natural to suppose that the adjustment in value is not in favour of women, leading to women missing out on promotions.[2]

In addition to offering a novel explanation for the glass-ceiling effect, our model makes a contribution by allowing the employer to influence the value distribution of workers on top of

[1] In banking, women achieved gender parity among employees, but only 38% of middle management were women. This gap widens further at the executive level where only 16% of employees are women, see Ferrary (2017). Women hold only 33% of tenured Professor positions in the US, see data from the Integrated Postsecondary Education Data System (IPEDS) for 2015-2016 from the National Center for Education Statistics. Female lawyers achieve gender parity among non-partners, while less than a quarter of partners are women, https://www.law360.com/articles/1162800/glass-ceiling-slow-to-break-for-female-attys-in-2018.

[2] Men and women display differences in the number and length of career interruptions and overall labour force experience (Bertrand, Goldin, and Katz (2010), Gayle and Golan (2011)), differences in performing job tasks with low-promotability (Babcock, Recalde, Vesterlund, and Weingart (2017)), differences in competitiveness (Gneezy, Niederle, and Rustichini (2003), Niederle and Vesterlund (2007), Dohmen and Falk (2011)) and exogenous differences in hours worked at home (Erosa, Fuster, Kambourov, and Rogerson (2017)).

designing the optimal mechanism. In our setting, the employer maximises total effort, that is the sum of effort across workers, subject to a constraint on the distribution.

If the employer does not face a constraint, he sets the value for the promotion as high as possible as a higher valuation leads to higher effort. Creating such a high valuation, however, may be prohibitively costly in terms of remuneration. Supposing that the employer has a limited budget to induce workers to exert effort, we ask whether allocating this budget differentially can be beneficial. It turns out, that this is indeed the case.

We consider two possible constraints, namely dispersing value and reallocating value. The former may be achieved by creating a toxic or cut-throat environment, which perpetuates beyond the current position and changes the original valuations or aspirations workers have. The latter could be attained by offering differential payment conditional on reaching the promotion, depending on a worker's label.

Workers start out with a value distribution. We assume that the value of the distribution cannot be negative and there is an upper bound on the distribution, a maximal value, $\overline{\omega}$.

If an employer can only influence the value distribution of a single worker, such that the average value for the promotion does not increase compared to the initial distribution, the employer always adjusts the distribution. The effort maximising distribution for the worker is binary, with mass at $\overline{\omega}$ and zero, such that the mean of the distribution equals that of the initial distribution. First, the employer prefers the highest possible expected value. As this value is bounded above by the expected value of the initial distribution, the expected value of the adjusted distribution is the same as that of the original one. In order to extract the entire expected value, the employer has to create a binary distribution with positive mass at exactly one positive value–reducing the information rent to zero. However, adjusting the distribution has also an impact on the effort choices of the other workers. Workers exert higher effort the greater their probability of attaining the promotion. Facing a co-worker with a very high value makes it less likely for them to receive the promotion, diminishing their effort. Therefore, it is optimal for the employer to set the value in the adjusted distribution as high as possible as this reduces the probability for the other worker to face such a high value competitor and therefore their effort remains high. The binary distribution with mass at zero and the maximal value, $\overline{\omega}$, is optimal for all workers, if the firm can adjust the distribution for all of them. However, in this case, the firm can also create a single worker as an "insurance": there can be a single worker whose distribution is degenerate with all mass at the average of the original distribution. If all of the other workers turn out to have zero valuation, the promotion is then given to the worker ,whose valuation equals the mean, with certainty.

Note that this setting encompasses a constraint that restricts the employer to adjust to distributions that are second order stochastically dominated by the original distribution. Choosing a riskier distribution is always beneficial for the firm. However, if the employer can only adjust to distributions that are first order stochastically dominated, then it turns out this is never optimal. In order to make the distribution sufficiently narrow so as to make the value of the worker sufficiently precise, the employer is required to reduce the average value by so much as to make the adjustment not worthwhile.

Therefore, the firm only has an incentive to adjust values if it can make values more extreme. Firms thus have an incentive to create a culture that require long hours and extremely high levels of dedication in order to attain a promotion, which is very appealing to some workers. At the same time, this leads to other workers disliking such an environment and less effort. This reduction, however, is more than compensated by the extensive effort by those striving for the promotion. In this setting, initial differences in distributions can be multiplied. If women start out with a slightly lower average value for the promotion, then it is less profitable for the employer to influence their distribution, especially if the employer is constrained in the how many distributions he can adjust.

We then turn to the case, where the employer can reallocate value between workers by paying them differently upon receiving the promotion. Workers start out with the same value distribution, denoted by $G(\cdot)$ and we assume that the overall distribution for workers must remain constant, that is for two workers the overall distribution is required to be $2G(\cdot)$. It turns out that in this setting it is again optimal to induce different distributions. For two workers, it is always better to create one worker whose values lie below the median of $2G(\cdot)$ and another one, whose values are above the median of $2G(\cdot)$, that is to create maximally distinct value distributions. By creating these distinct distributions, the employer again can extract higher effort due to a reduced information rent. The employer knows much more precisely, what value the worker has compared to keeping the distributions the same and thus has to leave less information rent on the table. This gain due to a reduction in information rent also offsets the reduced competition between workers.

Therefore, employers will also have an incentive to create differences between a priori identical workers, a result in line with the pronounced gender wage gap higher up in the hierarchy.[3]

Related Literature We contribute to different strands of the literature, which we discuss in turn.

Discrimination We contribute to the literature on discrimination, which traditionally assumes either taste-based (Becker (1957)) or statistical discrimination (Phelps (1972), Arrow (1973)); that is, discrimination emerges due to preferences or beliefs.[4] A notable exception is Winter (2004), where discrimination emerges in order to induce workers, whose effort is complementary, to exert effort in a team moral hazard problem. Like in Winter (2004), discrimination is belief- and preference-free. In our setting, discrimination induces higher effort due to a reduction in the information rent an employer has to pay to the worker, which outweighs the loss in competition. Further, our theory can shed light on why gender balance is reached at the entrance levels, while women fail to be promoted–contrary to existing theories. If there was a distaste

[3] See for instance, Merluzzi and Dobrev (2015), Bronson and Thoursie (2019).

[4] Models of statistical discrimination have been developed in Peski and Szentes (2013), Gu and Norman (2020), Lang (1986), Moro and Norman (2003), Lundberg and Startz (1983), Fershtman and Pavan (2020). Fang and Moro (2011) provides an overview of models of statistical discrimination. Kamphorst and Swank (2016) propose a different form of discrimination that still depends on beliefs. Akerlof (1985) allows for discrimination from the customer side.

toward women, they would not even be hired. Further, there is less information about the ability of women available at the hiring stage, which should lessen statistical discrimination– a feature that emerges in Bohren, Imas, and Rosenberg (2019)'s model of dynamic statistical discrimination. Thus, we propose a novel mechanism for discrimination, which is supported by empirical regularities. Discrimination in promotion has been explicitly analysed by Milgrom and Oster (1987) and Lazear and Rosen (1990), again in the presence of initial differences between workers, while we show that it optimal to introduce these differences.[5]

Information Design We also relate to the literature in information design in mechanisms. Contrary to our setting, the literature focuses on a buyer and seller relationship, see Condorelli and Szentes (2020), Roesler and Szentes (2017), Bobkova (2019). Additionally, in these settings, the buyer (our worker) can make a costly investment or an information choice, which the seller (our employer) takes as given. We focus on the choice of the *employer* to influence the valuation, subject to a *constraint*, thus departing in two crucial ways from previous work.

Inequality in Contests Last, we contribute to the literature on inequalities in contests. Mealem and Nitzan (2016) provide a survey of discrimination in contests, documenting that it is never optimal to introduce discrimination, while under some restrictive assumptions it may be beneficial to allow asymmetries to persist– in contrast to our setting, where it is optimal to *introduce* differences to a priori identical workers. Relatedly, in contests with sabotage, sabotage affects welfare adversely (Chowdhury and Gürtler (2015)). Specifically, Calsamiglia, Franke, and Rey-Biel (2013) show in an experiment that inequalities between contestants are harmful, while affirmative action is optimal (Franke (2012)). Li and Yu (2012) highlights that in a contest with two unequal participants, the designer makes them equal in order to obtain the highest payoff, contrary to our finding.

The remainder of the paper is organised as follows. In Section 2, we introduce the model. We solve for the optimal mechanism and distribution subject to a constrained in Section 3. We provide a benchmark result for the optimal distribution without constraints in Section 3.1. We analyse the case of dispersion, where we keep the average valuation weakly below the initial expected value in Section 3.2. In this context, we also characterise the optimal distribution if the employer is required to restrict attention to a distribution that is first-order and second-order stochastically dominated. We then turn to the setting in which the employer can reallocate value across workers in Section 3.3.

2 A Model of Discrimination in Promotion

We begin by introducing the worker's problem, before turning to the problem of the employer. The employer can influence workers' distributions of values when implementing an optimal mechanism. We first summarise classical contributions in independent private value mechanism

[5] Milgrom and Oster (1987) show that more able workers are not promoted to keep their high ability secret from competitors–but only if these workers are disadvantaged from the beginning. Lazear and Rosen (1990) reasons that gender differences outside the workplace lead to differential promotions.

design which constitute a key building block for our results, before focusing on the employer's problem of how to design value distributions.

2.1 Workers' Problem

Workers compete for a promotion by exerting effort. Every worker i, for $i \in \{A, B\}$, has a value for the promotion v_i which is private information and is independently distributed according to a cumulative distribution F_i with support $V_i = [\alpha_i, \omega_i]$. In the baseline setting, workers are initially identical in their value distribution and so $F_i = G$ for all i. Workers are faced with a direct mechanism, $(x(\mathbf{v}), e(\mathbf{v}))$ which for any profile of reported values $\mathbf{v} = (v_A, v_B)$ specifies the likelihood that i gets promoted, $x_i(\mathbf{v})$, and the effort to exert $e_i(\mathbf{v})$. Each worker maximises his expected payoff when choosing to report their valuation which amounts to

$$v_i x_i(\mathbf{v}) - e_i(\mathbf{v}). \tag{1}$$

2.2 Employer's Problem

The employer maximises the sum of efforts across workers, which we call *total effort*, by implementing an incentive compatible (all workers benefit by reporting their valuation truthfully) and individually rational (all workers benefit by participating) mechanism, while influencing the distribution of values for the promotion. We begin by revisiting the mechanism design problem while keeping the distribution as given, before endogenising the distributions of values.

Classical Mechanism Design We build on seminal results in Myerson (1981) to characterise total effort in the employer optimal mechanism for given value distributions. By the revelation principle, it is without loss for the employer to restrict attention to direct mechanisms in which the employer asks workers for their values and workers have an incentive to truthfully reveal these. Thus, the employer will set for all possible profiles of values $\mathbf{v}$ an effort rule $e(\mathbf{v})$ specifying the effort that each worker is expected to deliver and an allocation rule $x(\mathbf{v})$ pinning down the probability of promotion for any worker. The employer sets these two rules to maximise total effort subject to incentive compatibility and individual rationality. By the revenue equivalence principle we know that the effort rule will be fully pinned down by the allocation rule if the mechanism is incentive compatible. The characterisation of the optimal mechanism is based on the *virtual valuation*, which specifies the marginal contribution of worker i with value v_i to total effort. Formally, when denoting by f_i the probability distribution for worker i, the virtual value is defined as

$$\psi_i(v_i) = v_i - \frac{1 - F_i(v_i)}{f_i(v_i)}. \tag{2}$$

Taking expectations over the virtual valuations yields the *virtual surplus*,

$$\textstyle\sum_i \psi_i(v_i) x_i(\mathbf{v}). \tag{3}$$

Myerson (1981) establishes that total effort in an incentive compatible mechanism, in which the lowest type of the worker is indifferent about participating, amounts to expected virtual surplus,

$$TE\left(F_A, F_B\right) = \mathbb{E}_{\mathbf{v}}\left[\sum_i e_i(\mathbf{v})\right] = \mathbb{E}_{\mathbf{v}}\left[\sum_i \psi_i(v_i)x_i(\mathbf{v})\right]. \tag{4}$$

The optimal mechanism is therefore found by maximising expected virtual surplus subject to incentive compatibility which requires that the probability of receiving the promotion is increasing in worker's valuation, that is the virtual value is regular, $\psi_i'(v_i) \geq 0$ for all i. If the virtual value is regular, then the optimal mechanism awards the promotion with certainty to the worker with the highest non-negative virtual valuation. However, for irregular distributions, more complicated allocation rules need to be devised to fulfil incentive compatibility.

Quantile Space In order to characterise the optimal mechanism when regularity fails, it is useful to move away from the traditional approach in the value – virtual value space, and to instead translate the problem to the *quantile space*, see Hartline (2013). For any distribution F_i, define the quantile associated with value $v_i \in V_i$ as $q_i(v_i) = 1 - F_i(v_i)$. Similarly, define the value $v_i(q)$ associated to any quantile $q \in [0,1]$ as $v_i(q) = \sup\{v_i \in V_i | q \geq 1 - F_i(v_i)\}$. This definition encompasses cases in which the cumulative distribution is discontinuous. If F_i is continuous and strictly increasing, we obtain $v_i(q) = F_i^{-1}(1-q)$. In the quantile setting, virtual values amount to

$$\phi_i(q) = \psi_i(v_i(q)) = v_i(q) + v_i'(q)q = \frac{\partial\left(\Phi_i(q)\right)}{\partial q}, \tag{5}$$

where $\Phi_i(q) \equiv v_i(q)q$. Further, the interim promotion probability in the quantile space is denoted by $y_i(q) = \mathbb{E}_{\mathbf{v}_{-i}}\left[x_i(v_i(q), \mathbf{v}_{-i})\right]$ and must be non-increasing by incentive compatibility.

Value Design Problem Having found the optimal mechanism, conditional on the distribution, we can now focus on the employer's problem of designing the distribution. The employer can adjust the distribution of values for each worker, F_A and F_B, subject to some constraints, which to capture the ability to disperse or reallocate value between workers. We assume a maximal value $\overline{\omega} < \infty$ for the promotion. The employer's value design problem in the quantile space amounts to

$$\max_{F_A, F_B} \quad TE\left(F_A, F_B\right) = \mathbb{E}_q\left[\phi_A(q)y_A(q)\right] + \mathbb{E}_q\left[\phi_B(q)y_B(q)\right] \tag{6}$$
$$\text{s.t. constraints on distribution}$$

Next we discuss a classes of constraints the classes of constraints the analysis focuses on and provide some motivation.

Value Dispersion We consider constraints that allow to disperse value for each worker by creating the appropriate environment. We assume that the employer cannot increase the average value for the distribution, he is therefore restricted to select distributions with a weakly lower mean. Formally,

$$\mathbb{E}_{v_i \sim F_i}[v_i] \leq \mathbb{E}_{v \sim G}[v]. \tag{7}$$

Constraint (7) encompasses the set of distributions that are second order stochastic dominated by the worker's initial distribution. A further natural constraint is that of first order stochastic dominance

$$F_i(v_i) \geq G(v) \text{ for all } v \in [0, \overline{\omega}], \tag{8}$$

achieved by reducing the overall valuation for the promotion.

Value Reallocation We also consider settings in which value can be reallocated across workers, for example by paying differential wages after promotion. We consider a setting in which the employer can design any two distributions, F_A and F_B, provided that the sum of their means is smaller than the sum of the means of their undistorted distributions, formally,

$$\mathbb{E}_{v_A \sim F_A}[v_A] + \mathbb{E}_{v_B \sim F_B}[v_B] = 2\mathbb{E}_{v \sim G}[v]. \tag{9}$$

If the employer aims to avoid detection by keeping the overall distribution of valuations fixed, then he faces a more stringent constraint on the distributions, namely:

$$F_A(v_A) + F_B(v_B) = 2G(v) \text{ for all } v \in [0, \overline{\omega}]. \tag{10}$$

Both of these frameworks capture the ability of the employer to freely redistribute value among workers conditional on promotion either deterministically or stochastically, for instance by creating different compensation packages conditional on promotion.

3 Constrained Discrimination

We discuss in turn the optimal value distribution for the employer from value dispersion and reallocation, after highlighting the optimal distribution if the employer did not face any constraints.

3.1 Designing Value without Constraints

Suppose the employer maximises (6) without any constraints. If so, the employer design a value distribution such that at least one worker values the promotion at $\overline{\omega}$ with certainty. Such a value design would necessarily be optimal as the employer would generate the maximal surplus, which amounts to $\overline{\omega}$, and leave no information rents to workers.

Proposition 1. *If the employer can adjust the value distribution for both workers, then it sets measure one to value $\overline{\omega}$ for at least one worker.*

This result highlights two forces at play in our model that are present in a number of results. First, the employer would like for the value of the workers to be as precise as possible, leading to an atom. Knowing precisely what the value of the worker is reduces the information rent that the employer has to pay to the worker in order to ensure incentive compatibility. To see this, note that the *information rent* amounts to

$$v_i - \psi_i(v_i) = \frac{1 - F_i(v_i)}{f_i(v_i)}. \tag{11}$$

In the quantile space, it is straightforward to show that the information rent is zero if the distribution consist of a single atom. Therefore, creating a more narrow distribution, making the value easily recognisable leads to a higher total effort for the employer. Second, the employer aims to increase the value for the promotion as much as possible, because a higher value induces workers to exert higher effort. Therefore, the employer chooses to place all mass at the highest possible value for promotion. In the unconstrained optimal value design, it suffices to increase the value for one worker to $\overline{\omega}$, because the employer, knowing the value of that worker with certainty, is able extract the full surplus from them without ever promoting the other worker. If the employer adjusts the distribution of both workers, then each worker obtains the promotion probability 1/2, leading in expectation to the same total effort of $\overline{\omega}$.

3.2 Value Dispersion

We now assume that the employer faces a constraint in how he can adjust the distribution. We first consider constraint (7), which imposed weakly lower means, before turning to the optimal mechanism if the new distribution is required to be first order stochastically dominated.

Constraint: Lower Means It is instructive to keep the distribution of worker A fixed and focus on the adjustment of the value distribution of worker B, before turning to the optimal adjustment of both workers.

The optimal distribution of values for worker B allocates mass at the highest possible value of the distribution $\overline{\omega}$ and mass at zero, such that the expected value remains unchanged, compared to the initial distribution. This is formalised in Theorem 2, which characterises the optimal distribution.

Theorem 2. *Total effort is maximised among all distributions F_B with $\mathbb{E}_{v_B \sim F_B}[v_B] = \mathbb{E}_{v \sim G}[v]$ by*

$$F_B^*(v_B) = \begin{cases} 1 & \text{if } \; v_B = \overline{\omega} \\ 1 - \frac{\mathbb{E}_{v \sim G}[v]}{\overline{\omega}} & \text{if } \; v_B < \overline{\omega} \end{cases} \tag{12}$$

To see that the expected value remains constant, note that

$$\mathbb{E}_{v_B \sim F_B}[v_B] = F_B^*(\overline{\omega})\overline{\omega} + (1 - F_B^*(\overline{\omega}))\,0 = F_B^*(\overline{\omega})\overline{\omega} = \frac{\mathbb{E}_{v \sim G}[v]}{\overline{\omega}}\overline{\omega} = \mathbb{E}_{v \sim G}[v] \tag{13}$$

The employer chooses a maximal spread of valuations for one agent. As we already demonstrated in the benchmark case with no constraint, the firm would like to make a worker's distribution as precise as possible as this makes the workers's valuation easily identifiable. It therefore will allocate mass to two mass points, which replicate the average value of the original distribution. The firm knows exactly worker B's value for promotion and does not have to pay any information rent, and can thus increase the effort of the worker.

Beyond affecting the virtual value, a change in distribution affects the probability of promotion for *both* workers. For this reason, it is optimal to allocate a positive probability to the highest possible value instead of some other $\omega_B < \overline{\omega}$. The employer extracts the same effort from worker B independently of whether he allocates positive mass to some $\omega_B < \overline{\omega}$ or $\overline{\omega}$. However, allocating positive probability to the highest positive value has an effect on worker A's effort choice. As we keep the expected value of the distribution fixed, choosing positive probability for a higher value implies that the probability is lower than if the employer chose a lower value. But a lower probability associated with B's highest value leads to a higher probability that B has zero valuation and in turn, a higher probability for A to obtain the promotion. A's effort decreases in B's probability of having a high valuation and therefore it is optimal to keep this probability as low as possible – by choosing $\overline{\omega}$.

The intuition behind Theorem 2 is summarised graphically in Figure 1. The adjustment in distribution allows to extract the entire expected value for the promotion as the expected virtual valuation equals the expected value, see the shaded area in Figure 1a. This is more than what the employer can extract from worker A. A's expected virtual value is depicted by the striped area in Figure 1b, which is smaller than the expected value which is given by the shaded area under $v_A(q)$.

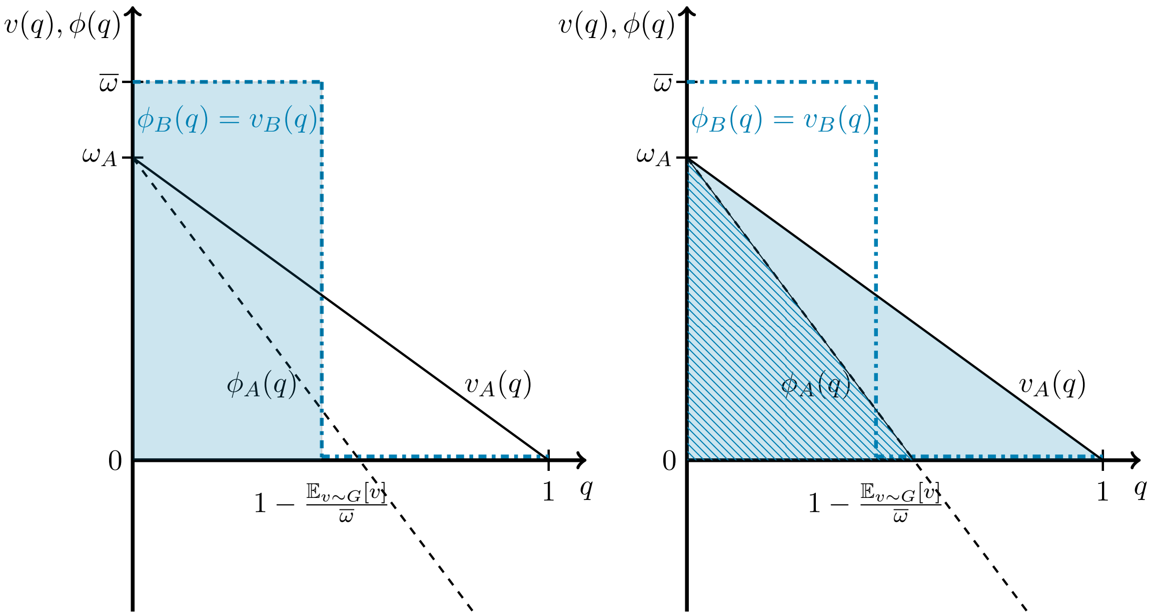
It is worth noting that Theorem 2 encompasses second order stochastic dominance, as the distribution F_B^* is second order stochastically dominated by $F_A = G$,

$$\int_0^v F_A(t)dt \leq \int_0^v F_B(t)dt \quad \text{for all } v \in [0, \overline{\omega}]. \tag{14}$$

Corollary 2.1. *Among all distributions that second order stochastically dominate G, F_B^* maximizes total effort.*

This implies that the employer always prefers a "riskier" value distribution for worker B. The result sheds light on an alternative interpretation of why the firm prefers to adjust the distribution. The employer already has a worker with a more smooth value distribution, worker A. For any realisation, worker A's value for the promotion is relatively similar. Therefore, the firm perceives this worker as a safety option in terms of promotion. To the contrary, worker B either has a very high or zero valuation for the promotion. If the employer discovers that B has

Figure 1: Value Dispersion for Worker B



(a) B's expected virtual value (b) A's expected virtual value

Note: B's expected virtual value is depicted by the shaded area in Figure 1a, which equals $\mathbb{E}_{v\sim G(v)}[v]$, while A's expected virtual value is given by the striped area in Figure 1b. As the shaded area under $v_A(q)$ equals $\mathbb{E}_{v\sim G(v)}[v]$, it follows that A's expected virtual value is smaller than B's. Example assumes that $\alpha = 0$.

a high valuation, then it can extract a high effort and promote worker B. The firm does make a loss if worker B turns out to have zero valuation for the promotion, but it still has worker A to fall back on.

Our results show that it is never optimal to reduce the expected valuation for the promotion. This raises the question of whether it can ever be optimal to adjust the value of a worker if this necessarily comes with a reduction in the expected value. It turns out this is indeed the case, as long as the expected value is not reduced by "too much", which is formalised in Proposition 3.

Proposition 3. *Spreading worker B's value such that*

$$F_B^+(v) = \begin{cases} 1 & \text{if } v = \overline{\omega} \\ 1 - \frac{m}{\overline{\omega}} & \text{if } v < \overline{\omega} \end{cases} \quad (15)$$

where $m \geq \mathbb{E}_{v\sim G(v)}[\psi(v)]$ yields higher total effort compared to no adjustment.

Note that $m \geq \mathbb{E}_{v\sim G(v)}[\psi(v)]$ is equivalent to $F_B^+(\overline{\omega})\overline{\omega} \leq E_{v\sim G(v)}[v]$. This implies that value dispersion is optimal for the firm a long as the costs of doing so, given by the reduction in expected value, is lower than the reduction in information rent the employer has to pay. Thus, even if the firm loses some value on average for the promotion by adjusting the distribution, doing so still increases the total effort, as long it is outweighed by lowering the information rent the employer needs to give to the worker in order to induce them to reveal their value.

Having discussed the case where the distribution of one worker can be adjusted, we now allow for the values of both workers to be influenced, keeping expected values fixed. It is still optimal to adjust the value of one worker as outlined in Theorem 2. For the second worker, there are then two optimal adjustments, as summarised in Proposition 4.

Proposition 4. *Assume* $\mathbb{E}[v_i] \leq \mathbb{E}_{v\sim G}[v]$, $\forall i \in \{A, B\}$. *Total effort is maximised if the employer sets for one worker* i,

$$F_i^*(v) = \begin{cases} 1 & \text{if } v = \overline{\omega} \\ 1 - \frac{\mathbb{E}_{v\sim G}[v]}{\overline{\omega}} & \text{if } v < \overline{\omega} \end{cases} \tag{16}$$

and for worker j *either (i)* $F_j^*(v) = F_i^*(v)$ *or (ii)*

$$F_j^*(v) = \begin{cases} 1 & \text{if } v \geq \mathbb{E}_{v\sim G} \\ 0 & \text{if } v < \mathbb{E}_{v\sim G} \end{cases} \tag{17}$$

We omit the proof as the result follows immediately from the proof of Theorem 2. If the employer can adjust the distribution for both workers, he will do so. There are two possible outcomes: the employer either creates only workers, who have with a fairly small probability a very high value for the promotion. If they have a high valuation, they also exert the highest possible effort. On the other hand, with a large probability, a worker does not aim for promotion and exerts zero effort. This translates into a work environment, with few superstars that work incredibly hard and a large set of workers that simply do the basics without striving for a promotion. Alternatively, the firm creates one value distribution which assigns a small probability to the highest value and for the other worker a distribution that assigns all mass to the expected value of the original distribution. The employer still aims to create a worker, who either exerts the highest possible effort if he turns out to have a high valuation or zero effort, creating again a superstar environment. But such an environment comes at the cost of not having a worker with positive valuation to be promoted. The firm can insure against this by creating a worker, who serves as an insurance. If the worker with the stark difference in values turns out to not value the promotion, then the firm can still allocate it to the worker with the average valuation.

Interestingly, the latter adjustment yields the same total effort as creating two workers with either high or zero valuation. However, there can never be more than one worker with value at the mean, see again the proof of Theorem 2. With many workers, the firm creates an environment where all workers are overachievers with a small probability.[6] The small probability of being an overachiever gives the illusion that once the worker has the high value, he obtains the promotion almost certainly. This allows the firm to extract high effort from all such workers, that is the firm can extract the expected value from every single worker.

Constraint: FOSD We keep again the value distribution of worker A fixed, and adjust the distribution of worker B such that $A's$ distribution first order stochastically dominates $B's$

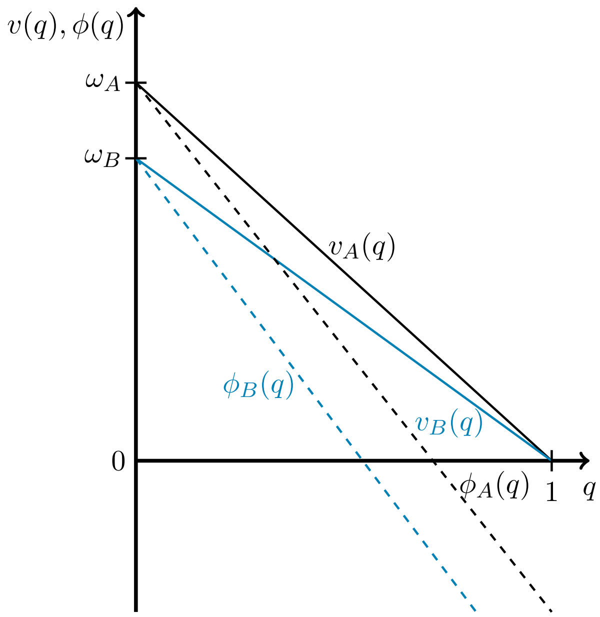
[6] The employer is indifferent between the nth worker having mass one at the mean or the same bimodal distribution as all the other workers.

distribution, $G(v) = F_A(v_A) \leq F_B(v_B)$. First order stochastic dominance implies a ranking of $q_A(v)$ and $q_B(v)$:

$$q_B(v) = 1 - F_B(v) \leq 1 - F_A(v) = q_A(v) \tag{18}$$

It follows that for a given quantile, the value for A is higher than that for B, formally $v_A(q) \geq v_B(q)$ for any $q \in [0,1]$. This observation is summarised in Figure 2. For any distribution B that is first order stochastically dominated by a distribution A, the value of A is higher than that of B for every quantile.

Figure 2: A FOSD B



This observation is instrumental in proving Theorem 5.

Theorem 5. *Total effort is maximised by* $F_B^*(v) = F_A(v) = G(v)$ *among all distributions* $F_B(v) \geq G(v)\ \forall v$.

If the employer can only adjust the distribution such that it is first order stochastically dominated by the original distribution, then it will never adjust the value. The firm can still make the distribution more precise, which reduces the information rent, but this comes at the cost of reducing the value for promotion agent's have. The gain in information rent never outweighs the loss in value.

To show this, we consider a distribution $F_B(v)$ that differs from and is first order stochastically dominated by $F_A(v)$. This implies that there are some values for which B's CDF is strictly above that of A. We then construct an alternative CDF for B, $\hat{F}_B(v)$, which first order stochastically dominates $F_B(v)$, again with strict inequality in distribution function for some v. We show that total effort is higher under $\hat{F}_B(v)$. Again, a change in distribution has two effects,

(i) it affects the virtual value and (ii) it affects the probability of promotion for *both* workers. To simplify our proof, we keep the allocation rule fixed and show that the optimal total effort increases due to the adjustment of virtual values from $F_B(v)$ to $\hat{F}_B(v)$. Total effort under the old allocation rule is a lower bound, as an adjustment in the allocation rule must increase effort – otherwise the firm would not select a new allocation rule. As this holds for any $\hat{F}_B(v)$, it follows that in optimum B's distribution must equal $A's$ distribution or rather, the original distribution G. This is also demonstrated in Figure 2, where the virtual values decrease for every quantile if the distribution is first order stochastically dominated by $G(v)$. This implies that $G(v)$ yields the highest information rent among all first order stochastically dominated distributions.

We have shown that if the employer can influence the worker's distribution by dispersing value, he will do so, unless he is restricted to distributions which are first order stochastically dominated. In particular, he will create a bimodal distribution, which leads to workers valuing the promotion either very highly or not at all. Such an environment can create gender disparities if women end up valuing the promotion under these circumstances less compared to men.

3.3 Value Reallocation

We turn to the case where the firm cannot only disperse the value We first consider the case of keeping the sum of expected values fixed before turning to the case where the sum of the original distributions remains constant.

Constraint: Lower Overall Means If the employer can reallocate value between workers, only keeping the overall expected value fixed, that is $\mathbb{E}[v_A] + \mathbb{E}[v_B] \leq 2\mathbb{E}_{v\sim G}[v]$, then it will be optimal to re-allocate value as described in Proposition 6.

Proposition 6. *Let* $\mathbb{E}[v_A] + \mathbb{E}[v_B] \leq \mathbb{E}_{v\sim G}[v]$. *The firm assigns mass 1 to worker i's valuation* $v_i = \mathbb{E}[v_A] + \mathbb{E}[v_B]$, $i \in \{A, B\}$.

We again omit the proof as this result follows immediately from Proposition 1, where we have shown that it is optimal to set the value of one worker as high as possible. First, we know that choosing a mass point at exactly one positive valuation is optimal as it reduces the information rent to zero. Next, it cannot be optimal to keep the values of both workers to be positive. Either one valuation is lower than the one of the other worker, resulting in a sure winner or loser of the promotion. Then, re-allocating value from the certain loser to the winner increases total effort. If the values are the same, then the allocation probability is not one, which reduces effort from both workers, making further re-distribution a profitable deviation. It therefore follows immediately that allocating all value to one worker maximises total effort.

Constraint: Fixed Overall Distribution It may however not always be feasible for the employer to reduce the valuation of their workers to some points. In order to capture this, we consider the reallocation of values keeping the overall distribution fixed, that is $F_A(v) + F_B(v) =$

$2G(v)$. It turns out to be beneficial for the employer to adjust the value of the distribution, creating two maximally different value distributions, see Theorem 7.

Theorem 7. *Reallocating value and creating differences between workers' valuations is always optimal.*

We show that splitting the distribution, such that worker B's values are below the median of $2G(v)$, while worker A's values are above the median of 2 $G(v)$, always yields a higher total effort than keeping workers' value distributions identical. If values are more spread out, then the employer is required to pay a higher information rent in order to induce the worker to reveal their valuation truthfully. Narrowing the distribution for both workers therefore reduces information rent. Interestingly, this is only due to the gain in information rent from worker B. Worker A's virtual values are the same as in the original distribution for the values above the median. Our proof arrives at this result by keeping the probability of receiving the promotion as under the initial distribution, which is identical for both workers. As noted before, the total effort derived under the adjusted virtual value, but fixed allocation probability is a lower bound on the total effort under the new distribution. We show that this lower bound still yields a higher total effort than keeping identical distributions. This holds if the promotion is always allocated, and is exacerbated if there is some exclusion of values. The values for which exclusion occurs is more limited with distinct distributions, making it more likely that the promotion is awarded, which further increases total effort.

Consequently, an employer who hires two identical workers, will induce differences in how they value the promotion, through differential wages or bonus payments upon promotion.These differences naturally emerge across two distinct groups, such as men and women, even if there are no differences in initial distributions, but employers display differences in beliefs or preferences. Therefore, marginal disparities in attitudes towards men and women can be multipled as it is optimal to induce disparities between workers.

4 Conclusion

We provide a novel explanation for the glass-ceiling effect, by allowing an employer to establish the optimal mechanism for the allocation of a promotion and additionally, to design the value for the promotion subject to a constraint. We allow for a large, natural set of constraints, including first and second order stochastic dominance. If the employer is restricted to value dispersion of a distribution, that is he can influence the distribution of each worker, but cannot increase their average valuation, then it is optimal to create a bimodal distribution, such that each worker has a very high value for the promotion with a small probability or is not interested in the promotion at all. If the employer can reallocate values for promotion by choosing differential wages or bonus schedules across workers, then it is optimal to always create differences between initially identical workers. We therefore provide a rationale for how small differences across gender or attitudes can evolve into large difference– through employers who benefit from unequal workers.

A Appendix

Proof of Proposition 1: No Constraints Note first that it is not possible to obtain a total effort higher than $\overline{\omega}$:

$$\mathbb{E}_q\left[\phi_A(q)y_A(q)\right] + \mathbb{E}_q\left[\phi_B(q)y_B(q)\right] \le \mathbb{E}_q\left[v_A(q)y_A(q)\right] + \mathbb{E}_q\left[v_B(q)y_B(q)\right] \tag{19}$$

$$\le \mathbb{E}_q\left[\overline{\omega}y_A(q)\right] + \mathbb{E}_q\left[\overline{\omega}y_B(q)\right] \le \overline{\omega}, \tag{20}$$

where the first inequality follows from $v(q) \ge \phi(q)$, the second equality from $v \le \overline{\omega}$ and the last inequality from $E_q\left[y_A(q)\right] + \mathbb{E}_q\left[y_B(q)\right] \le 1$ where the allocation rule is irrelevant. Therefore, as long as one worker has value $\overline{\omega}$ and he obtains the promotion, the firm extracts the highest possible total effort. ■

Proof of Proposition 2: Value Dispersion, 1 Worker We compare total effort when setting F_B^* to the total effort the seller would obtain when setting some other distribution F_B such that $\mathbb{E}_{v\sim F_b}[v] = \mathbb{E}_{v\sim G}[v]$. In a quantile setting, for $\overline{q} = 1 - F_B^*(0)$, the virtual value for the distribution F_B^* amounts to

$$\phi_B^*(q) = \begin{cases} \overline{\omega} & \text{if } q < \overline{q} \\ 0 & \text{if } q > \overline{q} \end{cases} . \tag{21}$$

Total effort under distribution F_B^* is given by

$$\mathbb{E}\left[\phi_A(q)y_A^*(q)\right] + \mathbb{E}\left[\phi_B^*(q)y_B^*(q)\right], \tag{22}$$

while the effort under the alternative F^B amounts to

$$\mathbb{E}\left[\phi_A(q)y_A(q)\right] + \mathbb{E}\left[\phi_B(q)y_B(q)\right]. \tag{23}$$

By optimality, we have that

$$\mathbb{E}\left[\phi_A(q)y_A^*(q)\right] + \mathbb{E}\left[\phi_B^*(q)y_B^*(q)\right] \ge \mathbb{E}\left[\phi_A(q)y_A(q)\right] + \mathbb{E}\left[\phi_B^*(q)y_B(q)\right]. \tag{24}$$

Therefore, to prove the result it suffices to establish that

$$\mathbb{E}\left[\phi_B^*(q)y_B(q)\right] - \mathbb{E}\left[\phi_B(q)y_B(q)\right] \ge 0. \tag{25}$$

For any $q < \overline{q}$, it must be that $\phi_B^*(q) = \overline{\omega} \ge \phi_B(q)$, since $\phi_B(q) = v_i(q) + v_i'(q)q \le v_i(q) \le \overline{\omega}$. Instead, for $q \ge \overline{q}$ and $y_B(q) > 0$, it must be that $\phi_B(q) \ge \phi_B^*(q) = 0$, since the promotion will only be given to someone with non-negative virtual value. As incentive compatibility requires $y_B(q)$ to be non-increasing, we have that if

$$\mathbb{E}\left[\phi_B^*(q) - \phi_B(q)\right] \ge 0 \quad \Rightarrow \quad \mathbb{E}\left[(\phi_B^*(q) - \phi_B(q))y_B(q)\right] \ge 0. \tag{26}$$

By construction, the expected virtual values satisfy

$$\mathbb{E}\left[\phi_B^*(q)\right] = \mathbb{E}\left[v_B^*(q)\right] = \overline{q\varpi} = \mathbb{E}\left[v_A(q)\right] = \mathbb{E}\left[v_B(q)\right]. \tag{27}$$

The result then obtains, because expected virtual value for F_B satisfies

$$\mathbb{E}\left[v_B(q)\right] \geq \mathbb{E}\left[\phi_B(q)\right], \tag{28}$$

given that $\phi_B(q) - v_B(q) = v_B'(q)q \leq 0$ for all $q > 0$. This establishes that it is never optimal to select a distribution for B such that the expected virtual value is below the expected value. However, there are multiple distribution that allow for the expected value to be equal to the expected virtual value, $\mathbb{E}\left[v_B(q)\right] = \mathbb{E}\left[\phi_B(q)\right]$. Note however that is never optimal to allocate mass to more than two values $v > 0$. Suppose to the contrary, the firm chose such a distribution. Then, the allocation probability must differ across the different valuations for this to be incentive compatible. Otherwise a worker with a higher valuation would pretend to be a worker with a lower valuation and still obtain the promotion with the same probability. It follows that it is not feasible to extract the entire valuation, yielding the contradiction.

We evaluate the effect of the different distributions that reduce information rent to zero on A's effort $\mathbb{E}[\phi_A(q)y_A(q)]$. Note that A only obtains the promotion if B's virtual valuation is smaller than A's:

$$\mathbb{E}[\phi_A(q)y_A(q)] = \int_{q_A} \phi_A(q_A)\left(1 - \hat{q}_B(v_B = \phi_A(q_A))\right) dq_A, \tag{29}$$

where $\hat{q}_B(v_B = \phi_A(q_A))$ is the highest value of q_B for which $v_B \geq \phi_A(q_A)$. Given that B's distribution only has positive mass at one positive value, it follows that A either obtains the promotion with certainty or not. Clearly, equation (29) is maximised for q_B approaching zero for any q_A, for which $\phi_A(q_A) \geq 0$. For $\phi_A(q_A) < 0$, the allocation probability becomes 0. This implies that the maximal effort is bounded above by

$$\mathbb{E}_{v\sim G}[v] + (1 - \hat{q}_B)\mathbb{E}[\phi_A(q)] < \mathbb{E}_{v\sim G}[v] + \mathbb{E}[\phi_A(q)] \tag{30}$$

There are different possible distributions for B that can potentially minimise the probability that B obtains the promotion. We first show that allocating mass at zero and ϖ can indeed lead for q_B to be arbitrary small, a feature other allocations do not posses, making it the optimal distribution.

1. *Mass at ϖ and 0* While it is never possible to set the probability that B obtains the promotion to zero for all positive virtual values of A, $B's$ probability can be made arbitrarily small, by choosing ϖ and allowing for this to become arbitrarily large. This implies that $q_B = \overline{q}$ becomes arbitrarily small. More formally,

$$F_B(0)\int_{t_A}^{\omega_A} \psi_A(v)dF_A(v), \tag{31}$$

where t_A addresses the exclusion restriction. As $\int_{t_A}^{\omega_A} \psi_A(v)dF_A(v)$ is independent of ω_B, the effect is purely driven by $F_B(0)$. As ω_B increases, $F_B(0)$ increases:

$$\frac{\partial\left((1-F(0))\omega_B\right)}{\partial\omega_B} = (1-F(0)) + \omega_B\frac{\partial(1-F(0))}{\partial\omega_B} = 0 \tag{32}$$

We can rewrite this as $(1-F(0))\omega_b = \mathbb{E}[v_A]$ or equivalently, $(1-F(0)) = \frac{\mathbb{E}[v_A]}{\omega_B}$

$$\frac{\mathbb{E}[v_A]}{\omega_B} = \omega_B\frac{\partial F(0)}{\partial\omega_B} \tag{33}$$

$$\Leftrightarrow \qquad \frac{\partial F(0)}{\partial\omega_B} = \frac{\mathbb{E}[v_A]}{\omega_B^2} > 0 \tag{34}$$

And therefore, the expected effort from A increases if ω_B increases. By setting $\overline{\omega}$ sufficiently large, there always exists a $\overline{q}$, such that $\overline{q} < \epsilon$.

2. *Allocate mass 1 to* $v_B = \mathbb{E}_{v\sim G}[v]$: Suppose first that G is regular, which implies no discontinuities. In this case, allocating mass at the mean always affects A's promotion probability as $\phi_A(0) = v_A(0) > \mathbb{E}_{v\sim G}[v]$ and $\phi_A(q)$ is strictly decreasing, such that there is a region for which $\phi_A(q) < \mathbb{E}_{v\sim G}[v]$. To see this note that there exists $v_A(q) \equiv v(q) < \mathbb{E}_{v\sim G}[v]$ and $\phi_A(q) < v_A(q)$. Therefore, the effect on the promotion probability is not arbitrarily small and so such an adjustment cannot be optimal.
Suppose next G has a discontinuity. Denote by q_A^M the quantile such that $\phi_A(q) > \mathbb{E}_{v\sim G}[v]$ for $q \in [0, q_A^M)$ and $\phi_A(q) \leq 0$ for $q \in [q_A^M, 1]$. Total effort is then given by

$$q_A^M\mathbb{E}[\phi_A[q]\big|q < q_A^M] + (1-q_A^M)\mathbb{E}_{v\sim G}[v] \tag{35}$$

Note that $\mathbb{E}[\phi_A[q]\big|q < q_A^M] \leq \mathbb{E}[v_A[q]\big|q < q_A^M]$. Additionally, $q_A^M\mathbb{E}[v_A[q]\big|q < q_A^M] \leq \mathbb{E}_{v\sim G}[v]$ as $\mathbb{E}_{v\sim G}[v] = q_A^M\mathbb{E}[v_A[q]\big|q < q_A^M] + (1-q_A^M)\mathbb{E}[v_A[q]\big|q \geq q_A^M]$. This implies that

$$q_A^M\mathbb{E}[\phi_A[q]\big|q < q_A^M] + (1-q_A^M)\mathbb{E}_{v\sim G}[v] \leq E_{v\sim G}[v] + (1-q_A^M)\mathbb{E}_{v\sim G}[v] \tag{36}$$

This condition only holds with equality if A's distribution is such that mass $\overline{q}$ is allocated to $\overline{\omega}$. Thus, generically, this condition holds with inequality and allocating mass at the mean does not yield optimal total effort.

■

Proof of Corollary 2.1: Second Order Stochastic Dominance We show that F_B^* is second stochastically dominated by $F_A = G$. First note that if $F_B^* = F_A$, then F_B^* is SOSD by F_A by definition. If both distributions are unequal, it holds that

$$\Delta(v) = F_A(v) - F_B^*(v) < 0 \tag{37}$$

for $v \in [0, \overline{v})$, $\overline{v} \in [\alpha_A, \omega_A]$. If follows that

$$\int_0^v \Delta(t)dt = \int_0^v F_A(t)dt - \int_0^v F_B^*(t)dt < 0; \tag{38}$$

for $v \in [0, \overline{v})$. Moreover,

$$\int_0^{\overline{\omega}} \Delta(t)dt = \left[t\left(F_A(t) - F_B^*(t)\right)\right]_0^{\overline{\omega}} - \int_0^{\overline{\omega}} td\left(F_A(t) - F_B^*(t)\right) = 0. \tag{39}$$

This implies that $\int_{\overline{v}}^{\overline{\omega}} \Delta(t)dt > 0$. To establish second order stochastic dominance, it then remains to be shown that $F_A(v) - F_B^*(v)$ is non-decreasing. This holds as F_A is non-decreasing and F_B^* is constant up until $\overline{\omega}$. Therefore for any $v \leq \overline{\omega}$ we established that A's distribution SOSDs B's distribution. ■

Proof of Theorem 5: FOSD Suppose by contradiction that the firm found it optimal to set $F_B \neq F_A$, which implies that $F_B(v) > F_A(v)$ for some v. We show that in this case, there exists a profitable deviation to a distribution $\hat{F}_B$, such that $F_B(v) \geq \hat{F}_B(v)$ with strict inequality for some v and $\hat{F}_B(v) \geq F_A(v)$.

The change from F_B to $\hat{F}_B$ has two effects: (i) it affects the virtual valuation of worker B and (ii) it affects the optimal allocation rule for the promotion. We begin by keeping the allocation rule fixed and show that a change from F_B to $\hat{F}_B$ increases total effort under the same allocation rule. Recall that optimal total effort is given by

$$\mathbb{E}[\phi_A(q)y_A(q)] + \mathbb{E}[\phi_B(q)y_B(q)] = \int_0^1 \phi_A(q)y_A(q)dq + \int_0^1 \phi_B(q)y_B(q)dq. \tag{40}$$

As the allocation rule is unchanged, $\mathbb{E}[\phi_A(q)y_A(q)]$ is not affected, and we only need to establish that

$$\mathbb{E}[\hat{\phi}_B(q)y_B(q)] \geq \mathbb{E}[\phi_B(q)y_B(q)] \quad \Leftrightarrow \quad \int_0^1 \left(\hat{\phi}_B(q) - \phi_B(q)\right) y_B(q)dq \geq 0. \tag{41}$$

Integration by parts implies that

$$\mathbb{E}[\phi_B(q)y_B(q)] = \alpha_B y_B(1) - \mathbb{E}[\Phi_B(q)y_B'(q)] = \alpha_B y_B(1) - \mathbb{E}[qv_B(q)y_B'(q)]. \tag{42}$$

But if so we get that

$$\int_0^1 \left(\hat{\phi}_B(q) - \phi_B(q)\right) y_B(q)dq = (\hat{\alpha}_B - \alpha_B)y_B(1) + \int_0^1 \left(qv_B(q) - q\hat{v}_B(q)\right) dy_B(q). \tag{43}$$

As $\hat{F}_B$ first order stochastically dominates F_B, we know that $\hat{\alpha}_B \geq \alpha_B$ and that $v_B(q) \leq \hat{v}_B(q)$. As the allocation probability is decreasing in q by incentive compatibility, $dy_B(q) \leq 0$, it follows that

$$\mathbb{E}[\phi_A(q)y_A(q)] + \mathbb{E}[\hat{\phi}_B(q)y_B(q)] \geq \mathbb{E}[\phi_A(q)y_A(q)] + \mathbb{E}[\phi_B(q)y_B(q)]. \tag{44}$$

Moreover, if the firm was allowed to set the allocation rule $\hat{y}(q)$ optimally, total effort further increases

$$\mathbb{E}[\phi_A(q)\hat{y}_A(q)] + \mathbb{E}[\hat{\phi}_B(q)\hat{y}_B(q)] \geq \mathbb{E}[\phi_A(q)y_A(q)] + \mathbb{E}[\hat{\phi}_B(q)y_B(q)], \tag{45}$$

or else the employer would prefer to leave allocation rule unchanged. ■

Proof of Theorem 7: Reallocating Value, Fixed Distribution We can restrict attention to regular $G(v)$. If this was not the case and the employer chooses a $\hat{G}(v)$ that is not regular, then we can consider an ironed function $G(v)$, which yields the same total effort, see Hartline (2013), Theorem 3.14, p.78. This then yields a distribution $H(v) = 2G(v)$, which is also regular.

Define v^M, such that

$$\int_{\alpha}^{v^M} dH(v) = 1. \tag{46}$$

This allows us to define the adjusted distributions,

$$F_B(v) = H(v) \qquad \text{if } v \in [\alpha, v^M) \tag{47}$$

$$F_A(v) = H(v) - 1 \qquad \text{if } v \in [v^M, \omega], \tag{48}$$

and we show that this distribution yields a higher total effort than $G(v)$.

The virtual values for H, G, F_A and F_B, respectively, are given by

$$\psi_H(v) = \psi_G(v) = \psi_A(v) = v - \frac{2 - H(v)}{h(v)}, \qquad \psi_B(v) = v - \frac{1 - H(v)}{h(v)} \tag{49}$$

This implies that $\psi_H(v) = \psi_G(v) = \psi_A(v)$ for $v \in [\alpha, v^M)$ and $\psi_H(v) = \psi_G(v) < \psi_B(v)$ for $v \in [v^M, \omega]$.

If distributions are equal for both workers, then the allocation probability is $G(v)$. Note that the allocation probability for A is given by

$$x(v_A) = P(\psi(v_A) > \psi(v_B)) = P(v_A > v_B) = G(v_A) \tag{50}$$

Dropping subscripts and noting that distributions are equal yields $x(v) = G(v)$.

Having calculated the virtual values and allocation probabilities, we can calculate total effort and show that

$$TE\,(F_A(v), F_B(v)) \geq TE\,(G(v), G(v)) \tag{51}$$

If there is no exclusion, it suffices to show that

$$\int_{v^M}^{\omega} \psi_A(v)\frac{1}{2}H(v)h(v)dv + \int_{\alpha}^{v^M} \psi_B(v)\frac{1}{2}H(v)h(v)dv > \int_{\alpha}^{\omega} \psi_H(v)\frac{1}{2}H(v)h(v)dv \tag{52}$$

as

$$\int_{v^M}^{\omega} \psi_A(v)x_A(v)h(v)dv + \int_{\alpha}^{v^M} \psi_B(v)x_B(v)h(v)dv \geq \int_{v^M}^{\omega} \psi_A(v)\frac{1}{2}H(v)h(v)dv + \int_{\alpha}^{v^M} \psi_B(v)\frac{1}{2}H(v)h(v)dv \tag{53}$$

The optimal allocation rule for the promotion must yield a weakly higher total effort than the allocation rule that is optimal if both workers have the same distribution.

Suppose first, that there is no exclusion. Then,

$$\begin{aligned}
& \int_{\alpha}^{\omega} \psi_H(v)H(v)h(v)dv < \int_{v^M}^{\omega} \psi_H(v)H(v)h(v)dv + \int_{\alpha}^{v^M} \psi_B(v)H(v)h(v)dv && (54)\\
\Leftrightarrow \quad & \int_{\alpha}^{v^M} \psi_H(v)H(v)h(v)dv < \int_{\alpha}^{v^M} \psi_B(v)H(v)h(v)dv && (55)\\
\Leftrightarrow \quad & \int_{\alpha}^{v^M} \left(v - \frac{2-H(v)}{h(v)}\right) H(v)h(v)dv < \int_{\alpha}^{v^M} \left(v - \frac{1-H(v)}{h(v)}\right) H(v)h(v)dv && (56)\\
\Leftrightarrow \quad & -\int_{\alpha}^{\omega} \left(2-H(v)\right) h(v)dv < -\int_{\alpha}^{v^M} \left(1-H(v)\right) H(v)dv && (57)\\
\Leftrightarrow \quad & \int_{\alpha}^{v^M} \left(2-H(v)\right) H(v)dv > \int_{\alpha}^{v^M} \left(1-H(v)\right) H(v)dv && (58)\\
\Leftrightarrow \quad & \int_{\alpha}^{v^M} H(v)dv > 0 && (59)
\end{aligned}$$

which always holds.

Suppose next there is exclusion. Then, adjusting the distribution to F_A and F_B yields again a strictly higher total effort, compared to both having distribution G. Let the cut off if both workers have the same distribution be denoted by $\hat{v}$ and suppose that $\hat{v} < v^M$. Then, the comparison of total effort with equal distributions versus distributions F_A and F_B becomes

$$\int_{\hat{v}}^{v^M} \psi_H(v)H(v)h(v)dv < \int_{\hat{v}}^{v^M} \psi_B(v)H(v)h(v)dv, \tag{60}$$

By the same logic as above, this inequality is always fulfilled. Note that the cutoff for B in the case of distinct distributions is at a lower value than $\hat{v}$, as $\phi_B(v) > \phi_G(v)$ for all v, thus increasing total effort under distinct distributions further.

If $\hat{v} > v^M$, then the comparison boils down to

$$\int_{\hat{v}}^{\omega} \psi_H(v)H(v)h(v)dv = \int_{\hat{v}}^{\omega} \psi_A(v)H(v)h(v)dv. \tag{61}$$

However, we know that the optimal allocation rule with the distinct distribution assigns the promotion with some probability to B if B's valuation is sufficiently close to v^M (as his virtual value at v^M equals v^M). This establishes that also with exclusion adjusting the distribution strictly increases total effort. ■

References

Akerlof, G. A. (1985). Discriminatory, status-based wages among tradition-oriented, stochastically trading coconut producers. *Journal of Political Economy 93*(2), 265–276.

Arrow, K. (1973). The theory of discrimination. In O. Ashenfelter and A. Rees (Eds.), *Discrimination in Labor Markets*, pp. 3–33. Princeton University Press.

Babcock, L., M. P. Recalde, L. Vesterlund, and L. Weingart (2017, March). Gender differences in accepting and receiving requests for tasks with low promotability. *American Economic Review 107*(3), 714–47.

Becker, G. S. (1957). *The Theory of Discrimination.* University of Chicago Press.

Bertrand, M., C. Goldin, and L. F. Katz (2010). Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors. *American Economic Journal: Applied Economics*, 228–255.

Bobkova, N. (2019). Information choice in auctions. Technical report, Working paper.

Bohren, J. A., A. Imas, and M. Rosenberg (2019). The dynamics of discrimination: Theory and evidence. *American economic review 109*(10), 3395–3436.

Bronson, M. A. and P. S. Thoursie (2019). The wage growth and within-firm mobility of men and women: New evidence and theory. Technical report, Working Paper.

Calsamiglia, C., J. Franke, and P. Rey-Biel (2013). The incentive effects of affirmative action in a real-effort tournament. *Journal of Public Economics 98*, 15–31.

Chowdhury, S. M. and O. Gürtler (2015). Sabotage in contests: a survey. *Public Choice 164*(1-2), 135–155.

Condorelli, D. and B. Szentes (2020). Information design in the holdup problem. *Journal of Political Economy 128*(2), 681–709.

Cotter, D. A., J. M. Hermsen, S. Ovadia, and R. Vanneman (2001). The glass ceiling effect. *Social forces 80*(2), 655–681.

Dohmen, T. and A. Falk (2011, April). Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender. *American Economic Review 101*(2), 556–90.

Erosa, Fuster, Kambourov, and Rogerson (2017). Hours, occupations, and gender differences in labor market outcomes. *NBER Working Paper 23636.*

Fang, H. and A. Moro (2011). Theories of statistical discrimination and affirmative action: A survey. In *Handbook of social economics*, Volume 1, pp. 133–200. Elsevier.

Ferrary, M. (2017). Gender diversity in the banking industry- an international comparison. Technical report, Skema Business School, https://www.skema.edu/research/Documents/chairs/women-and-business/gender-diversity-banking-industry-2017-english.pdf.

Fershtman, D. and A. Pavan (2020). Soft affirmative action and minority recruitment. *arXiv preprint arXiv:2004.14953*.

Franke, J. (2012). Affirmative action in contest games. *European Journal of Political Economy 28*(1), 105–118.

Gayle, G.-L. and L. Golan (2011, 09). Estimating a Dynamic Adverse-Selection Model: Labour-Force Experience and the Changing Gender Earnings Gap 1968–1997. *The Review of Economic Studies 79*(1), 227–267.

Gneezy, U., M. Niederle, and A. Rustichini (2003). Performance in competitive environments: Gender differences. *Quarterly Journal of Economics 118*(3), 1049–1074.

Gu, J. and P. Norman (2020). A search model of statistical discrimination. *Available at SSRN 3575199*.

Hartline, J. D. (2013). Mechanism design and approximation. *Book draft. October 122.*

Kamphorst, J. J. and O. H. Swank (2016). Don't demotivate, discriminate. *American Economic Journal: Microeconomics 8*(1), 140–65.

Lang, K. (1986). A language theory of discrimination. *The quarterly journal of economics 101*(2), 363–382.

Lazear, E. P. and S. Rosen (1990). Male-female wage differentials in job ladders. *Journal of Labor Economics 8*(1, Part 2), S106–S123.

Li, S. and J. Yu (2012). Contests with endogenous discrimination. *Economics Letters 117*(3), 834–836.

Lundberg, S. J. and R. Startz (1983). Private discrimination and social intervention in competitive labor market. *The American Economic Review 73*(3), 340–347.

Mealem, Y. and S. Nitzan (2016). Discrimination in contests: a survey. *Review of Economic Design 20*(2), 145–172.

Merluzzi, J. and S. D. Dobrev (2015). Unequal on top: Gender profiling and the income gap among high earner male and female professionals. *Social science research 53*, 45–58.

Milgrom, P. and S. Oster (1987). Job discrimination, market forces, and the invisibility hypothesis. *The Quarterly Journal of Economics 102*(3), 453–476.

Moro, A. and P. Norman (2003). Affirmative action in a competitive economy. *Journal of Public Economics 87*(3-4), 567–594.

Myerson, R. B. (1981). Optimal auction design. *Mathematics of operations research 6*(1), 58–73.

Niederle, M. and L. Vesterlund (2007). Do Women Shy Away From Competition? Do Men Compete Too Much? *The Quarterly Journal of Economics 122*(3), 1067–1101.

Peski, M. and B. Szentes (2013). Spontaneous discrimination. *American Economic Review 103*(6), 2412–36.

Phelps, E. S. (1972). The statistical theory of racism and sexism. *The american economic review 62*(4), 659–661.

Roesler, A.-K. and B. Szentes (2017). Buyer-optimal learning and monopoly pricing. *American Economic Review 107*(7), 2072–80.

Winter, E. (2004). Incentives and discrimination. *American Economic Review 94*(3), 764–773.

School of Economics and Finance

School of Economics and Finance Queen Mary University of London
Mile End Road
London E1 4NS
Tel: +44 (0)20 7882 7356
Fax: +44 (0)20 8983 3580
Web: www.econ.qmul.ac.uk/research/workingpapers/