Discrimination in Promotion*

Anja Prummer[†] Francesco Nava[‡]

January 29, 2021

Abstract

Does an employer benefit from inducing differential value distributions for a promotion among his workers? Workers compete by exerting effort and higher effort corresponds to higher profit for the employer. Introducing inequalities in valuations makes workers' values more easily recognisable, reducing their information rent, which in turn increases effort. At the same time, inequalities lead to differences in promotion attainment, even if realised values are identical. This corresponds to a decrease in competition. We show that if value is re-distributed, the reduction in information rent outweighs the loss in competition, making discrimination between workers optimal.

Keywords: Discrimination, Mechanism Design, Information Design, Culture

*We thank Arjada Bardhi, Chris Bidner, David McAdams, Balazs Szentes for excellent comments. We are grateful to seminar participants at the Berlin Mechanism Design Reading Group, Ca Foscari Venice, Duke, Herion-Watt Edinburgh, QMUL, Simon Fraser, University of Antwerp, UNC, University of Texas at Austin and the 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

[‡]Contact: Department of Economics, London School of Economics and Political Science, Houghton Street, London, WC2A 2AE, UK,f.nava@lse.ac.uk

Work culture tells us what to wear, how to talk, what to talk about-in short, how to interact[...]. And yet, traditional employment anti-discrimination efforts have largely ignored the role that work culture plays in perpetuating workplace discrimination [...].

This is a mistake.

Green (2005) p. 625

1 Introduction

The design of corporate culture as a key tool to foster worker's productivity has been identified in the business literature in the early 1980s (Ouchi and Cuchi (1981), Deal and Kennedy (1982), Peters and Waterman (1982)). It has since led employers to become increasingly concerned with a worker's *fit*. Employers determine features of the job (hours worked, dress codes) and additionally, create expectations around payoff-irrelevant characteristics, such as hobbies or language (dialects or accents).[1] Such expectation are especially important at the management level, where a worker may be seen as a representative of the firm, and in occupations that emphasise social relations, such as consulting, law and finance. At the same time, the management level in these sectors has remained fairly homogeneous, raising questions as to how employers define fit. Do employers prefer a culture that makes rising through the ranks accessible and thus appealing for all workers? Or do they instead choose to define fit more narrowly, giving some workers an advantage over others?

Our findings point towards the latter: firms prefer to induce distinct values for the promotion, which leads to differences in effort and ultimately a disparity in promotion attainment.[2] In our model, workers compete for a promotion by exerting effort. Their effort choices depend on their 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: this could be gender, race, background or simply name. The employer maximises the sum of workers' effort, their total effort, by designing an optimal mechanism and additionally, by influencing workers' value distributions for the promotion. He can influence these distributions by creating a work environment, an organisational culture that is a better fit for one worker, but ignores or disadvantages the other. His goal is to narrow distributions, making workers' values more recognisable. A recognisable value obviates incentivising the worker to reveal his value truthfully and thus the information rent vanishes. At the same time, such a value adjustment may increase one worker's prob-

[1] Examples of "fit" or rather of individuals not fitting in are women in an old boy's club; African Americans, who speak African American English, in a predominantly white company; black hairstyles that are ruled out in the companies dress code; a Northern accent in a London office/ a Southern accent in a New York office.

[2] In our setting, designing a worker's value distribution maps into designing a cost distribution.

ability of promotion, while disadvantaging the other, even if realised values are identical. These disparities in promotion attainment correspond to a decrease in competition. We show that as long as value is re-distributed the gain in information rent outweighs the loss in competition, increasing total effort. Even though employers do not display differences in beliefs or preferences, they still treat or reward workers distinctly, as it increases effort: discrimination is profitable.

Our main contribution is to allow the employer to design the value distribution on top of the optimal mechanism, so as to maximise total effort. Total effort is defined as the sum of effort across workers, which we consider as a proxy for the firm's profit, a natural assumption in the context of finance, consulting and law– all known for their competitive entry requirements, long hours and work hard mentality.

While it seems clear that firms impact worker's valuations, it is less obvious *how* workers' valuations change as a response to culture. Therefore, to allow for generality, we consider any value distribution and assume that the firm can induce it. We first analyse the unconstrained maximisation problem as a benchmark, which highlights the incentives of the employer to influence distributions. The employer aims to reduce the information rent. This makes it optimal to assign all mass to one value. The employer then knows the worker's value for the promotion exactly, leading to an information rent of zero. Given that higher values result in higher effort, the employer would like to set the value as high as possible. However, some values may be prohibitively costly, which leads us to impose an upper bound on the value for the promotion, denoted by $\overline{\omega}$. It is then optimal to assign mass one to $\overline{\omega}$ for at least one worker.

Creating such a high valuation for all workers, however, may not be feasible for the employer, as it may impose certain costs. Supposing that the employer has limited possibilities to increase all workers' valuations, we ask whether creating a culture that affects workers differentially can be beneficial. It turns out, that this is indeed the case, which motivates the two classes of constraints we consider: *value dispersion* and *value reallocation.*

Value dispersion focuses on value redistribution for a given worker, ignoring the other worker. This is in line with a culture, where some workers naturally fit in, while others receive pronounced attention to their payoff-irrelevant characteristics. The firm ignores the former, but tailors their culture design to a specific group.[3]

We first consider the constraint where the employer does not increase the average value for the promotion of the worker, but is otherwise unrestricted in his design. We restrict attention to this setting as we have already established that a costless increase

[3] An example could be double-standards for different demographic groups, dress codes that focus on the appearance of women or ethnic minorities.

in the expected value for the promotion will be beneficial for the employer. Eventually, such an increase in average value will no longer be costless, but the employer may still benefit from a distribution adjustment. Therefore, both workers start out with the same distribution–either because the employer has not yet influenced it or because it has been influenced in the same manner. We refer to this distribution as the initial distribution.

We show that it is optimal to adjust the initial distribution with the effort maximising distribution being binary, with mass at $\overline{\omega}$ and zero, such that the mean of the distribution equals that of the initial distribution. 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 creates 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 worker. Workers exert higher effort the greater their probability of attaining the promotion. Facing a co-worker with a high value with a high probability 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 the workers, if the firm can adjust the distribution for both of them – a result that may emerge if workers are affected in the same manner by the culture they face. Such an adjustment yields the same total effort as if the employer adjusted the distribution for one worker to the described, binary distribution, while using the other 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 the worker with the binary distribution turns out to have zero valuation, the promotion is then given to the worker, whose valuation equals the mean, with certainty. Such an adjustment may take place if the employer can, through corporate culture, impose different expectations for distinct workers.

Note that the constraint of non-increasing average values 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 the employer will refrain from any adjustments. 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 employer only has an incentive to adjust values if it can make values more extreme. This implies that employers favour cultures that induce an exceedingly high valuation by a few, who become very dedicated to the job, while it alienates others. This is typical of the culture observed in law, finance, or consulting, in which only few thrive.

In contrast to the constraints covered by value dispersion, value reallocation allows for redistribution of value *across* workers. Organisational culture may make one worker more comfortable in a given environment, to the detriment of the other. Such an environment can arise if employers define fit around worker's labels and they reward those that fit in, while punishing those who do not.

We assume that if the employer increases the valuation for one worker, he must decrease the valuation for the other worker by the same amount, a natural benchmark. This is equivalent to an employer who is constrained to keep the overall density of each value constant. The density of a worker's original value distribution is denoted by $g(v)$, with v the value of the promotion and it is identical across workers. Therefore, the overall density for two workers at valuation v is given by $2g(v)$ and this needs to be matched by the sum of the densities of the new distributions. As we are allowing for any original distribution, this boils down to matching an arbitrary measure with mass two, which we denote by $H(v)$. For two workers, it turns out to be optimal to create one worker whose values lie below the median of $H(v)$ and another one, whose values are above the median of $H(v)$, that is to create maximally distinct value distributions. By creating these distinct distributions, the employer can again extract higher effort due to a reduced information rent. The employer knows 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 reduction in information rent also offsets the reduced competition between workers. As the reduction in information rent always leads to a greater increase in total effort, it is then optimal to select a culture that induces the greatest distinction between workers.

This result highlights that employers have no incentive to treat their workers equally, but rather benefit from providing advantages to one worker, to the detriment of the other.

Thus, another contribution is to introduce a novel and distinct notion of discrimination, based on corporate culture which prescribes fit. In our model, differences do not reflect an employer's preferences or beliefs– they are distinct from taste-based discrimination (Becker (1957)) as well as statistical discrimination (Phelps (1972), Arrow (1973)). Rather, workers are treated distinctly, because it increases profits.

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

Information Design We relate to the literature on information design in mechanisms. Condorelli and Szentes (2020), Roesler and Szentes (2017), and Bobkova (2019) focus on a buyer and seller relationship.[4] In these settings, the buyer (our worker) can make a costly investment or an information choice, which the seller (our employer) takes as given. In contrast, we focus on the choice of the *employer* to influence the valuation of several workers, taking into account the effect on competition, subject to a *constraint*. In the context of selling a good, this can be interpreted as designing the valuation of the good through advertisement or more directly, the design of demand functions. The latter is explicitly considered by Johnson and Myatt (2006), when selling occurs through a market. While we focus on a work environment, our results are more broadly applicable: a seller who wishes to sell a good to buyers through an optimal mechanism and can influence the buyers' value distributions through advertising, will choose to emphasise features of the good that lead to the valuation becoming more easily recognisable.

Bergemann and Pesendorfer (2007) and Sorokin and Winter (2018) take a Bayesian Persuasion angle on revealing information to buyers about their valuations, subject to Bayes' plausibility. In contrast, we allow for the employer to adjust the value distributions directly and consider a larger class of constraints.

Inequality in Contests We contribute to the literature on inequalities in contests. Lazear and Rosen (1981) show that it is optimal to handicap a more able contestant. Mealem and Nitzan (2016) provide a survey of discrimination in contests, focusing on features of contest success functions, 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.

Specifically, Calsamiglia, Franke, and Rey-Biel (2013) show in an experiment that inequalities between contestants are harmful, while affirmative action is optimal (see also Franke (2012)). Li and Yu (2012) highlight that in a contest with two unequal participants, the designer makes them equal in order to obtain the highest payoff, contrary to our finding.

Organisational Incentives and Culture Our concept of discrimination relates to Winter (2004)'s discrimination among workers. There, discrimination in payment schedules emerges in order to induce workers, whose effort is complementary, to exert effort in a team moral hazard problem. In contrast, we focus on inequalities in a setting with substitutable effort induced through organisational culture. Kreps (1990) first acknowledged its importance and the desirability to incorporate corporate culture into economic theory,

[4] Relatedly, Ali, Haghpanah, Lin, and Siegel (2020) look at designing a distribution when selling through an intermediary.

which has spawned a small and distinguished literature, for an overview see Hermalin (1999, 2012). The focus of this literature is how culture ensures efficient operation.[5] In contrast, we focus on the effects of culture on worker's valuations for a promotion: workers, who fit in more easily, possess a higher valuation for the promotion.

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 constraint in Section 3. We provide a benchmark result for the optimal distribution without constraints in Section 3.1. We analyse the case of value 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. Section 4 provides a discussion and concludes.

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 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 $V_i = [\alpha_i, \omega_i]$. While the realisation is private, the distribution is common knowledge. 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 ex-post amounts to

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

[5] See Crémer (1993), Lazear (1995), Hodgson (1996) and more recently Martinez, Beaulieu, Gibbons, Pronovost, and Wang (2015), Gibbons and Kaplan (2015), Gibbons and Prusak (2020).

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. The solution concept is that of Bayesian Nash equilibrium. 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,

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

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, F_A and F_B, subject to some constraints, which 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

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

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

Value Dispersion Value dispersion focuses on the adjustment of values for a given worker, that is values can only be influenced *within* workers. As we analyse a benchmark case where the employer can adjust the distributions without any constraint, a natural second step is to assume that the employer cannot increase the average value for the promotion.

Formally,

$$\mathbb{E}_{F_i}[v] \leq \mathbb{E}_G[v], \tag{7}$$

where we index the expectation operator with the distribution v follows. Our main focus is on the case, where the employer can only adjust the distribution of one worker, keeping that of the other worker fixed. However, we also analyse the setting where the employer can adjust the distribution of both workers, without increasing the average promotion value. Constraint (7) encompasses the set of distributions that are second order stochastic dominated by the worker's initial distribution. A further obvious constraint is that of first order stochastic dominance,

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

Value Reallocation We also consider a setting in which value can be reallocated across workers, through work cultures that lead to an increase in valuation of one worker, while decreasing the valuation of the other worker. A natural constraint is for the employer to match an arbitrary distribution $H(v)$ which has mass two,

$$F_A(v) + F_B(v) = 2G(v) = H(v) \text{ for all } v \in [0, \overline{\omega}]. \tag{9}$$

Discussion of Modelling Assumptions We assume that both workers start out with the same distribution–either because the employer has not yet influenced it or because it has been influenced in the same manner. This serves as a benchmark to highlight that even if workers start out with the same distribution the employer has an incentive to adjust them. We first allow the employer to design valuations freely in our benchmark case. However, it is natural to assume that influencing workers' valuations comes at a cost. These costs are captured by the constraints imposed by value dispersion and reallocation.[6]

Value dispersion captures an adjustment in culture, which focuses on changes in valuations within a worker, ignoring the other. This corresponds to a culture that focusses on the fit of one worker more than the other. It may prescribe a certain code of conduct or dress code that only affects one of the workers. Alternatively, the code of conduct may affect workers differentially. We aim to capture the latter by allowing for changes in values across workers. We will also briefly discuss a constraint that reallocates *expected* value across workers. The solution to this problem is straightforward, and so such a constraint did not seem to warrant a detailed discussion.

Given how arbitrary work place culture is, with varying dress codes, appropriate ways

[6] It is straightforward to map the constraints we consider to cost functions.

of speaking and interacting, which are also country-specific, our modelling approach allows for a general design of valuations, with natural and unrestrictive constraints that encompass first order and second order stochastic dominance. It turns out that the value distributions that emerge capture features of the corporate culture well, validating our approach.

If the value for promotion is zero, then the worker exerts zero effort. This is a normalisation. We view effort as additional effort on top of a benchmark effort needed to retain the job.

Further, we assume that workers' effort is observable. Effort can be replaced by a measurable output, such as billable hours, projects or profits generated that is related to effort. It may be the case that effort does not map linearly into some measurable output. However, even if costs of exerting effort are convex, our approach remains valid (Greenwald, Oyakawa, and Syrgkanis (2017)).

3 Constrained Discrimination

We discuss in turn the optimal value distribution for the employer if he can disperse or reallocate value, after highlighting the optimal distribution if the employer did not face any constraints. All our proofs are collected in the Appendix.

3.1 Designing Value without Constraints

Suppose the employer maximises (6) with a bounded support, $v \in [0, \overline{\omega}]$ the only constraint. If so, the employer designs 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. Formally, distribution*

$$F^{NC}(v) = \begin{cases} 0 & \text{if } v < \overline{\omega} \\ 1 & \text{if } v = \overline{\omega} \end{cases}, \tag{10}$$

assigned to at least one worker, maximises total effort.

This result highlights two forces at play in our model that emerge in all our 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* in the quantile space amounts to $v_i(q) - \phi_i(q) = v_i'(q)q$. If the distribution consists of a single atom, the value does not change across the quantile space and therefore the information rent is reduced to zero. Therefore, creating a more narrow distribution and 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. Thus, 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 one worker obtains the promotion probability $p \in [0, 1]$, while the other receives it with probability $1 - p$. This reduces the effort of the worker, who previously obtained the promotion with certainty, while simultaneously increasing the effort of the worker, who did not receive the promotion, by the *same amount*. Thus, in expectation total effort is the same, independently of whether the employer increases the value for one or both workers.

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 We keep the distribution of worker A fixed and focus on the adjustment of the value distribution of worker B, assuming that the workplace culture affects only one worker.

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.[7] This is formalised in Proposition 2, which characterises the optimal distribution.

Proposition 2. *Total effort is maximised among all distributions F with $\mathbb{E}_F[v] \leq \mathbb{E}_G[v]$*

[7]This result resembles the perfect information case in information design, where the posterior jumps to either zero or one. Similarly, the constant mean translates to Bayes Plausibility, see Kamenica and Gentzkow (2011).

by

$$F^*(v) = \begin{cases} 1 - \frac{\mathbb{E}_G[v]}{\overline{\omega}} & \textit{if} \;\; v < \overline{\omega} \\ 1 & \textit{if} \;\; v = \overline{\omega} \end{cases} \tag{11}$$

To see that the expected value remains constant, note that

$$\mathbb{E}_{F^*}[v] = \frac{\mathbb{E}_G\,[v]}{\overline{\omega}}\overline{\omega} + \left(1 - \frac{\mathbb{E}_G\,[v]}{\overline{\omega}})\right) 0 = \mathbb{E}_G[v] \tag{12}$$

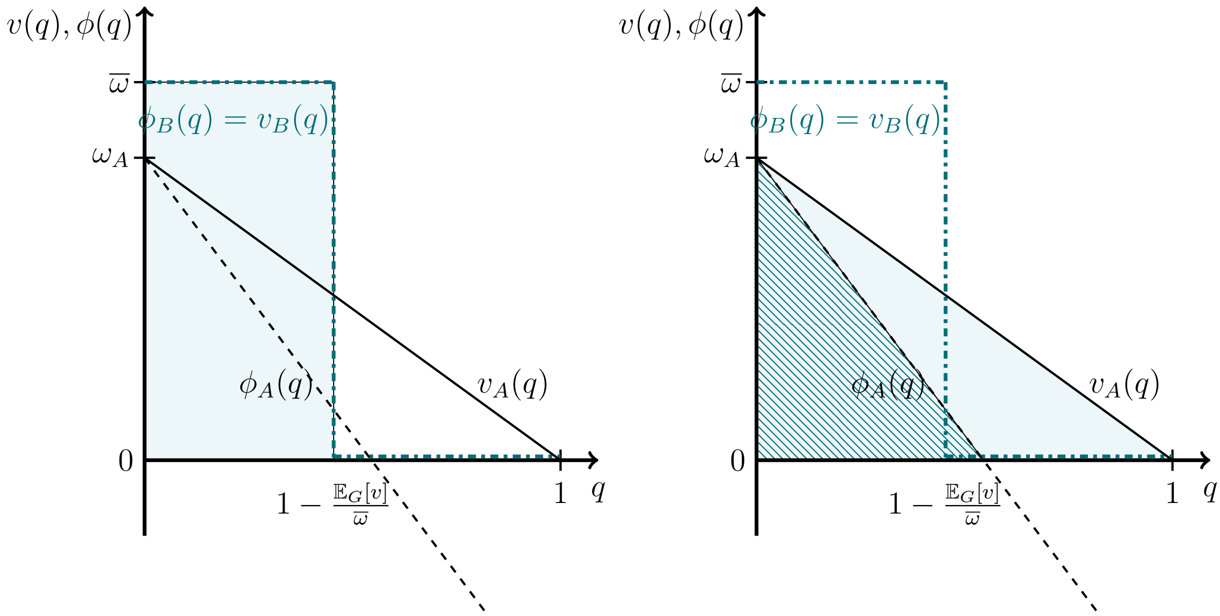
The employer chooses a maximal spread of valuations for one agent. As we already demonstrated in the benchmark case, the firm would like to create as narrow a distribution as possible as this makes the worker's valuation easily identifiable. A worker's valuation is only identifiable without the firm having to pay information rent if there is mass at a single positive value. This implies that there can be mass one at the original expected value or alternatively, mass allocated to a single positive value and zero so as to match the average value of the initial distribution. At zero the firm weakly prefers to not allocate the promotion and therefore does not incentivise the worker to reveal his type. As for each of these distributions he can extract the entire valuation, the employer selects the highest possible average valuation, the expected value of the initial distribution, $\mathbb{E}_G[v]$. While the effort associated with each of the adjustments is the same, 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 $v < \overline{\omega}$. 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, a higher value implies a lower probability that the value is positive and not zero. Therefore, 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 Proposition 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 $v_A(q)$ (and which is the same size as the shaded area in 1a).

It is worth noting that Proposition 2 encompasses second order stochastic dominance, where B's distribution is second order stochastically dominated if and only if the following

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

holds:

$$\int_0^v G(t)dt \leq \int_0^v F_B(t)dt \quad \text{for all } v \in [0, \overline{\omega}]. \tag{13}$$

As F^* is second order stochastically dominated by G, it is also the optimal distribution among all F_B to which definition (13) applies.

Corollary 2.1. *Among all distributions that are second order stochastically dominated by G, F^* maximises 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 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 Corollary 2.2.

Corollary 2.2. *For any* $m \geq \mathbb{E}_G[\max\{\psi(v), 0\}]$ *the distribution*

$$F^+(v) = \begin{cases} 1 - \frac{m}{\overline{\omega}} & \textit{if} \;\; v < \overline{\omega} \\ 1 & \textit{if} \;\; v = \overline{\omega} \end{cases} \tag{14}$$

yields higher total effort compared to no adjustment.

The condition implies that spreading value is optimal for the employer as long as the reduction in expected value, $\mathbb{E}_G[v] - m$, is more than compensated by the reduction in information rent, $\mathbb{E}_G[v] - \mathbb{E}_G[\max\{\psi(v), 0\}]$. Thus, even if the firm loses some value for the promotion on average by adjusting the distribution, doing so still increases the total effort for the firm.

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 to the distribution described in Proposition 2. For the second worker, there are then two optimal adjustments, as summarised in Corollary 2.3.

Corollary 2.3. *Assume* $\mathbb{E}_{F_i}[v] \leq \mathbb{E}_G[v]$, $\forall i \in \{A, B\}$. *Total effort is maximised by setting for worker* i, $F_i^*(v) = F^*(v)$ *and for worker* j *either (i)* $F_j^*(v) = F^*(v)$ *or (ii)*

$$F_j^*(v) = \begin{cases} 0 & \textit{if} \;\; v < \mathbb{E}_G[v] \\ 1 & \textit{if} \;\; v \geq \mathbb{E}_G[v] \end{cases} \tag{15}$$

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 do not strive 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. Such an environment comes at the cost of not having a worker with positive valuation with a high probability and thus no one 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 Proposition 2. With many workers, the firm creates an environment where all workers (potentially with the exception of one) are overachievers with a small probability.[8] 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 distribution, $G(v) = F_A(v_A) \leq F_B(v_B)$. First order stochastic dominance implies a ranking of quantiles $q_A(v)$ and $q_B(v)$:

$$q_B(v) = 1 - F_B(v) \leq 1 - F_A(v) = q_A(v) \tag{16}$$

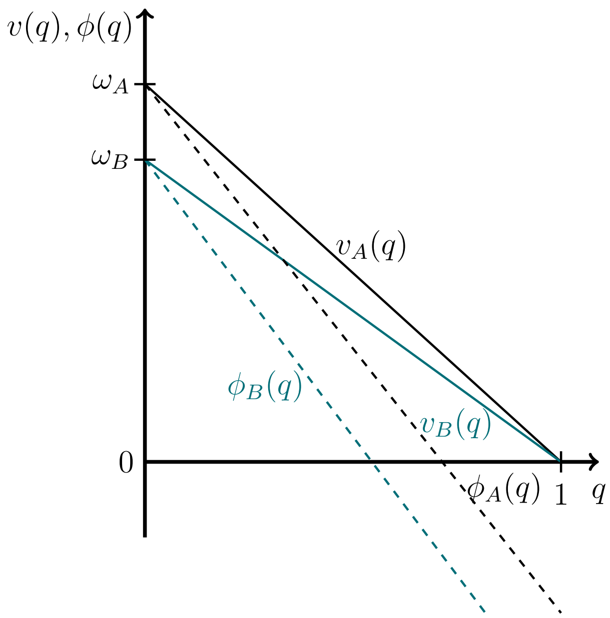
It follows that for a given quantile, the value for A is higher than that for B, formally $v_A(q) \geq v_B(q)$ for any $q \in [0, 1]$. This observation is summarised in Figure 2. For any distribution B that is first order stochastically dominated by a distribution A, the value of A is higher than that of B for every quantile. This allows us to work directly with the values, instead of virtual values through integration by parts. This observation turns out to be instrumental in proving Proposition 3.

Proposition 3. *Total effort is maximised by having no discrimination and setting* $F_B^*(v) = F_A(v) = G(v)$ *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 he 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 dramatically. In this case, the gain in information rent never outweighs the loss in value.

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

Figure 2: A FOSD B



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 is first order stochastically dominated by $F_A(v)$, but first order stochastically dominates $F_B(v)$, again with strict inequality 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 fix the allocation rule to the one that is optimal if the distribution is $F_B(v)$ and show that the optimal total effort increases due to the adjustment of virtual values from $F_B(v)$ to $\hat{F}_B(v)$. Formally, we show the second inequality in expression (17).

$$\underbrace{\mathbb{E}[\hat{\phi}_B(q)\hat{y}_B(q)]}_{\text{Distribution:}\hat{F}_B,\ \text{Allocation:}\hat{F}_B} \geq \underbrace{\mathbb{E}[\hat{\phi}_B(q)y_B(q)]}_{\text{Distribution:}\hat{F}_B,\ \text{Allocation:}F_B} > \underbrace{\mathbb{E}[\phi_B(q)y_B(q)]}_{\text{Distribution:}F_B,\ \text{Allocation:}F_B} \tag{17}$$

The first inequality holds as total effort under the old allocation rule, associated with $F_B(v)$, and new virtual value derived from $\hat{F}_B(v)$ is a lower bound on total effort, obtained from the virtual value and allocation rule associated with $\hat{F}_B(v)$. This follows as an adjustment of the allocation rule must increase effort – otherwise the firm would not select a new allocation rule. We then turn to a comparison of virtual values, to show that $\mathbb{E}\left[\left(\hat{\phi}_B(q) - \phi_B(q)\right) y_B(q)\right] > 0$. Integrating by parts leads to a comparison of values as $\phi(q) = \frac{\partial(v(q)q)}{\partial q}$. We know that distribution $\hat{F}_B(v)$, is associated with a higher value for each quantile compared to distribution $F_B(v)$, as $\hat{F}_B(v)$ first order stochastically dominates

$F_B(v)$, see also Figure 2. 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.

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. This is in line with the environment we see in consulting, law and finance: some workers value a promotion to full professor, partner, manager a lot and are willing to work incredibly hard for it. Others dislike this culture and opt out of the race, doing the minimum to keep their job. Even if such a culture comes at a cost, it can still be beneficial to implement it, as long as the cost is not too high.

3.3 Value Reallocation

We turn to a setting where the employer designs cultures that cater to one worker, while disadvantaging the other. A natural assumption is that an increase in valuations for one worker is compensated by a decrease in valuations for the other worker, that is value is reallocated. This is equivalent to keeping the overall density of each value constant or alternatively, $F_A(v) + F_B(v) = 2G(v) = H(v)$, where $H(v)$ is an arbitrary measure with mass *two*. If workers' distributions are identical, they are again denoted by $G(v)$. However, it is not optimal to create identical distributions. Rather, the employer maximises total effort by creating two maximally different value distributions. This is formalised in Proposition 4, for which it is useful to define v^M as the median of $H(v)$, $\int_\alpha^{v^M} dH(v) = 1$.

Proposition 4. *If $F_A(v) + F_B(v) = H(v)$, total effort is maximised by choosing distributions*

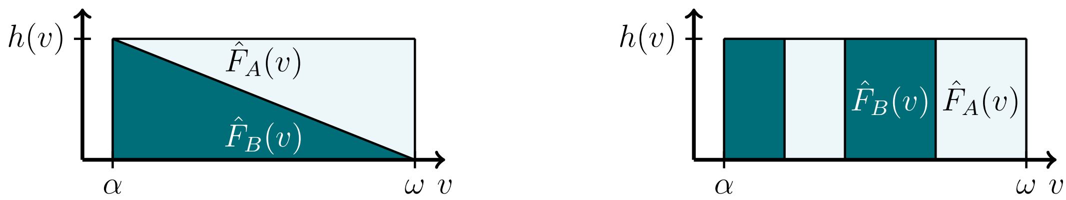
$$F_A^*(v) = H(v) - 1 \qquad \text{if } v \in [v^M, \omega], \tag{18}$$

$$F_B^*(v) = H(v) \qquad \text{if } v \in [\alpha, v^M). \tag{19}$$

Maximal discrimination, such that worker B's values are below the median of $H(v)$, while worker A's values are above the median of $H(v)$, always yields a higher total effort than any other two distributions. First, this adjustment narrows the distribution, by reducing the values workers can have. Additionally, it increases the probability of receiving the promotion for the worker with high values, inducing him to work harder, while lowering the probability of receiving the promotion for the low value worker. This leads to lower effort from worker B. As it turns out, the gain in information rent again outweighs the loss in competition and maximal discrimination is optimal as it minimises the information rent.

To show the optimality of maximal discrimination, we first compare maximal discrimination to any two distributions that split the densities $h(v)$ for all v before turning to disjoint supports, which also encompass mass points. These comparisons are illustrated in Figure 3, with an example of split densities illustrated in Figure 3a. The example splits the densities in a smooth manner. However, in principle, each density can be split in an arbitrary manner. Further, we require maximal discrimination to lead to a higher total effort compared to any other distributions with disjoint support, see Figure 3b. It suffices to show that these two classes of distributions do not lead to a higher total effort compared to maximal discrimination, as any combination between split densities and disjoint support leads to a linear combination of the effort of split densities and disjoint supports– and we show that total effort is lower for each of these cases.

Figure 3: Alternative Distributions



(a) Splitting Densities (b) Disjoint Support

We turn to comparing the total effort with maximal discrimination to the total effort with split densities. In order to capture that densities can be split arbitrarily, we assign a share $a_A(v) \in (0,1)$ of the density to $\hat{F}_A$, while the remainder, $a_B(v) = 1 - a_B(v)$ is allocated to $\hat{F}_B$. In order to compare total effort, we again need to keep track of changes in virtual values and allocation probabilities. We therefore impose some structure on the allocation probabilities. Namely, we construct for the distributions $\hat{F}$ auxiliary distributions that rank the allocation probabilities and are equivalent in total effort. More precisely, we define by $\overline{x}_A(v)$ the maximum of the allocation probabilities $\hat{x}_A(v)$, $\hat{x}_B(v)$ if $v \geq v^M$ and equivalently, by $\overline{x}_B(v)$ the maximum of $\hat{x}_A(v)$, $\hat{x}_B(v)$ if $v < v^M$. The corresponding virtual values are denoted by $\overline{\psi}_i$. We then compare

$$\int_{v^M}^{\omega} \psi_A^*(v)\overline{x}_A(v)h(v)dv + \int_{\alpha}^{v^M} \psi_B^*(v)\overline{x}_B(v)h(v)dv$$
$$> \int_{\alpha}^{\omega} \left(a(v)\overline{\psi}_A(v)\overline{x}_A(v) + (1-a(v))\overline{\psi}_B(v)\overline{x}_B(v)\right) h(v)dv. \tag{20}$$

The LHS of (20) presents the lower bound of total effort with maximal discrimination, given by the virtual values of distributions F_A^* and F_B^* combined with the allocation probabilities of the auxiliary distributions.[9] The RHS of (20) equals total effort with split densities. The problem can then be reframed as one of signing the following differences:

$$\psi_i(v)\overline{x}_i(v) - a_A(v)\overline{\psi}_A(v)\overline{x}_A(v) - a_B(v)\overline{\psi}_B(v)\overline{x}_B(v) \tag{21}$$

$$> \left(\psi_i(v) - a_A(v)\overline{\psi}_A(v) - a_B(v))\overline{\psi}_B(v)\right)\overline{x}_i(v), \qquad i \in \{A, B\}, \tag{22}$$

where the inequality follows from the definition of the allocation probabilities. We therefore have simplified the problem to a comparison of virtual values. If the difference in virtual values is positive for A and B, for all v, then we have shown that total effort is higher under maximal discrimination compared to any split. It turns out that this is straightforward and we obtain that the difference in virtual values is zero for values above the median v^M, while it is equal to one for values below the median.

This implies that the gain in virtual valuation occurs in the lower half of the value distribution. In the upper half of the value distribution, the information rent is always low, and therefore the exact distributions covering $H(v)$ are irrelevant. In the bottom half, however, the information rent is high. Recall that the information rent amounts to $\frac{1-F(v)}{f(v)}$, which is high if $F(v)$ and $f(v)$ are low. Therefore maximal discrimination is optimal as it assigns the highest mass possible to low values, where it is particularly valuable to minimise information rent.

Moreover, it is not optimal to divide the distributions into finer intervals, without splitting densities. As with splitting densities, our proof amounts to a comparison of virtual values. In contrast to splitting densities, the employer can now reduce the information rent for values above the median, compared to maximal discrimination: there can always be a distribution with sufficiently high mass up to some value v, such that the information rent is lower at v. This is countered by a loss in information rent at some other, lower values. However, the gain in information rent for higher values cannot make up for the loss at lower values–as at lower values, the information rent is highest.

As was the case for previous constraints, competition is of secondary importance. It is crucial for the employer to minimise the information rent and he achieves this through maximal discrimination. In particular, he saves on information rent for those workers who value the promotion little and thus discrimination disadvantages them not only as they face lower values, but also as for these values the information rent is reduced.

Discrimination remains optimal if the employer instead reallocates expected valuations across workers, $2\mathbb{E}_G[v] = \mathbb{E}_{F_A}[v] + \mathbb{E}_{F_B}[v]$. Under this constraint, the firm allocates all expected value to one worker, such that $\mathbb{E}_{F_A^*}[v] = 2\mathbb{E}_G[v]$, while $\mathbb{E}_{F_B^*}[v] = 0$. The intuition

[9] As in the case of first order stochastic dominance assigning different allocation probabilities weakly lowers total effort. Note that we verify that $\overline{x}_A(v)$ and $\overline{x}_B(v)$ constitute allocation probabilities for distributions F_A^* and F_B^* when we check interim feasibility.

is similar to the benchmark case: the firm only requires one worker with a positive valuation and therefore selects worker A to receive all expected value, while the other worker's expected value is reduced to zero. Consequently, an employer who hires two identical workers, will induce differences in how they value the promotion, through creating the appropriate work place culture.

4 Discussion and Conclusion

Our analysis highlights that an employer can induce higher effort by introducing a culture with a focus on one worker, creating a bi-modal value distribution for him, while ignoring the other worker. Additionally, the employer finds it beneficial to reallocate value such that one worker displays a high value for the promotion, while the other one ends up with a lower valuation.

In practice, we would expect a mix of both features to emerge. This seems to be true for law, banking and management consultancies, where culture is often described to be cut-throat, a toxic environment with extremely long work hours. Such a culture is loved by few and disliked by many and matches the bi-modal value distribution that emerges in the value dispersion case. Additionally in these environments, fit is important. And for some workers, fitting in is more challenging than others according to Padavic, Ely, and Reid (2020): they argue that such an environment disproportionately excludes women, as they face both an expectation to spend time with their family as well as to focus on work, creating a lose-lose situation for them. The latter could be viewed as an example of value reallocation. Our model therefore proposes a novel explanation for why some groups, such as women, fail to raise through the ranks, in line with the empirical evidence.[10]

Our model also sheds light on why minorities and women are not as underrepresented at entry level jobs. In our model, it is optimal to hire workers with the highest value for promotion, independently of their gender or other characteristics. A higher average value results in higher value extraction if the employer is bound by the expected value of the initial distribution. If the employer is constrained by the initial distribution, he prefers workers with distributions with low information rent, a narrow distribution. This translates into screening individuals to select those with the highest expected value and low variation. Such a hiring strategy is arguably followed by law firms, consultancies and in finance, who prefer graduates from top programmes.

Nevertheless, the employer has an incentive to create certain cultures for their entry

[10]Our explanation for the underrepresentation of women in management positions is complementary to those presented in Bertrand, Goldin, and Katz (2010), Gayle and Golan (2011), Gneezy, Niederle, and Rustichini (2003), Niederle and Vesterlund (2007), Dohmen and Falk (2011), Erosa, Fuster, Kambourov, and Rogerson (2017).

level employees– a feature contrary to existing theories of discrimination. If there was a distaste toward certain types of employees, they would not even be hired. Further, there is less information about a worker's ability available at the hiring stage, which should lessen statistical discrimination at the promotion stage compared to the entry level. This pattern indeed emerges in Bohren, Imas, and Rosenberg (2019)'s model of dynamic statistical discrimination based on gender and is confirmed through an experiment: statistical discrimination is less important at later career stages. This is in contrast to empirically observed patterns of gender inequality (Bertrand et al. (2010)): the gender wage gap increases across career time, which is related to workplace culture and women missing out on promotions.[11] Similarly Altonji and Pierret (2001) document that race matters more further down the career path, less so at the entry level. Thus, we propose a novel mechanism of why discrimination arises, which is supported by empirical regularities.

To conclude, we allow the employer to establish the optimal mechanism for the allocation of a promotion and additionally, to design a value distribution for the promotion by implementing a certain workplace culture subject to a constraint. We allow for a large, natural set of constraints, including first and second order stochastic dominance and obtain that introducing inequalities is generally profitable.

We focus throughout on two workers, although it is straightforward to extend the value dispersion to multiple workers. Similarly, we speculate that an extension to multiple workers in the value reallocation setting leads to a split support similar to what we find in the case with two workers as it will still be optimal to minimise the information rent. While we have focused here on independent values, we expect our insight that discrimination is optimal to carry over to a setting with correlated values (Crémer and McLean (1985)). In this case, the information rent is zero even when the employer is faced with a non-degenerate distribution, but he still prefers a worker to have the highest valuation possible- even if this is at the expense of the other worker. Further extensions could allow for (i) multiple promotions, (ii) a sequence of promotions, and (iii) multiple employers that compete among each other. The solutions for these extensions are involved, as evidenced by the expansive literature on auctions. We therefore leave these extensions to future research.

[11] Merluzzi and Dobrev (2015) highlight the importance of culture for women not performing better, while Bronson and Thoursie (2019) relate the gender wage gap to (lack of) promotions.

A Appendix

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

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

where the first inequality follows from $\phi_i(q) \leq \overline{\omega}$ for all $i \in A, B$, and the second inequality from $E\left[y_A(q)\right] + \mathbb{E}\left[y_B(q)\right] \leq 1$ for any allocation rule. Therefore, as long as at least one worker has value $\overline{\omega}$ with certainty, the firm extracts the highest possible total effort. ■

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

$$\phi_B^*(q) = \begin{cases} \overline{\omega} & \text{if } q < \overline{q} \\ 0 & \text{if } q > \overline{q} \end{cases}. \tag{24}$$

Total effort under distribution F^* is given by

$$\mathbb{E}\left[\phi_A(q)y_A^*(q)\right] + \mathbb{E}\left[\phi_B^*(q)y_B^*(q)\right], \tag{25}$$

while the effort under any alternative F_B amounts to

$$\mathbb{E}\left[\phi_A(q)y_A(q)\right] + \mathbb{E}\left[\phi_B(q)y_B(q)\right]. \tag{26}$$

The change from F_B to F^* 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 in the quantile space and show that a change from F_B to F^* increases total effort under the optimal allocation rule for distribution F_B. Such insight then immediately delivers the result, since optimality implies that

$$\mathbb{E}\left[\phi_A(q)y_A^*(q)\right] + \mathbb{E}\left[\phi_B^*(q)y_B^*(q)\right] \geq \mathbb{E}\left[\phi_A(q)y_A(q)\right] + \mathbb{E}\left[\phi_B^*(q)y_B(q)\right]. \tag{27}$$

Note that fixing the allocation rule in the quantile space to $\mathbf{y}(\mathbf{q}) = \mathbf{x}(\mathbf{v}(\mathbf{q}))$ implies that interim allocation rules are also unchanged in the quantile space, since $y_i(q_i) = \int_0^1 \mathbf{y}_i(\mathbf{q})dq_j$, which was implicitly assumed in the previous expression.
Therefore, to prove the result, it suffices to establish that

$$\mathbb{E}\left[\phi_B^*(q)y_B(q)\right] - \mathbb{E}\left[\phi_B(q)y_B(q)\right] \geq 0. \tag{28}$$

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

$$\mathbb{E}\left[\phi^*_B(q) - \max\{\phi_B(q), 0\}\right] \geq 0 \quad \Rightarrow \quad \mathbb{E}\left[(\phi^*_B(q) - \phi_B(q))y_B(q)\right] \geq 0. \tag{29}$$

Moreover, by construction, the expected virtual values satisfy

$$\mathbb{E}\left[\phi^*_B(q)\right] = \mathbb{E}\left[v^*_B(q)\right] = \overline{q\omega} = \mathbb{E}\left[v_A(q)\right] = \mathbb{E}\left[v_B(q)\right]. \tag{30}$$

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

$$\mathbb{E}\left[v_B(q)\right] \geq \mathbb{E}\left[\max\left\{\phi_B(q), 0\right\}\right], \tag{31}$$

given that $\max\left\{\phi_B(q), 0\right\} \leq v_B(q)$ for all $q > 0$ since $v_B(q) \geq 0$ and $\phi_B(q) - v_B(q) = qv_B'(q) \leq 0$.
This establishes that it is never optimal to select a distribution for B with an expected virtual value which is strictly smaller than 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 that it 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 would need to differ across the different valuations for the mechanism 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 with such a distribution it would not be feasible to extract the entire valuation, yielding the contradiction.

To show that F^* is the unique optimum, consider any other distribution F_B having the same mean as G and a single atom on positive values at ω_B, meaning that

$$F_B(v) = \begin{cases} 1 & \text{if} \quad v = \omega_B \\ 1 - \frac{\mathbb{E}_G[v]}{\omega_B} & \text{if} \quad v < \omega_B \end{cases}. \tag{32}$$

Letting $p_B = \mathbb{E}_G[v]/\omega_B$ denote the probability of having value ω_B, total effort can be rewritten as

$$\mathbb{E}[\phi_A(q)y_A(q)] + \mathbb{E}[\phi_B(q)y_B(q)] \tag{33}$$

$$= (1-p_B)\int_{V_A} \max\{\psi_A(v_A), 0\}dF_A(v_A) + p_B\int_{V_A} \max\{\psi_A(v_A), \omega_B\}dF_A(v_A) \tag{34}$$

$$= (1-p_B)\int_{V_A} \max\{\psi_A(v_A), 0\}dF_A(v_A) + \int_{V_A} \max\{p_B\psi_A(v_A), \mathbb{E}_G[v]\}dF_A(v_A). \tag{35}$$

The first equality follows as the optimal mechanism allocates the promotion to A when their virtual valuation is positive and $v_B = 0$, and when their virtual valuation exceeds ω_B if $v_B = \omega_B$.[12]

The last expression is differentiable in p_B. Next we differentiate such expression with respect to p_B and establish that total effort decreases in p_B, meaning that the the optimal distribution will set p_B to be as small as possible, or equivalently ω_B as large as possible. Letting $V_+(k) = \{v_A \in [0, \overline{\omega}] | \psi_A(v_A) \geq k\}$ for $k \geq 0$ denote the set of values for which A's virtual valuation weakly exceeds k, we find that

$$\frac{\partial \mathbb{E}[\phi_A(q)y_A(q) + \phi_B(q)y_B(q)]}{\partial p_B} = -\int_{V_+(0)} \psi_A(v_A)dF_A(v_A) + \int_{V_+(\omega_B)} \psi_A(v_A)dF_A(v_A) \leq 0, \tag{36}$$

where the inequality holds, since $V_+(\omega_B) \subseteq V_+(0)$. Moreover, the inequality is strict whenever $\psi_A(v_A) \in (0, \omega_B)$ for some $v_A \in V_A$. Thus, it is optimal to minimize p_B which is accomplished by setting $\omega_B = \overline{\omega} \geq \omega_A$. This establishes the result. ■

Proof of Corollary 2.1: Second Order Stochastic Dominance We show that F^* is second order stochastically dominated by $F_A = G$. If $F^* = F_A$, then F^* is second order stochastically dominated by F_A trivially. If $F^* \neq F_A$, it must be that for some $\overline{v} \in [\alpha_A, \omega_A]$ and all $v \in [0, \overline{v})$,

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

since F^* places the maximal possible measure on $v = 0$ amongst all distributions with mean equal to $\mathbb{E}_G[v]$. It follows that for any $v \in [0, \overline{v})$, we have that

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

[12] If G was not regular, the previous expression for revenues would still apply. In such scenarios, F_A would denote the ironed distribution of values yielding the same total effort, rather than G itself – see Hartline (2013), Theorem 3.14, p.78.

Moreover at $\overline{\omega}$, Riemann Stieltjes integration by parts yields

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

To establish second order stochastic dominance, it then suffices to show that $\Delta(v)$ is non-decreasing, since that implies $\int_0^v \Delta(t)dt < 0$ for all $v \in [0, \overline{\omega})$. This holds as F_A is non-decreasing and F^* is constant up until $\overline{\omega}$. Therefore, F_A second order stochastically dominates F^*. ■

Proof of Proposition 3: First Order Stochastic Dominance 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 $\hat{F}_B(v) \geq F_A(v)$.

As in the proof for value dispersion, 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 again 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[13]

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

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

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

Denote by $D \subseteq [0,1]$ the set of points at which y_B in non-differentiable, and by $C = [0,1] \setminus D$. The interim allocation rule y_B can be discontinuous, when the distribution F_B is not continuously differentiable and if there are gaps in the support. We first consider the case where only the allocation rule is discontinuous, before turning to gaps in valuations. Integration by parts using Riemann Stieltjes integrals implies that

$$\mathbb{E}[\phi_B(q)y_B(q)] = \alpha_B y_B(1) + \sum_{q \in D} qv_B(q)\left(-J(q)\right) - \int_C qv_B(q)dy_B(q), \tag{42}$$

where $J(q) = y_B^+(q) - y_B^-(q) \equiv \lim_{\epsilon \to 0}[y_B(q+\epsilon) - y_B(q-\epsilon)] < 0$ and the inequality follows

[13] If there are gaps in the support, the following expression requires a slight amendment, which does not affect results, but adds additional notation.

from y_B being decreasing. Exploiting the latter, we find that

$$\mathbb{E}[(\hat{\phi}_B(q) - \phi_B(q))y_B(q)] = \int_0^1 (\hat{\phi}_B(q) - \phi_B(q))y_B(q)dq \tag{43}$$

$$= (\hat{\alpha}_B - \alpha_B)y_B(1) + \sum_{q \in D} q(\hat{v}_B(q) - v_B(q))\,(-J(q)) + \int_C (q\hat{v}_B(q) - qv_B(q))\,(-y'_B(q))\,dq. \tag{44}$$

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

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

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

$$\mathbb{E}[\phi_A(q)\hat{y}_A(q)] + \mathbb{E}[\hat{\phi}_B(q)\hat{y}_B(q)] \geq \mathbb{E}[\phi_A(q)y_A(q)] + \mathbb{E}[\hat{\phi}_B(q)y_B(q)], \tag{46}$$

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

Suppose now $F_B(v)$ has gaps in its support. Then, for each of these discontinuities in D, we need to show that

$$v^-_B(q)y^-_B(q) - v^+_B(q)y^+_B(q) \leq \hat{v}^-_B(q)y^-_B(q) - \hat{v}^+_B(q)y^+_B(q), \tag{47}$$

where $v^-_B(q)$ and $v^+_B(q)$ are defined in line with the allocation probabilities. If $G(v)$ does not have gaps in its support, there always exists a distribution $\hat{F}_B$ that first order stochastically dominates F_B and has continuous support. In this case, $\hat{v}^-_B(q)y^-_B(q) - \hat{v}^+_B(q)y^+_B(q) = \hat{v}^-_B(q)\,(-J(q))$ and as values are decreasing in q, $\hat{v}^-_B(q)\,(-J(q)) > v^-_B(q)y^-_B(q) - v^+_B(q)y^+_B(q)$. Thus, there exists once again a deviation that leads to a higher total effort.

Last, suppose the support of the initial distribution has gaps. In this case, for a given q, there can be a jump for F_B, a jump for $\hat{F}_B$ or a jump for both. If there is a jump for exactly one distribution, first order stochastic dominance implies that $\hat{v}$ lies above v ($\hat{v}^+, \hat{v}^- > v$ or $v^+, v^- < \hat{v}$). If there is a jump for both distributions at the same q with the jump in values is larger under F_B, then inequality (47) is violated. We therefore require a more sophisticated approach. In this case, we need to amend expression (42) to

$$\mathbb{E}[\phi_B(q)y_B(q)] = \alpha_B y_B(1) + \sum_{q \in D} q\left[v^-_B(q)y^-_B(q) - v^+_B(q)y^+_B(q)\right] - \int_C qv_B(q)dy_B(q). \tag{48}$$

We compare distribution F_B to distribution $\hat{F}_B(v) = F_B(v)$ for $v \notin [v^+, \hat{v}^+]$, where $v^+ < \hat{v}^+$. This implies that $\hat{F}_B(\hat{v}^+) = F_B(v^+)$. Such a F_B yields a higher total effort than another distribution that displays a gap between v^+ and $\hat{v}^+$.

We can now flip this expression into the value-quantile-space, which yields

$$\alpha_B x_B(1) + \sum_{v \in D} v q_B(v) \left[x_B^+(v) - x_B^-(v) \right] + \int_{V_A \setminus D} q_B(v) v x_B'(v) dv, \tag{49}$$

Integrating over values allows to capture gaps in the support directly and thus a correction term is no longer needed. Thus, comparing total effort under $\hat{F}_B$ and F_B amounts to showing that

$$\int_{v^+}^{\hat{v}^+} \hat{q}_B(v) v \hat{x}_B'(v) dv \geq 0, \tag{50}$$

which always holds. Note that we keep here $\hat{x}_B'(v)$ to emphasise that this is not the allocation probability $x_B(v)$, but rather the allocation probability at the q associated with v_B, which differs from $\hat{v}_B$. ■

Proof of Proposition 4: Reallocating Value, Fixed Distribution We want to show that total effort is maximised by selecting distributions which maximise discrimination

$$F_B^*(v) = H(v) \qquad \text{if } v \in [\alpha, v^M) \tag{51}$$

$$F_A^*(v) = H(v) - 1 \qquad \text{if } v \in [v^M, \omega]. \tag{52}$$

To simplify notation, we drop the star to indicate the optimal adjustment and simply refer to these two distributions by $F_A(v)$ and $F_B(v)$. We proceed as follows:
(i) We first assume that each value v admits a density and show that $F_A(v)$ and $F_B(v)$ yield higher effort than any other two distributions, $\hat{F}_A(v)$ and $\hat{F}_B(v)$, that allocate strictly positive density at each $v \in [\alpha, \omega]$ – meaning that $\min\{\hat{f}_A(v), \hat{f}_B(v)\} > 0$ for all v.
(ii) We prove that $F_A(v)$ and $F_B(v)$ lead to higher total effort than any other two distributions with disjoint support. This allows to account for mass points.
(iii) We combine these insights to show that $F_A(v)$ and $F_B(v)$ yield a higher total effort than any other two distributions.

Splitting Densities We want to compare total effort with maximal discrimination to total effort with split densities, where $\min\{\hat{f}_A(v), \hat{f}_B(v)\} > 0$ for all $v \in [\alpha, \omega]$. Denote by

$a_i(v)$ the share of the density $h(v)$ assigned to worker $i \in \{A, B\}$. We want to show that

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

$$\geq \int_{\alpha}^{\omega} \hat{\psi}_A(v)\hat{x}_A(v)a_A(v)h(v)dv + \int_{\alpha}^{\omega} \hat{\psi}_B(v)\hat{x}_B(v)a_B(v)h(v)dv. \tag{54}$$

For convenience, define

$$\overline{x}_A(v) = \begin{cases} \max\{\hat{x}_A(v), \hat{x}_B(v)\} & \text{if } v \geq v^M \\ \min\{\hat{x}_A(v), \hat{x}_B(v)\} & \text{if } v < v^M \end{cases}, \tag{55}$$

$$\overline{x}_B(v) = \{\hat{x}_A(v), \hat{x}_B(v)\} \setminus \overline{x}_A(v). \tag{56}$$

Similarly, define for any $i, j \in \{A, B\}$ such that $i \neq j$

$$\overline{\psi}_i(v) = \begin{cases} \hat{\psi}_i(v) & \text{if } \hat{x}_i(v) = \overline{x}_i(v) \\ \hat{\psi}_j(v) & \text{if } \hat{x}_i(v) \neq \overline{x}_i(v) \end{cases}, \tag{57}$$

$$\overline{a}_i(v) = \begin{cases} a_i(v) & \text{if } \hat{x}_i(v) = \overline{x}_i(v) \\ a_j(v) & \text{if } \hat{x}_i(v) \neq \overline{x}_i(v) \end{cases}. \tag{58}$$

These definitions immediately imply that

$$\int_{\alpha}^{\omega} \hat{\psi}_A(v)\hat{x}_A(v)a_A(v)h(v)dv + \int_{\alpha}^{\omega} \hat{\psi}_B(v)\hat{x}_B(v)a_B(v)h(v)dv \tag{59}$$

$$= \int_{\alpha}^{\omega} \overline{\psi}_A(v)\overline{x}_A(v)\overline{a}_A(v)h(v)dv + \int_{\alpha}^{\omega} \overline{\psi}_B(v)\overline{x}_B(v)\overline{a}_B(v)h(v)dv. \tag{60}$$

As in previous proofs, we again compare total effort for distributions F_A and F_B with virtual values $\psi_A(v)$,$\psi_B(v)$ and allocation probabilities $\overline{x}_A$, $\overline{x}_B$ to total effort for any other distribution. To do so, we need to establish that $\overline{x}_A$ and $\overline{x}_B$ satisfy interim feasibility for distributions F_A and F_B.

Interim Feasibility Interim feasibility is satisfied if and only if

$$\int_{\max\{\overline{v}, v^M\}}^{\omega} \overline{x}_A(v)h(v)dv + \int_{\min\{\overline{v}, v^M\}}^{v^M} \overline{x}_B(v)h(v)dv \leq 1 - F_A(\overline{v})F_B(\overline{v}), \tag{61}$$

see Border (1991). First, let $\overline{v} > v^M$. Then, $F_B(\overline{v}) = 1$ and the problem simplifies to

$$\int_{\overline{v}}^{\omega} \overline{x}_A(v)h(v)dv \leq 2 - H(v), \tag{62}$$

which trivially holds. Next, let $\overline{v} < v^M$. In this case,

$$\int_{v^m}^{\omega} \overline{x}_A(v)h(v)dv + \int_{\overline{v}}^{v^M} \overline{x}_B(v)h(v)dv \leq 1 \tag{63}$$

as $F_A(\overline{v}) = 0$. Therefore, it suffices to show that

$$\int_{v^m}^{\omega} \overline{x}_A(v)h(v)dv + \int_{\alpha}^{v^M} \overline{x}_B(v)h(v)dv \leq 1 \tag{64}$$

Inequality (64) corresponds to the following inequality in the quantile-probability space:

$$\int_q \overline{y}_A(q)dq + \int_q \overline{y}_B(q)dq \leq 1 \tag{65}$$

We can alternatively integrate over allocation probabilities, which transforms the inequality to

$$\overline{y}_A(1) + \int_{\overline{y}_A} q_A(y)dy + \overline{y}_B(1) + \int_{\overline{y}_B} q_B(y)dy \leq 1 \tag{66}$$

The allocation probability for A, $\overline{y}_A$, lies between 1 and the allocation probability at the median value $\overline{y}_A(1)$. Inequality (66) takes into account that $\overline{y}_A(1) > 0$. This is as if the allocation probability displays a jump, in which case the quantile does not change. Formally, if there is a jump at some q we define

$$y^- \equiv \lim_{\epsilon \to 0} y(q - \epsilon) \tag{67}$$

$$y^+ \equiv \lim_{\epsilon \to 0} y(q + \epsilon) \tag{68}$$

and for every $y \in [y^-, y^+]$, $q(y)$ remains constant. The probability of promotion for B at the median is denoted by $\overline{y}_B(0)$ and it goes down to $\overline{y}_B(1)$.

We can express any $q_A = \hat{q}_A + \hat{q}_B$ and $q_B = \hat{q}_A + \hat{q}_B - 1$. To see this note that $q_A(v) = 1 - F_A(v) = 2 - H(v)$, $q_B(v) = 1 - F_B(v) = 1 - H(v)$ and $\hat{q}_A(v) + \hat{q}_B(v) = 2 - \hat{F}_A(v) + \hat{F}_B(v) = 2 - H(v)$. Note that in general, it will not hold that $q_A(y) = \hat{q}_A(y) + \hat{q}_B(y)$. However, there always exists a y' such that $q_A(y) = \hat{q}_A(y) + \hat{q}_B(y')$. As we integrate over all y and thus all y', we omit the dependence on y'. We can therefore replace q_A and q_B by $\hat{q}_A$ and $\hat{q}_B$ as follows:

$$\overline{y}_A(1) + \int_{\overline{y}_A} (\hat{q}_A(y) + \hat{q}_B(y))\, dy + \overline{y}_B(1) + \int_{\overline{y}_B} (\hat{q}_A(y) + \hat{q}_B(y) - 1)\, dy \leq 1 \tag{69}$$

Note that $\int_{\overline{y}_B} dy = \overline{y}_B(0) - \overline{y}_B(1)$. This implies

$$\overline{y}_A(1) - \overline{y}_B(0) + 2\overline{y}_B(1) + \int_{\overline{y}_A} (\hat{q}_A(y) + \hat{q}_B(y))\, dy + \int_{\overline{y}_B} (\hat{q}_A(y) + \hat{q}_B(y))\, dy \leq 1 \tag{70}$$

If $\overline{y}_A(1) = \overline{y}_B(0)$, that is there is no jump at the median value, then we can rewrite inequality (70) as

$$2\overline{y}_B(1) + \int_{\overline{y}_B(1)}^{\overline{y}_A(0)} \hat{q}_A(y)dy + \int_{\overline{y}_B(1)}^{\overline{y}_A(0)} \hat{q}_B(y)dy \leq 1 \tag{71}$$

$$\Leftrightarrow \qquad \int_q \hat{y}_A(q)dq + \int_y \hat{y}_B(q)dq \leq 1, \tag{72}$$

where the latter holds as $\hat{y}_A$ and $\hat{y}_B$ are the allocation probabilities for distributions $\hat{F}_A$ and $\hat{F}_B$. If $\overline{y}_A(1) > \overline{y}_B(0)$, that is there is a jump at the median value, inequality (70) can be expressed as

$$\begin{aligned}\overline{y}_A(1) - \overline{y}_B(0) + 2\overline{y}_B(1) + \int_{\overline{y}_B(1)}^{\overline{y}_B(0)} \hat{q}_A(y)dy + \int_{\overline{y}_A(1)}^{\overline{y}_A(0)} \hat{q}_A(y)dy \\ + \int_{\overline{y}_B(1)}^{\overline{y}_B(0)} \hat{q}_B(y)dy + \int_{\overline{y}_A(1)}^{\overline{y}_A(0)} \hat{q}_B(y)dy \leq 1\end{aligned} \tag{73}$$

Note that $\overline{y}_A(1) - \overline{y}_B(0) = \int_{\overline{y}_B(0)}^{\overline{y}_A(1)} q dy$, with $q = 1 = \hat{q}_A(y) + \hat{q}_B(y)$, for all $y \in [\overline{y}_B(0), \overline{y}_A(1)]$. We can then amend inequality (73) to

$$2\overline{y}_B(1) + \int_{\overline{y}_B(1)}^{\overline{y}_A(0)} \hat{q}_A(y)dy + \int_{\overline{y}_B(1)}^{\overline{y}_A(0)} \hat{q}_B(y)dy \leq 1 \tag{74}$$

$$\Leftrightarrow \qquad \int_q \hat{y}_A(q)dq + \int_y \hat{y}_B(q)dq \leq 1, \tag{75}$$

where the latter holds once again as we fixed an interim feasible allocation for $\hat{F}_A$ and $\hat{F}_B$.

As $\overline{x}_A$, $\overline{x}_B$ satisfy interim feasibility, it is sufficient to show that

$$\int_{v^M}^{\omega} \psi_A(v)\overline{x}_A(v)h(v)dv + \int_{\alpha}^{v^M} \psi_B(v)\overline{x}_B(v)h(v)dv \tag{76}$$

$$\geq \int_{\alpha}^{\omega} \overline{\psi}_A(v)\overline{x}_A(v)\overline{a}_A(v)h(v)dv + \int_{\alpha}^{\omega} \overline{\psi}_B(v)\overline{x}_B(v)\overline{a}_B(v)h(v)dv. \tag{77}$$

This is equivalent to establishing that

$$\underbrace{\int_{v^M}^{\omega} \left(\left(\psi_A(v) - \overline{\psi}_A(v)\overline{a}_A(v)\right)\overline{x}_A(v) - \overline{\psi}_B(v)\overline{x}_B(v)\overline{a}_B(v)\right) h(v)dv}_{\text{Part 1}} \tag{78}$$

$$+\underbrace{\int_{\alpha}^{v^M} \left(\left(\psi_B(v) - \overline{\psi}_B(v)\overline{a}_B(v)\right)\overline{x}_B(v) - \overline{\psi}_A(v)\overline{x}_A(v)\overline{a}_A(v)\right) h(v)dv}_{\text{Part 2}} \geq 0. \tag{79}$$

We first focus on the case of virtual values being weakly positive, both under maximal discrimination and for distributions that present a deviation, for all values and establish the inequality by signing the two parts in turn.

Part 1 Note that since $\overline{x}_A(v) \geq \overline{x}_B(v)$ when $v \geq v^M$, we have that

$$\int_{v^M}^{\omega} \left(\left(\psi_A(v) - \overline{\psi}_A(v)\overline{a}_A(v)\right)\overline{x}_A(v) - \overline{\psi}_B(v)\overline{x}_B(v)\overline{a}_B(v)\right) h(v)dv \tag{80}$$

$$\geq \int_{v^M}^{\omega} \left(\psi_A(v) - \overline{\psi}_A(v)\overline{a}_A(v) - \overline{\psi}_B(v)\overline{a}_B(v)\right)\overline{x}_A(v)h(v)dv = 0. \tag{81}$$

The right hand side of the last inequality is equal to zero because

$$\psi_A(v) - \overline{\psi}_A(v)\overline{a}_A(v) - \overline{\psi}_B(v)\overline{a}_B(v) = \psi_A(v) - \hat{\psi}_A(v)\hat{a}_A(v) - \hat{\psi}_B(v)\hat{a}_B(v) \tag{82}$$

$$= v - \frac{1-F_A(v)}{h(v)} - \hat{a}_A(v)\left(v - \frac{1-\hat{F}_A(v)}{\hat{a}_A(v)h(v)}\right) - \hat{a}_B(v)\left(v - \frac{1-\hat{F}_B(v)}{\hat{a}_B(v)h(v)}\right) \tag{83}$$

$$= \frac{1}{h(v)}\left(1 + F_A(v) - \hat{F}_A(v) - \hat{F}_B(v)\right) = 0, \tag{84}$$

where the final equality follows from $\hat{F}_A(v) + \hat{F}_B(v) = H(v)$ and $F_A(v) = H(v) - 1$. Therefore, the integral in Part 1 is necessarily non-negative as it is bounded below by zero. Moreover, the integral is strictly positive provided that $\overline{x}_A(v) \neq \overline{x}_B(v)$ for a positive measure of $v \geq v^M$.

Part 2 As in Part 1, note that since $\overline{x}_A(v) \leq \overline{x}_B(v)$ when $v < v^M$, we have that

$$\int_{\alpha}^{v^M} \left(\left(\psi_B(v) - \overline{\psi}_B(v)\overline{a}_B(v)\right)\overline{x}_B(v) - \overline{\psi}_A(v)\overline{x}_A(v)\overline{a}_A(v)\right) h(v)dv \tag{85}$$

$$\geq \int_{\alpha}^{v^M} \left(\psi_B(v) - \overline{\psi}_A(v)\overline{a}_A(v) - \overline{\psi}_B(v)\overline{a}_B(v)\right)\overline{x}_B(v)h(v)dv = \int_{\alpha}^{v^M} \overline{x}_B(v)dv \geq 0. \tag{86}$$

As in the previous part, the equality in the previous expression follows because

$$\psi_B(v) - \overline{\psi}_A(v)\overline{a}_A(v) - \overline{\psi}_B(v)\overline{a}_B(v) = \frac{1}{h(v)}\left(1 + F_B(v) - \hat{F}_A(v) - \hat{F}_B(v)\right) = \frac{1}{h(v)}, \quad (87)$$

where the final equality follows from $\hat{F}_A(v)+\hat{F}_B(v) = H(v)$ and $F_B(v) = H(v)$. Therefore, the integral in Part 2 is also non-negative as it is bounded below by zero. Moreover, the integral is strictly positive whenever either $\overline{x}_A(v) \neq \overline{x}_B(v)$ or $\overline{x}_B(v) > 0$ for a positive measure of $v < v^M$.

The two parts establish that splitting densities is strictly worse than maximal discrimination when $\overline{x}_B(v) > 0$ for some $v < v^M$, since $\int_\alpha^{v^M} \overline{x}_B(v)dv > 0$ given that $\overline{x}_B(v)$ is increasing by incentive compatibility. However, splitting densities is strictly worse than maximal discrimination even when $\overline{x}_B(v) = 0$ for all $v < v^M$, because the optimal allocation under maximal discrimination must differ from $\overline{x}$ given that worker B must be promoted with positive probability when v_B is smaller but close v^M – meaning that for such values v_B we have that $x_B(v_B) > \overline{x}_B(v_B) = 0$.

So far we assumed that all virtual values are weakly positive. We relax this assumption and show that even with negative virtual values our result continues to hold.

Case 1 $\psi_A(v), \psi_B(v) > 0$ for all v, $\overline{\psi}_i(v) \geq 0$ for all $v \geq v^M$, $\overline{\psi}_i(v) < 0$ for some $v < v^M$

First, note that $\psi_B(v) > \overline{\psi}_i(v)$ for all $v < v^M$. In this case, Part 1 remains unchanged, while Part 2 needs to be amended. If both $\overline{\psi}_A(v), \overline{\psi}_B(v) < 0$ for some v, the difference in virtual values for each v is trivially positive. If $\overline{\psi}_B(v) > 0 > \overline{\psi}_A(v)$ for some v, then for these values

$$\left(\psi_B(v) - \overline{\psi}_B(v)\overline{a}_B(v)\right)\overline{x}_B(v) - \overline{\psi}_A(v)\overline{x}_A(v)\overline{a}_A(v) > 0, \quad (88)$$

as $\psi_B(v) - \overline{\psi}_B(v)\overline{a}_B(v) > 0$ and $-\overline{\psi}_A(v)\overline{x}_A(v)\overline{a}_A(v) \geq 0$. If for some v, $\overline{\psi}_A(v) > 0 > \overline{\psi}_B(v)$

$$\begin{aligned}&\left(\psi_B(v) - \overline{\psi}_B(v)\overline{a}_B(v)\right)\overline{x}_B(v) - \overline{\psi}_A(v)\overline{x}_A(v)\overline{a}_A(v) && (89)\\ &\quad > \left(\psi_B(v) - \overline{\psi}_B(v)\overline{a}_B(v) - \overline{\psi}_A(v)\overline{a}_A(v)\right)\overline{x}_B(v) > 0, && (90)\end{aligned}$$

where the last inequality holds as $\psi_B(v) > \overline{\psi}_A(v)\overline{a}_A(v)$.

Case 2 $\psi_A(v), \overline{\psi}_i(v) > 0$ for $v \geq v^M$, $\psi_B(v) < 0$ for some $v < v^M$

If the virtual value is negative under maximal discrimination for some v, it holds that $0 > \psi_B(v) > \overline{\psi}_i(v)$. Note that the allocation probability must not necessarily be zero, as regularity is not assumed. Now instead of assigning $\overline{x}_B$ to $\psi_B(v)$, assign $\overline{x}_A$ for all values for which $\psi_B(v) < 0$. Given that $\overline{x}_A(v) < \overline{x}_B$ interim feasibility continues to hold. Then, for these v

$$\psi_B(v)\overline{x}_A(v) - \overline{\psi}_B(v)\overline{a}_B(v)\overline{x}_B(v) - \overline{\psi}_A(v)\overline{x}_A(v)\overline{a}_A(v) \tag{91}$$

$$\geq \psi_B(v)\overline{x}_A(v) - \overline{\psi}_B(v)\overline{a}_B(v)\overline{x}_A(v) - \overline{\psi}_A(v)\overline{x}_A(v)\overline{a}_A(v), \tag{92}$$

which, by the same logic as in Part 2 is positive.

Case 3 $\psi_A(v) \geq 0$ for all $v \geq v^M$, but $\overline{\psi}_i(v) < 0$ for some $v \geq v^M$. Suppose first that both $\overline{\psi}_A(v), \overline{\psi}_B(v) < 0$. In this case, Part 1 is trivially fulfilled. Assume next that there exists some v such that $\overline{\psi}_A(v) < 0 < \overline{\psi}_B(v)$. In this case, we construct a profitable deviation, that coincides with $\overline{F}_A, \overline{F}_B$ for all values below some threshold $\underline{v}$ and maximal discrimination above $\underline{v}$. We proceed with this approach until the candidate for the profitable deviation does not contain negative virtual values above the median anymore, in which case either Case 1 or Case 2 apply.

To construct such a deviation, note that $\overline{\psi}_A(\omega) = \overline{\psi}_B(\omega) = \omega$, by assumption. This implies that there exists a $\overline{v}$, such that for all $v > \overline{v}$, $\overline{\psi}_A(v), \overline{\psi}_B(v) > 0$. The mass between $\overline{v}$ and ω is given by $m = 2 - H(\overline{v}) > 0$. Construct another cutoff $\underline{v}$ such that $H(\overline{v}) - H(\underline{v}) = m$. Now consider a distribution $\tilde{F}_A(v), \tilde{F}_B(v)$ which corresponds to $\overline{F}_A(v), \overline{F}_B(v)$ for $v < \underline{v}$ and

$$\tilde{F}_A(v) = \begin{cases} \overline{F}_A(v) & \forall \quad v \leq \underline{v} \\ \overline{F}_A(\underline{v}) & \forall \quad \underline{v} < v \leq \overline{v} \\ H(v) - \overline{F}_A(\underline{v}) & \forall \quad \overline{v} < v, \end{cases} \tag{93}$$

$$\tilde{F}_B(v) = \begin{cases} \overline{F}_B(v) & \forall \quad v \leq \underline{v} \\ H(v) - \overline{F}_B(\underline{v}) & \forall \quad \underline{v} < v \leq \overline{v} \\ 1 & \forall \quad \overline{v} < v, \end{cases} \tag{94}$$

As $\tilde{F}$ and $\overline{F}$ coincide for values below $\underline{v}$, we focus on $v > \underline{v}$. Applying the same approach as in Part 1, for $v > \overline{v}$ it must hold that

$$1 + \tilde{F}_A(v) - \hat{F}_A(v) - \hat{F}_B(v) = 1 + H(v) - \overline{F}_A(\underline{v}) - H(v) > 0. \tag{95}$$

Part 2 can be similarly amended. We have now a new candidate for a profitable deviation $\tilde{F}$. Repeating the same steps if virtual values are negative for some values $v > v^M$, $\tilde{\psi}_i(v) < 0$ leads again to maximal discrimination being optimal.

Case 4 Last, suppose that $\psi_A(v) < 0$ for some $v \geq v^M$. In this case, $\overline{\psi}_i(v) < 0$ for all i.

Assign $\overline{x}_B(v)$ to F_A. Then, we obtain

$$\psi_A(v)\overline{x}_B(v) - \overline{\psi}_A(v)\overline{a}_A(v)\overline{x}_A(v) - \overline{\psi}_B(v)\overline{x}_B(v)\overline{a}_B(v) \tag{96}$$

$$> \left(\psi_A(v) - \overline{\psi}_A(v)\overline{a}_A(v) - \overline{\psi}_B(v)\overline{a}_B(v)\right)\overline{x}_B(v) = 0, \tag{97}$$

as before.

Disjoint Support This proof allows for distributions $H(v)$ for which densities are not defined. If the density is defined, we assign each density to one distribution. This implies that the support of at least one distribution is disjoint. We restrict attention to alternative distributions with $\hat{v}_B \in [\alpha, t_1] \cup [t_2, t_3]$ and $\hat{v}_A \in [t_1, t_2] \cup [t_3, \omega]$, with $\alpha < t_1 < t_2 < t_3 \leq \omega$. If these distributions do not yield higher total effort than maximal discrimination, then splitting the distribution further cannot be optimal either. Note that $t_1 < v^M < t_3$.
Recall that total effort in the quantile space if given by

$$\int_q \phi_A(q)y_A(q)dq + \phi_B(q)y_B(q)dq \tag{98}$$

Integrating each component by parts yields

$$v_i(q)qy_i(q)\Big|_0^1 + \int_q v_i(q)\left(-y_i'(q)\right)qdq \tag{99}$$

$$= v_i(1)y_i(1) + \int_C v_i(q)\left(-y_i'(q)\right)qdq + \sum_{q\in D} q\left[v^-(q)y^-(q) - v^+(q)y^+(q)\right] \tag{100}$$

Flipping this to a $v-q-$ space yields

$$\alpha_A x_A(\alpha_A) + \int_{V_A\setminus D} vq_A(v)x_A'(v)dv \tag{101}$$

$$+\alpha_B x_B(\alpha_B) + \int_{V_B\setminus D} vq_B(v)x_B'(v)dv + \sum_{v\in D} q_{A\setminus B}(v)v\left[x^+_{A\setminus B}(v) - x^-_{A\setminus B}(v)\right], \tag{102}$$

which allows us once more to directly account for gaps in the support. We use $A\setminus B$ to signify that the variable may either belong to distribution A or B.

We consider three distinct cases:

1. both distributions with disjoint support, cutoff t_2 below v^M, $t_2 \leq v^M$
2. both distributions with disjoint support, cutoff t_2 above v^M, $t_2 > v^M$
3. one distribution with disjoint support: $\hat{v}_B \in [\alpha, t_1] \cup [t_2, \omega]$ and $\hat{v}_A \in [t_1, t_2]$

We relabel $\hat{x}_i(v)$, $\hat{q}_i(v)$ as $\tilde{x}_A(v)$, $\tilde{q}_A(v)$ if $v > v^M$ and as $\tilde{x}_B(v)$, $\tilde{q}_B(v)$ if $v < v^M$. Then, it is sufficient to show that

$$v^M \tilde{x}_A(v^M) + \int_{v^M}^{\omega} q_A(v)\tilde{x}'_A(v)vdv + \int_{\alpha}^{v^M} q_B(v)\tilde{x}'_B(v)vdv + \sum_{v\in D} q_{A\setminus B}(v)v\left[x^+_{A\setminus B}(v) - x^-_{A\setminus B}(v)\right], \tag{103}$$

$$> t_1\tilde{x}_B(t_1) + \int_{v^M}^{\omega} \tilde{q}_A(v)\tilde{x}'_A(v)vdv + \int_{\alpha}^{v^M} \tilde{q}_B(v)\tilde{x}'_B(v)vdv + \sum_{v\in D} \tilde{q}(v)v\left[\tilde{x}^+_{A\setminus B}(v) - \tilde{x}^-_{A\setminus B}(v)\right]. \tag{104}$$

Disjoint Support for $\hat{F}_A, \hat{F}_B$: $t_2 \leq v^M$ In this case inequality (103) can be amended to

$$v^M \tilde{x}_A(v^M) + \int_{v^M}^{t_3} q_A(v)\tilde{x}'_A(v)vdv + \int_{t_1}^{v^M} q_B(v)\tilde{x}'_B(v)vdv + \sum_{v\in D} q_{A\setminus B}(v)v\left[x^+_{A\setminus B}(v) - x^-_{A\setminus B}(v)\right], \tag{105}$$

$$> t_1\tilde{x}_B(t_1) + \int_{v^M}^{t_3} \tilde{q}_A(v)\tilde{x}'_A(v)vdv + \int_{t_1}^{v^M} \tilde{q}_B(v)\tilde{x}'_B(v)vdv + \sum_{v\in D} \tilde{q}_{A\setminus B}(v)v\left[\tilde{x}^+_{A\setminus B}(v) - \tilde{x}^-_{A\setminus B}(v)\right], \tag{106}$$

as the distributions are identical below t_1 and above t_3. Note that $\tilde{q}_A(v) = \hat{q}_B(v)$ for $v \in [v^M, t_3]$. Taking the difference yields

$$\int_{v^M}^{t_3} \left(q_A(v) - \hat{q}_B(v)\right)\tilde{x}'_A(v)vdv \tag{107}$$

where

$$q_A(v) = 2 - H(v) \tag{108}$$

while

$$\hat{q}_B(v) = 1 - \hat{F}_B = 1 - (H(v) - H(t_2) + H(t_1)) \tag{109}$$

Then, for a given $v \in [v^M, t_3]$

$$q_A(v) - \hat{q}_B(v) = 1 - H(v) + (H(v) - H(t_2) + H(t_1)) = 1 - H(t_2) + H(t_1) > 0 \tag{110}$$

implying that the difference is constant. Expression (107) then becomes:

$$(1 - H(t_2) + H(t_1))\int_{v^M}^{t_3} \tilde{x}'_A(v)vdv \tag{111}$$

We then turn to

$$\int_{t_1}^{v^M} \left(q_B(v) - \tilde{q}_B(v)\right) \tilde{x}'_B(v) v dv \tag{112}$$

Between t_2 and v^M, $\tilde{q}_B(v) = \hat{q}_B(v)$:

$$q_B(v) - \hat{q}_B(v) = 1 - H(v) - (1 - (H(v) - H(t_2) + H(t_1))) \tag{113}$$

$$= 1 - H(v) - 1 + (H(v) - H(t_2) + H(t_1)) = -\left(H(t_2) - H(t_1)\right) < 0 \tag{114}$$

which is again constant. For $v \in [t_1, t_2]$, this difference is given by

$$q_B(v) - \hat{q}_A(v) = 1 - H(v) - (1 - (H(v) - H(t_1))) \tag{115}$$

$$= -H(v) + (H(v) - H(t_1)) = -H(t_1) \tag{116}$$

Expression (112) is then equivalent to

$$\int_{t_1}^{v^M} \left(q_B(v) - \tilde{q}_B(v)\right) \tilde{x}'_B(v) v dv = -\left(H(t_2) - H(t_1)\right) \int_{t_2}^{v^M} \tilde{x}'_B(v) v dv - H(t_1) \int_{t_1}^{t_2} \tilde{x}'_B(v) v dv \tag{117}$$

Collecting all terms leads to the following comparison

$$v^M \tilde{x}_A(v^M) + (1 - H(t_2) + H(t_1)) \int_{v^M}^{t_3} \tilde{x}'_A(v) v dv \tag{118}$$

$$-\left(H(t_2) - H(t_1)\right) \int_{t_2}^{v^M} \tilde{x}'_B(v) v dv - H(t_1) \int_{t_1}^{t_2} \tilde{x}'_B(v) v dv \tag{119}$$

$$+\left(1 - H(t_2) + H(t_1)\right) \sum_{v \in D \cap [v^M, t_3]} v \left[\tilde{x}^+_A(v) - \tilde{x}^-_A(v)\right] \tag{120}$$

$$-\left(H(t_2) - H(t_1)\right) \sum_{v \in D \cap [t_2, v^M]} v \left[\tilde{x}^+_B(v) - \tilde{x}^-_B(v)\right] \tag{121}$$

$$-H(t_1) \sum_{v \in D \cap [t_1, t_2]} v \left[\tilde{x}^+_B(v) - \tilde{x}^-_B(v)\right] > t_1 \tilde{x}_B(t_1) \tag{122}$$

Rearranging yields

$$v^M \tilde{x}_A(v^M) + \int_{v^M}^{t_3} \tilde{x}'_A(v) v dv - (H(t_2) - H(t_1)) \int_{t_2}^{t_3} \tilde{x}'_{A\setminus B}(v) v dv - H(t_1) \int_{t_1}^{t_2} \tilde{x}'_B(v) v dv \tag{123}$$

$$+ \sum_{v \in D \cap [v^M, t_3]} v \left[\tilde{x}_A^+(v) - \tilde{x}_A^-(v)\right] - (H(t_2) - H(t_1)) \sum_{v \in D \cap [t_2, t_3]} v \left[\tilde{x}_{A\setminus B}^+(v) - \tilde{x}_{A\setminus B}^-(v)\right] \tag{124}$$

$$-H(t_1) \sum_{v \in D \cap [t_1, t_2]} v \left[\tilde{x}_B^+(v) - \tilde{x}_B^-(v)\right] > t_1 \tilde{x}_B(t_1) \tag{125}$$

As $H(t_1), H(t_2) - H(t_1) < 1$, it suffices to show that

$$t_3 \tilde{x}_A(t_3) - \int_{v^M}^{t_3} \tilde{x}_A(v) dv - (t_3 \tilde{x}_A(t_3) - t_1 \tilde{x}_B(t_1)) + \int_{t_1}^{t_3} \tilde{x}(v) dv > \tilde{x}_B(t_1) t_1 \tag{126}$$

$$\Leftrightarrow \qquad \int_{t_1}^{v^M} \tilde{x}_B(v) dv > 0, \tag{127}$$

which always holds and establishes that any other disjoint distributions do not yield a higher total effort. As we integrate over allocation probabilities, potential discontinuities are then once again subsumed.

Disjoint Support for $\hat{F}_A, \hat{F}_B$: $t_2 > v^M$ By the same logic as in the previous case, it is sufficient to show

$$v^M \tilde{x}_A(v^M) + \int_{v^M}^{t_3} q_A(v) \tilde{x}'_A(v) v dv + \int_{t_1}^{v^M} q_B(v) \tilde{x}'_B(v) v dv \tag{128}$$

$$> t_1 \tilde{x}_B(t_1) + \int_{v^M}^{t_3} \tilde{q}_A(v) \tilde{x}'_A(v) v dv + \int_{t_1}^{v^M} \tilde{q}_B(v) \tilde{x}'_B(v) v dv \tag{129}$$

We can ignore any potential discontinuities in allocation probabilities here, as we integrate at the end, meaning the discontinuities will vanish once again. As before, with a different integration bound below

$$\int_{t_2}^{t_3} q_A(v) - \hat{q}_B(v) \tilde{x}'_A(v) v dv = (1 - H(t_2) + H(t_1)) \int_{t_2}^{t_3} \tilde{x}'_A(v) v dv \tag{130}$$

We turn to

$$\int_{v^M}^{t_2} (q_A(v) - \hat{q}_A(v)) \, \tilde{x}'_A(v) v dv \tag{131}$$

$$q_A(v) - \hat{q}_A(v) = 2 - H(v) - (1 - (H(v) - H(t_1))) \tag{132}$$

$$= 1 - H(v) + (H(v) - H(t_1)) = 1 - H(t_1) \tag{133}$$

Then,

$$\int_{v^M}^{t_2} \left(q_A(v) - \hat{q}_A(v)\right) \tilde{x}'_A(v)vdv = (1 - H(t_1)) \int_{v^M}^{t_2} \tilde{x}'_A(v)vdv \tag{134}$$

Last,

$$\int_{t_1}^{v^M} \left(q_B(v) - \hat{q}_A(v)\right) \tilde{x}'_A(v)vdv = -H(t_1) \int_{t_1}^{v^M} \tilde{x}'_B(v)vdv \tag{135}$$

Collecting terms yields

$$v^M \tilde{x}_A(v^M) + (1 - H(t_2) + H(t_1)) \int_{t_2}^{t_3} \tilde{x}'_A(v)vdv + (1 - H(t_1)) \int_{v^M}^{t_2} \tilde{x}'_A(v)vdv \tag{136}$$

$$- H(t_1) \int_{t_1}^{v^M} \tilde{x}'_B(v)vdv > t_1 \tilde{x}_B(t_1) \tag{137}$$

As before it is sufficient to show that

$$t_3 \tilde{x}_A(t_3) - \int_{v^M}^{t_3} \tilde{x}_A(v)dv - (t_3 \tilde{x}_A(t_3) - t_1 \tilde{x}_B(t_1)) + \int_{t_1}^{t_3} \tilde{x}_{A \setminus B}(v)dv > \tilde{x}_B(t_1)t_1 \tag{138}$$

$$\Leftrightarrow \qquad \int_{t_1}^{v^M} \tilde{x}_B(v)dv > 0 \tag{139}$$

This establishes that also in this case, maximal discrimination yields the highest effort. *Disjoint Support for $\hat{F}_B$: $t_2 > v^M$* Similar to the previous case, we can ignore discontinuities in the allocation probabilities as they will once again vanish in the end. It is sufficient to show

$$v^M \tilde{x}_A(v^M) + \int_{v^M}^{t_2} q_A(v)\tilde{x}'_A(v)vdv + \int_{t_1}^{v^M} q_B(v)\tilde{x}'_B(v)vdv \tag{140}$$

$$> t_1 \tilde{x}_B(t_1) + \int_{v^M}^{t_2} \tilde{q}_A(v)\tilde{x}'_A(v)vdv + \int_{t_1}^{v^M} \tilde{q}_B(v)\tilde{x}'_B(v)vdv \tag{141}$$

As before,

$$\int_{v^M}^{t_2} q_A(v) - \hat{q}_A(v)\tilde{x}'_A(v)vdv = (1 - H(t_1)) \int_{v^M}^{t_2} \tilde{x}'_A(v)vdv \tag{142}$$

$$\int_{t_1}^{v^M} q_B(v) - \hat{q}_A(v)\tilde{x}'_A(v)vdv = -H(t_1) \int_{t_1}^{v^M} \tilde{x}'_A(v)vdv \tag{143}$$

Following the same steps as before establishes that maximal discrimination yields higher total effort than this candidate.

Mixing Split Densities and Disjoint Supports It follows that there cannot be any other two distributions that combine mixing split densities and disjoint supports that yield higher total effort compared to maximal discrimination. We can always treat subsets of the distribution as the entire distribution. This just requires an adjustment of the mass in a certain subset. Then we can perform the same analysis as we did for a subset, which yields lower total effort for this subset by the same logic as above.

Therefore, we have established that maximal discrimination is optimal and yields highest total effort among all possible distributions.■

References

Ali, S. N., N. Haghpanah, X. Lin, and R. Siegel (2020). How to sell hard information. *arXiv preprint arXiv:2010.08037*.

Altonji, J. G. and C. R. Pierret (2001). Employer learning and statistical discrimination. *The quarterly journal of economics 116*(1), 313–350.

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

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

Bergemann, D. and M. Pesendorfer (2007). Information structures in optimal auctions. *Journal of economic theory 137*(1), 580–609.

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.

Border, K. C. (1991). Implementation of reduced form auctions: A geometric approach. *Econometrica: Journal of the Econometric Society*, 1175–1187.

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.

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

Crémer, J. (1993). Corporate culture and shared knowledge. *Industrial and corporate change 2*(3), 351–386.

Crémer, J. and R. P. McLean (1985). Optimal selling strategies under uncertainty for a discriminating monopolist when demands are interdependent. *Econometrica 53*, 345–361.

Deal, T. E. and A. A. Kennedy (1982). *Corporate cultures: The rites and rituals of corporate life.* Addison-Wesley.

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.*

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.

Gibbons, R. and R. S. Kaplan (2015). Formal measures in informal management: can a balanced scorecard change a culture? *American Economic Review 105*(5), 447–51.

Gibbons, R. and L. Prusak (2020). Knowledge, stories, and culture in organizations. In *AEA Papers and Proceedings*, Volume 110, pp. 187–92.

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

Green, T. K. (2005). Work culture and discrimination. *Calif. L. Rev. 93*, 623.

Greenwald, A., T. Oyakawa, and V. Syrgkanis (2017). Simple vs optimal mechanisms in auctions with convex payments. *arXiv preprint arXiv:1702.06062*.

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

Hermalin, B. E. (1999). Economics & corporate culture. *Available at SSRN 162549*.

Hermalin, B. E. (2012). Leadership and corporate culture. *Handbook of organizational economics*, 432–78.

Hodgson, G. M. (1996). Corporate culture and the nature of the firm. In *Transaction cost economics and beyond*, pp. 249–269. Springer.

Johnson, J. P. and D. P. Myatt (2006). On the simple economics of advertising, marketing, and product design. *American Economic Review 96*(3), 756–784.

Kamenica, E. and M. Gentzkow (2011). Bayesian persuasion. *American Economic Review 101*(6), 2590–2615.

Kreps, D. M. (1990). Corporate culture and economic theory. *Perspectives on positive political economy 90*(109-110), 8.

Lazear, E. P. (1995). Corporate culture and the diffusion of values. *Trends in business organization: do participation and cooperation increase competitiveness*, 89–133.

Lazear, E. P. and S. Rosen (1981). Rank-order tournaments as optimum labor contracts. *Journal of political Economy 89*(5), 841–864.

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

Martinez, E. A., N. Beaulieu, R. Gibbons, P. Pronovost, and T. Wang (2015). Organizational culture and performance. *American Economic Review 105*(5), 331–35.

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.

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.

Ouchi, W. G. and W. G. Cuchi (1981). *Theory Z: How American business can meet the Japanese challenge*, Volume 1081. Addison-Wesley Reading, MA.

Padavic, I., R. J. Ely, and E. M. Reid (2020). Explaining the persistence of gender inequality: The work–family narrative as a social defense against the 24/7 work culture. *Administrative Science Quarterly 65*(1), 61–111.

Peters, T. J. and R. H. Waterman (1982). In search of excellence: Lessons from america's best-run companies.

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.

Sorokin, C. and E. Winter (2018). Pure information design in classical auctions. Technical report, Working paper.

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