Profitable Inequality*

Anja Prummer[†] Francesco Nava[‡]

October 29, 2021

Abstract

What is an employer's optimal design of values for promotion among his workers? Workers compete by exerting effort, and higher effort induced by greater valuations corresponds to higher profits 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, potentially reducing effort. We show that if value is redistributed within or across workers, the reduction in information rent outweighs potential losses due to inequality. Maximal dispersion and maximal discrimination emerge as features of optimal designs. We confirm our theoretical predictions in an empirical application.

Keywords: Mechanism Design, Information Design, Culture, Discrimination

*We thank Jerome Adda, Arjada Bardhi, Chris Bidner, Francesca Cornaglia, David McAdams, Bernhard Schmidpeter, 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, JKU Linz, King's College London, Penn State, POLECONUK, PUC-Rio, QMUL, Simon Fraser, University of Antwerp, UNC, University College Dublin, University of Texas at Austin and the Southern Economic Association. All remaining errors are ours. First draft: February 2020.

[†]Johannes-Kepler-University Linz and Queen Mary University of London, anja.prummer@jku.at

[‡]London School of Economics, f.nava@lse.ac.uk

1 Introduction

Employers incentivise their workers to exert effort using, for instance, competition for a promotion. How much a worker applies himself depends on his value for the promised compensation package, the work environment as well as the anticipated tasks and requirements of the new job– features determined by the firm. Workers may value the promotion differently either due to idiosyncratic preferences or variations in skills and ability, all of which are unobservable to the firm. The employer may treat workers distinctly and can tailor the compensation package to each of them, generating further disparities in values for the promotion. Thus, the employer obtains an induced value distribution, which is a function of preferences, skills as well as the potentially targeted characteristics of the promotion. Does the employer prefer to induce distinct value distributions for the promotion across workers?

Our main contribution is a full characterisation of the optimal value distributions subject to potential constraints the employer may face. In our setting, the employer maximises total effort, defined as the sum of effort across workers. We consider total effort as a proxy for the firm's profit, a natural assumption for numerous sectors such as finance, consulting and law– all known for their long hours and work hard mentality. Effort is observable, in line with new technologies facilitating monitoring, but can also be interpreted as billable hours, returns on investments or projects completed. The employer implements the optimal mechanism (Myerson (1981)) to award a promotion, while inducing workers to reveal their private net value for the promotion. The optimal mechanism prescribes a specific effort level and probability of promotion for each induced valuation. Departing from canonical mechanism design, we seem to be the first to equip the employer with influence on the valuation. Perhaps surprisingly, maximal dispersion of values is optimal if the induced distribution cannot exceed a fixed expected value for each worker individually. If the employer can design the distributions of his workers jointly, then he generates maximally distinct distributions.[1] Inequality turns out to be profitable. We confirm that value distributions exhibit the predicted features in an empirical application.

The design of the value distribution has three effects: (i) it affects the worker's effort choice directly;[2] (ii) it impacts the promotion probability, which in turn influences the effort choices; (iii) it changes the *information rent*, the employer provides to the worker to reveal his induced valuation truthfully, ensuring that the mechanism is incentive compatible. These forces determine the optimal value distribution for the employer, where we distinguish between individually optimal distributions and jointly optimal distributions.

Individually Optimal Distributions Suppose first that the employer can generate any distribution for a given worker. To minimise the information rent, the employer assigns all mass to one value. In this case, he 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 prohibitive, which leads us to impose

[1] The joint distribution does not imply that values are correlated, for a discussion of why our analysis readily extends to correlated values, see Appendix B.

[2] In more technical terms, this is the contribution to total surplus.

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. If there is only worker with value $\overline{\omega}$ the employer specifies effort $\overline{\omega}$ for the worker to attain the promotion. The firm therefore extracts the entire surplus and reduces the utility of the worker to zero. The employer cannot do better than obtain the entire surplus, establishing that one worker with a high value is sufficient.

Second, we consider that the employer can generate any distribution as long as its expected value is below some constant $k < \overline{\omega}$, a constraint mirroring Kamenica and Gentzkow (2011)'s *Bayes' plausibility.* For ease of exposition, we restrict attention for now to the employer designing the value distribution of exactly one worker. In this case it is optimal to generate a binary distribution, with mass at $\overline{\omega}$ and zero, such that the mean of the distribution is equal to k. The employer prefers the highest possible expected value and therefore sets it equal to k, the upper bound. 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. Additionally, the distribution has an impact on the effort choices of the other workers. Workers exert higher effort the greater their probability of attaining the promotion. Facing a co-worker with a 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 of the generated distribution as high as possible as this reduces the probability for the other worker to face such a high value competitor. This implies the other worker's effort remains high. In contrast to the common perception that higher variance means less information about the worker's values, our result highlights that this is only the case for sufficiently high values: values for which the worker can attain the promotion with some probability. If variance is determined by the spread between values the employer cares about and excluded values, then higher variance does not imply worse information.

Bounding the expected value encompasses a constraint that restricts the employer to generate distributions that are second order stochastically dominated by some given distribution. Choosing a riskier distribution is always beneficial for the firm. However, if the employer can only create distributions that are first order stochastically dominated by some specified distribution, 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.

Jointly Optimal Distributions The employer can now generate distributions for two workers jointly, by re-allocating value across workers. In this context, we first assume that sum of the expected values of the workers cannot exceed a constant. In this case, the employer assigns all value to one worker, reducing the other worker's value to zero. The intuition for this result mirrors that of the benchmark case without any constraint: it is sufficient to have one worker with a high valuation. This result is stark and it raises questions as to whether it holds more generally. To assess this, we fix an arbitrary measure of mass two, with cumulative $H(v)$, where v denotes the induced value for the promotion. The employer can create any two distributions such that these distributions match $H(v)$. Put differently, the sum of any two distributions amounts to $H(v)$. Again, it turns out to be optimal to create maximally distinct value distributions. One worker's values lies below the median of $H(v)$, while for the other worker they are above the median of $H(v)$. By creating these distinct distributions, the employer can extract higher effort due to a reduced information rent. The employer knows more precisely what value the worker has compared to implementing more similar distributions across workers and thus has to leave less information rent on the table. However, such inequality results in a reduction of effort by the disadvantaged worker. Surprisingly, the loss in effort from the worker whose values are lower is more than countered by the increase in effort by the advantaged worker, making maximal discrimination optimal.

This results highlights that if employers can introduce inequalities across workers or alternatively distinct groups of workers (these interpretations are interchangeable in the context of our model), they find it optimal to do so. Firms profit from providing advantages to one group of workers, to the detriment of the other. In particular, they create an environment, that helps them learn about the underlying values of their workers, a result typical of adverse selection models.

Notably, employers are not required to exhibit differences in tastes (Becker (1957)) or beliefs (Phelps (1972), Arrow (1973)) regarding their employees, to generate disparities. However, differences in tastes or beliefs in combination with profitable inequality makes it plausible that disparities arise along old and established fault lines, such as gender or race.

Empirical Application We propose a novel application of this theory by investigating disparities in promotion outcomes in law using surveys conducted by the American Bar Association. Lawyers are a particularly good fit for our model for a number of reasons. First, the associate-partner ratio is 2:1, implying that there is a competition for the promotion. In addition, law is a gender balanced profession. This allows us to split the sample of lawyers into two equally sized groups, men and women, in line with our theory. Notably, female lawyers match their male colleagues well, in terms of observables at the beginning of their career. While there is certainly a selection at the hiring stage it is not systematically different across groups. We measure the value for the promotion by aspirations to become an equity partner: aspirations capture how much an employee would like to rise through the ranks, how much he values becoming partner. These aspirations are influenced by workplace culture: if a lawyer has experienced demeaning or discriminatory comments within the first two years in the profession, then their aspirations are significantly lower. We measure the prevalence of demeaning and discriminatory comments in Wave I of the survey, but aspirations five years later, ruling out it is low aspirations that generate these negative experiences. In contrast, time spent with a partner is associated with a rise in aspirations. This finding highlights that experiences at the workplace, shaped by the firm, are tied to the value for the promotion.

Having demonstrated that the law sector matches our framework well, we turn to investigating the predictions of our model. We first consider the distributions of aspirations across groups. Our theory predicts that (i) when firms are able to reallocate value across groups, average levels of aspirations differ across groups, and that (ii) when firms can further redistribute value within groups, bi-modal distributions of aspiration arise. Indeed, we find that average values of aspirations are higher for men than women. The distributions are bi-modal, with the highest mass at the highest and lowest level of aspirations. These patterns are consistent with our model.

Linking aspirations and effort, we show that aspirations have a significantly positive effect on any of our measures of effort, such as billable hours. Additionally, as predicted the effort levels differ across groups. Last, we find that aspirations are an important determinant of becoming partner. Effort has a limited effect as well. In contrast, gender and variables that measure family constraints, such as the number of children are no longer significant.

Based on these findings, we argue that firms benefitting from inequalities in valuations for rising through the ranks, may be an overlooked source of the well known gender promotion gap. For instance, Bertrand, Goldin, and Katz (2010) document that the gender promotion gap among former MBA students is driven by women's lower hours as well as family constraints. Such a result would also emerge in our setting, if we did not control for aspirations. This finding has lead to calls for more childcare support and flexible hours.[3] Looking through the lens of our model generates a very different policy advice: if we aim for women to take on leadership positions, then it is crucial to create a welcoming work environment. The environment and workplace culture should be evaluated not based on whether it works for a select few, but rather whether it keeps everyone engaged.

Related Literature Methodologically, we contribute to the literature on value and information design in mechanisms. Condorelli and Szentes (2020), Roesler and Szentes (2017), and Bobkova (2019) focus on a buyer and seller relationship. 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.[4]

[3]Our explanation for the underrepresentation of women in management positions is complementary to those presented in Gayle and Golan (2011), Gneezy, Niederle, and Rustichini (2003), Niederle and Vesterlund (2007), Dohmen and Falk (2011), Erosa, Fuster, Kambourov, and Rogerson (2017). Independently, Azmat, Cuñat, and Henry (2020) ties aspirations to becoming partner for the same data set. While their focus is on understanding the determinants of becoming partner, ours is on the distribution of aspirations and demonstrating that valuations are a product of workplace experiences.

[4]Our result is in contrast to Cantillon (2008) who shows for first and second price auctions that asymmetries between bidders reduce revenue. Deb and Pai (2017) obtain discrimination while designing symmetric auctions, while we design value distributions.

Bergemann and Pesendorfer (2007) and Sorokin and Winter (2018) consider similar problems but in which the designer chooses which information about values to reveal to their privately-informed counter-parties, rather than directly designing the distribution of values. Such a Bayesian persuasion (Kamenica and Gentzkow (2011)) approach is closely related to our analysis because it amounts to designing posterior value distributions subject to a Bayes' plausibility constraint. Conclusions differ however considerably as the designer has more tools available to shape value distributions in each of our constrained design exercises.

Competition among workers for a prize, our promotion, has first been introduced by the seminal paper of Lazear and Rosen (1981). There, the promotion is given to the worker, who exerts the highest effort. The prize encourages workers to generate output; it provides them with incentives to work. Our setting departs from the usual rank order tournaments in two crucial ways: (i) we allow for unobserved heterogeneity among workers, which can be considered skill or ability, the employer would like to learn, in addition to generating incentives to exert effort and (ii) the optimal mechanism does not necessarily promote the worker with the highest effort and may not award the promotion at all. As total effort in an optimal mechanism cannot be improved upon by definition, it follows that the rank order tournament awards the prize too often. Most importantly, in contrast to much of the contest literature, in our setting it is optimal to introduce inequalities between workers, while in contests it is generally optimal to create equal agents, for an overview, see Mealem and Nitzan (2016). A notable exception to this is Pérez-Castrillo and Wettstein (2016), who provide numerical simulations, to show that under certain conditions asymmetric rewards can be optimal.[5]

Organisational Incentives and Culture Our concept of inequality across workers relates to Winter (2004)'s discrimination. 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 and allow for general tools to induce these asymmetries, one of them being corporate culture. Kreps (1990) first acknowledged its importance and the desirability to incorporate corporate culture into economic theory, 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.[6] In contrast, we focus on the effects of culture on worker's valuations for a promotion.

The remainder of the paper is organised as follows. We begin with an example to illustrate the forces at play in Section 2. In Section 3, we introduce the model. We solve for the optimal mechanism and distribution subject to a constraint in Section 4. We provide a benchmark result for the optimal distribution without constraints in Section 4.1. We analyse the case of individually optimal distributions, where we keep the average valuation weakly below some constant in Section 4.2. In this context, we also characterise the optimal distribution if the

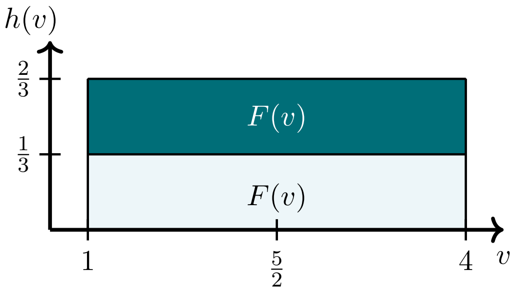
[5]Interestingly, in their contest, agents have private information, and the designer is interested in learning it, linking their set up closely to a mechanism design problem with additional restrictions.

[6]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).

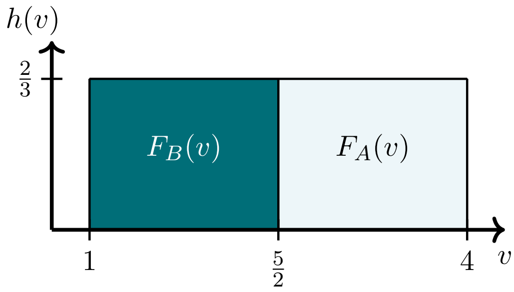
employer is required to restrict attention to a distribution that is first-order and second-order stochastically dominated. We then turn to the setting in which the employer can determine the distributions of workers jointly in Section 4.3. Section 5 matches our model to an empirical application, before testing our theoretical predictions. Section 6 provides a discussion and concludes.

2 An Example Of Jointly Optimal Distributions

For the purpose of this example, we allow the firm to generate the value distributions for two workers, A and B, jointly to match $H(v) = \frac{2(v-1)}{3}$ with $v \in [1, 4]$. We contrast the total effort if the firm treats their two workers equally and generates cumulative distribution functions $F(v) = \frac{(v-1)}{3}$ for both workers, with maximal discrimination. Maximal discrimination refers to two maximally distinct distributions, where one worker receives all the values below the median of $H(v)$, his distribution equals $F_B(v) = H(v)$ for $v < \frac{5}{2}$, while the other worker's distribution is given by $F_A(v) = H(v) - 1$ for $v > \frac{5}{2}$. These distributions are depicted in Figure 1.



(a) Equal Induced Distributions



(b) Maximal Discrimination

Figure 1: Equality versus Maximal Discrimination

The employer implements a mechanism which specifies the probability of promotion as well as an effort level for each induced valuation such that workers find it optimal to reveal their net value for the promotion (instead of pretending they have a different induced value).[7] We follow Myerson (1981) in mapping value distributions into effort and promotion probabilities. He showed that maximising the sum of efforts is equivalent to maximising the *expected virtual surplus*, which is a function of allocation probabilities and *virtual values*. The virtual value for distributions with defined densities is given by $\psi(v) = v - \frac{1-F(v)}{f(v)}$. The virtual surplus for a given profile of values v_A, v_B is then $\psi_A(v_A)x_A(v_A, v_B) + \psi_B(v_B)x_B(v_A, v_B)$, where x_A and x_B specify the expected promotion probabilities for A and B, respectively.[8] The worker with the highest realised virtual value receives the promotion, if his virtual value is positive. Otherwise, the promotion is not allocated.

[7]The mechanism needs to be incentive compatible and individually rational, we make this precise in Section 3

[8]When calculating the probability of obtaining the promotion, the worker knows his valuation, but only the distribution of the other worker's values. Note that the expected virtual surplus here is calculated for a *given* value v. That is, the expectation is taken over the other worker's value distribution. In our main analysis, we refer to expected virtual surplus when taking expectations over *all* value distributions.

We depict the expected virtual surplus of each worker, given his realised value, in Figure 2a. We calculate for each value the virtual value and the probability that the worker obtains the promotion. The expected virtual surplus for each value with maximal discrimination is above the expected virtual surplus if distributions are equal. However, there are *two* individuals with distribution F, but only one with F_A and F_B, each. Two workers with distribution F and realised values close to four generate a higher expected surplus, compared to two workers whose valuations are drawn from F_A and F_B. Nevertheless, maximal discrimination yields a higher expected surplus if one takes the densities of the different values into account, determined by their respective distributions. Formally, we want to show the following:

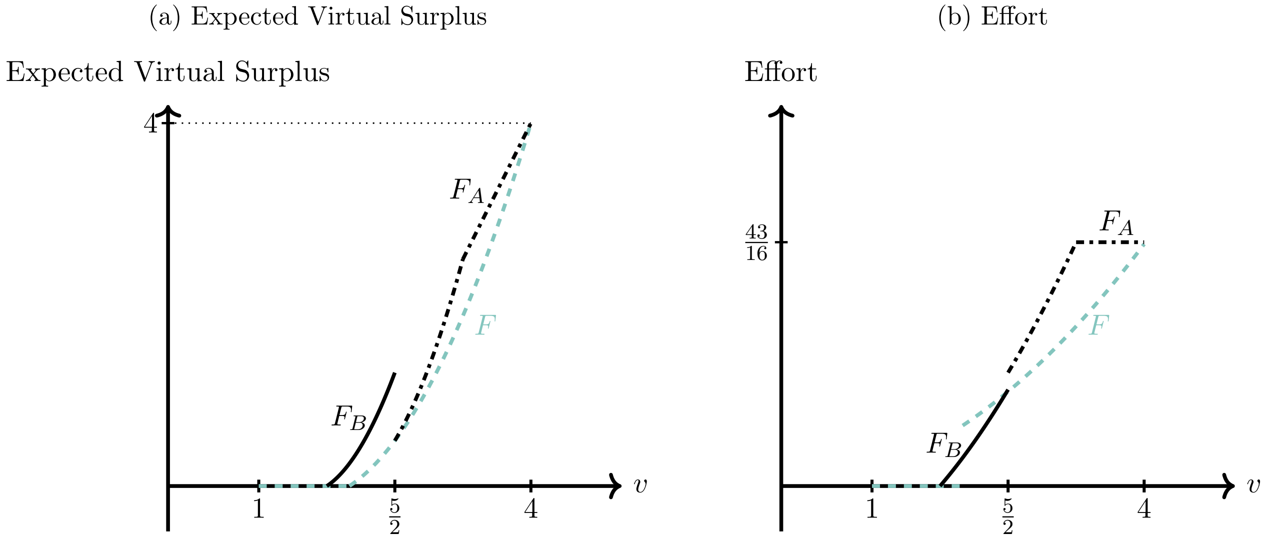
$$\int_{\frac{5}{2}}^{4} \psi_A(v)x_A(v)h(v)dv + \int_{1}^{\frac{5}{2}} \psi_B(v)x_B(v)h(v)dv > 2\int_{1}^{4} \psi(v)x(v)f(v)dv.$$

Note that $2f(v) = h(v) = H'(v)$ by definition and so the previous inequality simplifies to

$$\int_{\frac{5}{2}}^{4} \psi_A(v)x_A(v)h(v)dv + \int_{1}^{\frac{5}{2}} \psi_B(v)x_B(v)h(v)dv > \int_{1}^{4} \psi(v)x(v)h(v)dv.$$

It is as if one is comparing both F_A and F_B workers with *one* worker with distribution H as each value is associated with the same density. Therefore, 2a does in fact reveal that maximal discrimination is associated with a higher virtual surplus and in turn a higher effort.

Figure 2: Expected Virtual Surplus vs Effort



Note: Expected Virtual Surplus of Worker i: $\psi_i(v_i)x_i(v_i)$. The light, dashed line represents F, the solid line F_B and the dashed dotted line F_A.

The effort is made explicit in Figure 2b. Maximal discrimination generates a higher effort from those with high valuations, a lower effort for intermediate values and a higher effort for low values. The last effect is due to different exclusion thresholds. Maximal discrimination leads to fewer values being excluded, as the information rent is lower for each value. This is evident from

Figure 2a: the value for which the expected virtual surplus becomes positive is higher if both workers have the same distribution compared to maximally distinct distributions. If there is no exclusion, maximal discrimination reduces the effort of the worker with lower values. However, this reduction is more than compensated for by the increase in effort by the high value worker, making discrimination optimal.

3 A Theory of Promotion

We first revisit classical independent private value mechanism design contributions for given value distributions. We begin by introducing the worker's problem, before turning to the problem of the employer. There, we present the value design problems at the heart of the analysis.

3.1 Workers' Problem

Two workers, A and B, compete for a promotion by exerting effort. Worker i's, $i \in \{A, B\}$, induced value for the promotion v_i is private information and is independently distributed according to a cumulative distribution F_i with support $V_i \subseteq [\alpha_i, \omega_i]$. The distributions of values F_A and F_B are commonly known by the two workers and by the employer. The promotion is awarded according to a direct mechanism, $(x, e) : V_A \times V_B \to \Delta^2_- \times \mathbb{R}^2$, where $x_i(\mathbf{v})$ specifies the probability that i gets promoted, $e_i(\mathbf{v})$ denotes the effort that worker i has to exert for any profile of reported values $\mathbf{v} \in V_A \times V_B$, and Δ^2_- denotes the set of non-negative two-dimensional vectors whose entries add to no more than 1.

The mechanism we consider is incentive compatible (all workers benefit by reporting their valuation truthfully) and individually rational (all workers benefit by participating). This implies the existence of a Bayes Nash equilibrium in which all workers prefer to report their value truthfully and participate, meaning that for worker i with value v_i it must be that

$$v_i x_i(v_i) - e_i(v_i) \geq \max\{v_i x_i(z_i) - e_i(z_i), 0\} \quad \text{for any} \quad z_i \neq v_i, \tag{1}$$

where $x_i(v_i) = \mathbb{E}_{F_{-i}}[x_i(\mathbf{v})]$ denotes the expected probability of promotion and $e_i(v_i) = \mathbb{E}_{F_{-i}}[e_i(\mathbf{v})]$ the expected effort.[9] Further, when cumulatives are discontinuous on the support, Riemann-Stijltes integrals will be used to calculate the expectations implicitly.

3.2 Employer's Problem

The employer maximises *total effort*, the expected sum of efforts across workers, by implementing the described direct mechanism to assign the promotion. In addition, he designs the value distributions for the promotion subject to constraints. We begin by revisiting the mechanism design problem with given distribution of values, before addressing the value design problem.

[9] With a slight abuse of notation, we use the same operator to denote the allocation rule and the expected (or interim) allocation rule, but it should be understood that $x_i(v_i) = \int_{\alpha_j}^{\omega_j} x_i(\mathbf{v}) dF_j(v_j)$ for $j \neq i$. Similarly for the effort rule, $e_i(v_i) = \int_{\alpha_j}^{\omega_j} e_i(\mathbf{v}) dF_j(v_j)$ for $j \neq i$.

Classical Mechanism Design Seminal results in Myerson (1981) characterise total effort in the employer's optimal mechanism when values are independent and private, and distributions are given. By the revelation principle, it is without loss for the employer to restrict attention to incentive compatible and individually rational direct mechanisms in which workers reveal their type truthfully.[10] Thus, the employer sets for all possible profiles of values $\mathbf{v} \in V_A \times V_B$ an effort rule $e(\mathbf{v})$ and an allocation rule $x(\mathbf{v})$ in order to maximise total effort subject to incentive compatibility and individual rationality constraints expressed in equation (1). The revenue equivalence principle for independent private value settings implies that in any incentive compatible mechanism, the effort rule is fully pinned down by the allocation rule when individual rationality holds with equality for the lowest value type of each worker.

The characterisation of the optimal mechanism relies on the notion of *virtual valuation*, which identifies the marginal contribution of worker i with value v_i to total effort. Formally, at any v_i at which the cumulative F_i is differentiable, the virtual valuation is defined as

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

where $f_i = F_i'$ denotes the density. We refer to the difference between the value and virtual value as the *information rent*.

Relatedly, for any profile of values $\mathbf{v} \in V_A \times V_B$, the *virtual surplus* of an incentive compatible mechanism can be defined as the expected virtual valuation of the worker being awarded the promotion,

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

Total effort in an incentive compatible mechanism, in which the lowest type of the worker is indifferent about participating, amounts to expected virtual surplus,

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

which cannot exceed *surplus*, $\mathbb{E}_{\mathbf{v}}[\max\{v_A, v_B\}]$. The optimal mechanism is therefore a direct mechanism maximising expected virtual surplus subject to (i) incentive compatibility, which implies that the expected probability of being promoted, $x_i(v_i)$, is increasing in the value v_i for any worker i and (ii) individual rationality, which pins down the expected effort for the lowest value of any worker i at $e_i(\alpha_i) = \alpha_i x_i(\alpha_i)$. The optimal mechanism can display both exclusion, meaning that the promotion may not be awarded to any worker despite values being positive, and inequity, meaning that the worker with the highest value may not be promoted even when the promotion is awarded. When the virtual values are increasing for all workers, the optimal mechanism always awards the promotion with certainty to the worker with the highest

[10]The revelation principle for these settings states that for any mechanism and any Bayes Nash equilibrium of that mechanism, there exists a Bayes Nash equilibrium in a corresponding direct mechanism with the same outcomes and in which workers reveal their value truthfully.

non-negative virtual valuation. However, more complicated allocation rules need to be devised to fulfil incentive compatibility in more general settings. For any pair of marginal distributions (F_A, F_B), denote total effort at the associated optimal mechanism by $TE\,(F_A, F_B)$.

To establish our result when cumulative distributions are not continuously differentiable on their support, it is convenient to translate the problem to the *quantile space*, as pioneered in Bulow and Roberts (1989).[11] 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 with 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 or contains atoms.[12] At quantiles at which v_i is differentiable, the virtual value can then be restated as

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

At quantiles at which v_i is not differentiable, the virtual value is bounded above by the value and is always equal to v_i at the highest value in the support. The expected promotion probability for worker i in the quantile space can similarly be restated as $y_i(q) = x_i(v_i(q))$ and must be non-increasing by incentive compatibility.

Value Design Problem Having provided a partial characterisation of the optimal mechanism for given value distributions, we now turn to the design of value distributions. To keep the problem compact, we posit throughout that the maximal value for the promotion is bounded and equal to $\overline{\omega} < \infty$. We refer to any pair of distributions (F_A, F_B) with support in $[0, \overline{\omega}]$ as a *value design*. Denoting by $y_A(q)$ and $y_B(q)$ the expected allocation probabilities in the optimal mechanism for given a value design, (F_A, F_B), the employer's value design problem in the quantile space amounts to

$$\begin{aligned}\max_{F_A, F_B} \quad & TE\,(F_A, F_B) = \mathbb{E}_q\left[\phi_A(q)y_A(q)\right] + \mathbb{E}_q\left[\phi_B(q)y_B(q)\right] \\ & \text{s.t. Distributional Constraints}\end{aligned} \tag{6}$$

A pair of distributions (F_A, F_B) is said to be an *optimal value design* if it maximises total effort subject to some distributional constraints.

These constraints can be parsed in two classes: individual design constraints and joint design constraints. The former class looks at the possibility of influencing values for each worker individually. The latter class considers design constraints that allow for value reallocation across workers and shows that discrimination is a natural feature of such settings.

We analyse two *individual design* problems. We allow for any possible design of value distributions with expected value below some constant $k < \overline{\omega}$. Formally, the constraint allows the

[11] See Hartline (2013) for a recent comprehensive survey.

[12] When a cumulative distribution is continuous and strictly increasing, $v_i(q) = F_i^{-1}(1-q)$.

employer to design any distribution of values F_i for player i satisfying

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

where $\mathbb{E}_{F_i}[v]$ denotes the expected value of the distribution F_i. The key insights for this design problem will be showcased in a setting in which 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 subject to this constraint. Constraint (7) encompasses the set of distributions that are second order stochastically dominated by some bounding distribution G with $\mathbb{E}_G[v] = k$. The second natural constraint considered in this class is first order stochastic dominance which allows the employer to design any distribution of values F_i for player i satisfying

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

In this setting the cumulative distribution G provides a point-wise upper-bound on the distributions that the employer can choose to design.

We further analyse two *joint design* problems. These capture settings in which value can be reallocated across workers, meaning that enhancing the distribution of values of one worker comes at the cost of deteriorating the distribution of values for the other worker.

Mirroring constraint (7) in the individual design problem, we first allow the employer to generate any two distributions for which the sum of expected values across (F_A, F_B) is below some constant $k < \overline{\omega}$. Formally, the constraint allows the employer to design any two distributions of values satisfying

$$\mathbb{E}_{F_A}[v] + \mathbb{E}_{F_B}[v] \leq k. \tag{9}$$

Such a constraint gives substantial freedom to the employer. However, there may be distributional bounds. We therefore take an arbitrary measure with mass two and an associated cumulative distribution $H(v)$, and ask how the employer would split such a measure to create the two value distributions for the workers.[13] Formally, the constraint allows the employer to design any two value distributions satisfying

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

3.3 Discussion of Modelling Assumptions

Induced Valuations In our modelling interpretation, the employer can directly affect a worker's value for the promotion (either through the compensation package conditional on promotion or through workplace culture). Formally, the induced value can be described as

[13] Formally, let $\mathcal{H}$ denote the measure and define $H(v) = \mathcal{H}(V \leq v)$.

$v = \gamma(\nu, \epsilon, \zeta)$, where ν is the intrinsic valuation for the promotion driven by their skills and preferences, while ϵ describes the valuation for the compensation package as well as features of the new job after promotion. Last, the costs and benefits of dealing with a certain workplace culture are given by ζ. We allow for the valuation for the promotion package as well as workplace culture to depend on intrinsic, unobservable characteristics of the employer. Therefore, the exact value for the promotion remains private, even though the firm may be able to generate more information about the worker's distribution. If more able workers value the promotion more and have an easier time of fitting in, then the real private information of a worker is their quantile in the value distribution. In this case, the quantile showcases the extent to which the worker is willing to contribute to effort provision relative to other workers. The employer's design of values then transforms such an implicit motive into an explicit willingness to contribute effort.

Our modelling approach does not impose any difference across workers, in that the quantiles for both workers are simply uniformly distributed and asks which designs are optimal for the employer.

Effort We assume that effort is observable, in line with modern technology that has improved monitoring of workers. Alternatively, effort can be interpreted as some measurable output. In this case, the model requires workers' efforts to map linearly and separably into such an output measure. However, our approach would remain valid even if costs of exerting effort were convex, see Greenwald, Oyakawa, and Syrgkanis (2017). Additionally, effort can be interpreted as additional effort contributions on top of the minimal effort needed to retain the job.

Individual versus Joint Distribution Design We focus on a set of abstract and general constraints. It seems natural to assume that the employer cannot increase the value infinitely, that there is cap. Such a cap also seems reasonable in terms of expectations: workers, in expectation, do not value a promotion infinitely. However, by only restricting the employer to limit the average, we may be giving him too much freedom. This motivates a pointwise restriction of the value– captured by first order stochastic dominance. Given our constraints at the individual level, we once again allow for a cap on expectations and a pointwise value restriction in the form of some distribution. The twist in the joint problem is that an increase in valuation for one worker is matched by a decrease for the other worker. Note that any policy that increases the valuation for both workers is optimal. Our paper asks what happens if all these policies have been implemented and the remaining choices lead to lower values for one worker and not for the other.

Overall, our constraints provide a full characterisation of optimal designs: we show the best distribution if the employer is free to do as he pleases, if he faces restrictions on the average values or if he is even more constrained by certain functional shapes. We document what happens if he must focus on each worker individually, or whether he is able to jointly affect outcomes.

While our approach is abstract, it turns out that our constraints carry meaning in the workplace, see Section 5. There, we confirm the predictions generated by our model.

Implementation For the employment application analysed here, it is best to consider an all-pay implementation in which worker i's effort when their value v_i simply amounts to the expected effort in the optimal mechanism, $e(v_i)$, irrespective of whether they are awarded the promotion. The employer can offer a schedule of allocation probabilities based on the effort exerted. Whenever the value distributions are asymmetric, it may not be the case that the worker with the highest valuation or effort will be promoted with certainty. Workers may need to be compensated differentially to get them to exert the desired effort and thereby reveal their private information, which can lead to a worker with lower valuation and effort receiving the promotion. Further, it may well be the case that no worker is promoted when effort contributions are insufficient–a feature common in the workplace.

4 Optimal Value Design

We begin by characterising the optimal distribution when no constraints are in place, and then analyse the individual and joint design problems introduced in the previous section. All proofs of propositions and corollaries can be found in Appendix C. The proofs of remarks are omitted as they follow immediately from the other results.

4.1 Designing Value without Constraints

Suppose the employer can design any value distributions with support in $[0, \overline{\omega}]$. Then the employer's optimal designs are the ones in which at least one worker values the promotion at $\overline{\omega}$ with certainty. This value design is optimal as the surplus cannot exceed the maximal value for the promotion $\overline{\omega}$. If a worker has such value with certainty the employer can simply promote that worker for sure while asking them to exert effort $\overline{\omega}$, extracting the maximal surplus.

Proposition 1. *In any optimal value design, there is at least one worker whose value distribution amounts to*

$$F^{NC}(v) = \begin{cases} 0 & \textit{if } v < \overline{\omega} \\ 1 & \textit{if } v = \overline{\omega} \end{cases} \tag{11}$$

The result highlights two forces at play throughout the analysis. First, the employer would like to increase total surplus, or equivalently the maximal value for the promotion, as much as possible, because a higher value induces workers to exert higher effort.

Second, 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.

It suffices to increase the value for one worker to $\overline{\omega}$, as 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.

4.2 Individually Optimal Value Design

We turn our attention to optimal value distributions for the individual design problems set out earlier. We begin by considering the constraint (7), which simply bounds the mean of the possible distributions by a constant $k < \overline{\omega}$. In individual design problems, it is instructive to initially fix a distribution for worker A satisfying the constraint and focus on the adjustment of the value distribution for worker B.

Our first key insight establishes that the optimal design of values for worker B is a two atom distribution with mass at the highest possible value of the distribution, $\overline{\omega}$, and at zero, and with expected value equal to the upper-bound k.[14]

Proposition 2. *An optimal value design among all designs satisfying $\mathbb{E}_{F_B}[v] \leq k$ is given by*

$$F^*(v) = \begin{cases} 1 - \frac{k}{\overline{\omega}} & \textit{if } \; v < \overline{\omega} \\ 1 & \textit{if } \; v = \overline{\omega} \end{cases} \tag{12}$$

Such a design is uniquely optimal if $F_A \neq F^*$. When $F_A = F^*$, then a design is optimal if and only if for some $K \in [k, \overline{\omega}]$ it satisfies

$$F^K(v) = \begin{cases} 1 - \frac{k}{K} & \textit{if } \; v < K \\ 1 & \textit{if } \; v = K \end{cases} \tag{13}$$

Before providing some intuition for the result, let us briefly comment on its content. The result establishes that the optimal design is one in which the employer does not promote or reward worker B unless they contribute the maximal possible effort, $\overline{\omega}$. In that event the employer rewards the worker with the highest possible induced value, namely $\overline{\omega}$. The worker does not benefit from this high compensation because the surplus of worker B is equal to zero irrespective of their quantile or type. This is in line with compensation schemes observed in sectors such as finance, law and consulting, which give huge benefits to those promoted. However, all workers suffer being kept at their reservation utility through excessive effort contributions at these high stakes jobs.[15] The result also implies that quite dispersed effort contributions will be observed

[14]This result is evocative of the perfect information case in information design, where the posterior jumps to either zero or one.

[15]See for example, Empson (2017), for a description of the working conditions in these environments.

among workers vying for the same promotion.

From a mechanism design perspective, the result is perhaps surprising as it demonstrates that the optimal design is one that maximises variance (the next corollary makes this point explicit). The employer minimises the dispersion in valuations, and consequently the information rent, but only for those types of the worker it intends to promote with positive probability. The employer does not care about workers who never get promoted and are thus excluded, since it leaves no information rent to such workers.

As to the intuition for the proof of the result, note that employer has once again two, possibly conflicting, motives: maximising surplus and reducing information rents. In the proposed design problem, both motives are however aligned. The first motive is driven by the employer's desire to design a promotion which is highly valued, in order to extract more surplus from the worker through their effort contribution. This aim is met by having the mean constraint bind at the optimal design, so that $\mathbb{E}_{F^*}[v] = k$.[16] The second motive is driven by the employer's desire to make the worker's valuation easily identifiable to minimise the information rent paid to workers for truthfully revealing their type. As the value of player B is deterministic conditional on ever being promoted, no information rent is left to the worker. As the employer does not benefit by leaving any value to types that are not promoted, the lower atom must be at zero. This establishes that a bimodal distribution is optimal, with one atom at a strictly positive value and another zero.

Why does the strictly positive value equal $\overline{\omega}$? To see this, note that a change in B's 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 B's distribution at k, a higher value implies a lower probability that the value is strictly 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 $\overline{\omega}$.

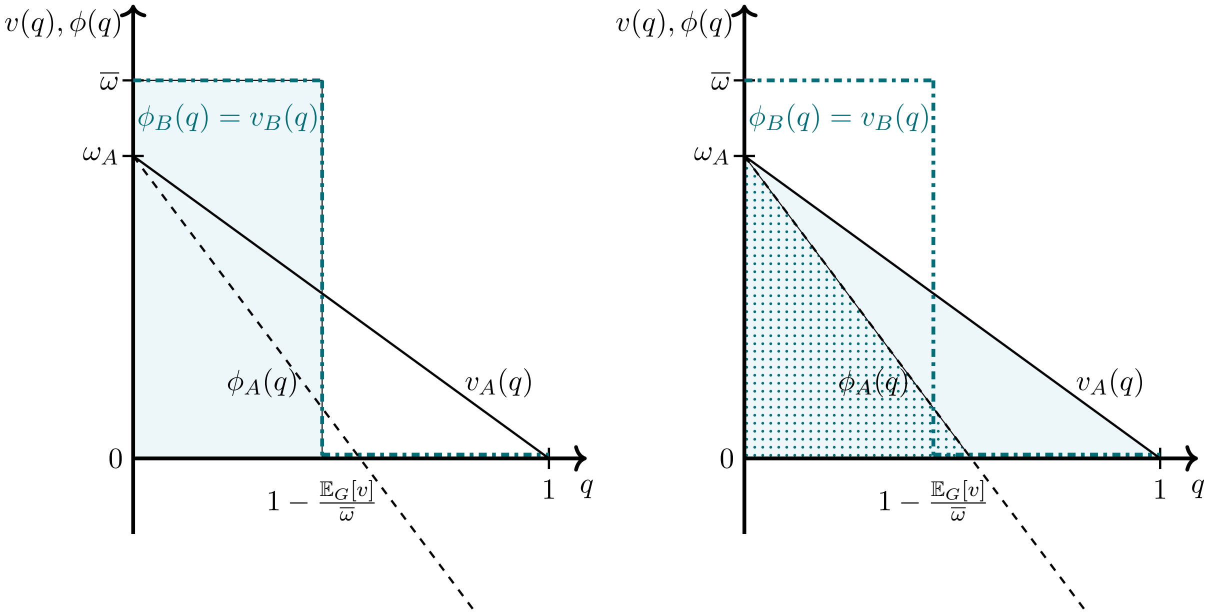
This argument yields a unique solution as long as F_A differs from F.* If F_A equals F,* then a multiplicity of designs yield the same total effort: any distribution, which assigns mass to exactly one K with $K \in [k, \overline{\omega}]$ and matches k in expectation is optimal. If A's valuation is $\overline{\omega}$, then the firm always awards him the promotion. In expectation, the firm extracts k from him. If A's valuation is zero, worker B attains the promotion. Again, the firm extracts k in expectation as there is no information rent: the exact value of K is irrelevant.

The intuition behind Proposition 2 is summarised graphically in Figure 3. The adjustment in distribution allows the employer to extract the entire expected value for the promotion from worker B, since the expected virtual value coincides with the expected value which is depicted

[16] To see that the expected value remains constant, note that $\mathbb{E}_{F^*}[v] = \frac{k}{\overline{\omega}}\overline{\omega} + \left(1 - \frac{k}{\overline{\omega}}\right) 0 = k$.

by the shaded area in Figure 3a. This is more than what the employer can possibly extract from worker A to whom we assign some alternative continuous distribution with expected value k. Worker A's expected virtual value, the dotted area in Figure 3b, is smaller than the expected value of F_A, the shaded area under $v_A(q)$, which by construction is no larger than k and thus no larger than the expected virtual value of worker B.

Figure 3: 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 dotted area in Figure 3a, which equals k, while A's expected virtual value is given by the dotted area in Figure 3b. As the shaded area under $v_A(q)$ equals k, it follows that A's expected virtual value is smaller than B's. The example posits that $\alpha = 0$.

It is worth noting that Proposition 2 also establishes that F^* is optimal design relative to any other distribution that is second order stochastically dominated by a distribution G with mean $\mathbb{E}_G[v] = k$. This follows by two insights. First, any distribution F_B that is second order stochastically dominated by G satisfies

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

and therefore has a lower mean than G, thereby fulfilling the constraint (7). Second, F^* is second order stochastically dominated by G. Thus, F^* is also the optimal distribution among all F_B satisfying constraint in (14).

Corollary 2.1. *The distribution F^* is an optimal value design for worker B among all distributions satisfying (14). It is uniquely optimal if $F_A \neq F^*$.*

This implies that the employer always prefers a "riskier" value distribution for worker B.

Proposition 2 shows that it is sub-optimal to reduce the expected valuation for the promotion.

This raises the question of whether the employer would ever be willing to adjust the value distribution for a worker if this necessarily came at the expense of a reduction in the expected value. The next remark exploits the proof of Proposition 2 to show that this is indeed the case, as long as the reduction in expected value is sufficiently small.

Remark 1. *Consider any value design* (F_A, F_B)*. For any* $m \geq \mathbb{E}_{F_B}[\max\{\psi_B(v), 0\}]$*, the value design* (F_A, F^+) *where*

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

increases total effort compared to the initial value design (F_A, F_B)*.*

The condition implies that spreading value is optimal for the employer as long as the reduction in expected value, $\mathbb{E}_{F_B}[v] - m$, is more than compensated by the reduction in information rent, $\mathbb{E}_{F_B}[v] - \mathbb{E}_{F_B}[\max\{\psi_B(v), 0\}]$. 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 only the distribution of one worker can be adjusted, we now allow for the values of both workers to be influenced subject to the constraint requiring the mean of the new distributions not to exceed $\mathbb{E}_G[v]$. The next remark establishes that it is still optimal to adjust the value of one worker to the distribution described in Proposition 2. But for the second worker two designs can be optimal.

Remark 2. *The optimal value designs,* (F_A^*, F_B^*)*, among all designs satisfying* $\mathbb{E}_{F_i}[v] \leq k$ *for all* $i \in \{A, B\}$ *are such that: for one worker* $i \in \{A, B\}$*,* $F_i^*(v) = F^*(v)$*; while for the other worker* $j \neq i$ $F_j^*(v) = F^K(v)$ *for some* $K \in [k, \overline{\omega}]$*.*

When both value distributions can be designed, the result matches Proposition 2 where A's distribution is $F.^*$ One worker displays the most extreme bi-modal distribution, while for the second worker it merely matters that the information rent is reduced to zero. However, there can never be more than one worker with value K for the promotion, as the employer would then just be able to extract k from all workers, whereas by introducing variance total effort increases and converges to $2k$ as $\overline{\omega}$ diverges to infinity. With $n > 2$ workers, the firm creates an environment where all workers (potentially with the exception of one) have the high variance distribution F^*.[17]

In such a design, when workers have high value, they exert the highest possible effort. But with a large probability, workers have low value and do not aim for promotion; instead, they exert zero effort. This translates into a work environment, with few competitive employees who work incredibly hard striving to be promoted, while the remaining workers exert no effort and give up at their chance to be promoted. The designs lead to a high variance in observed effort contributions and are exploitative to workers, leaving all of them without any surplus. However,

[17] As in the two player case, the employer designs the high variance distribution F^* for $n-1$ workers. The remaining worker's distribution equals $F^K(v)$.

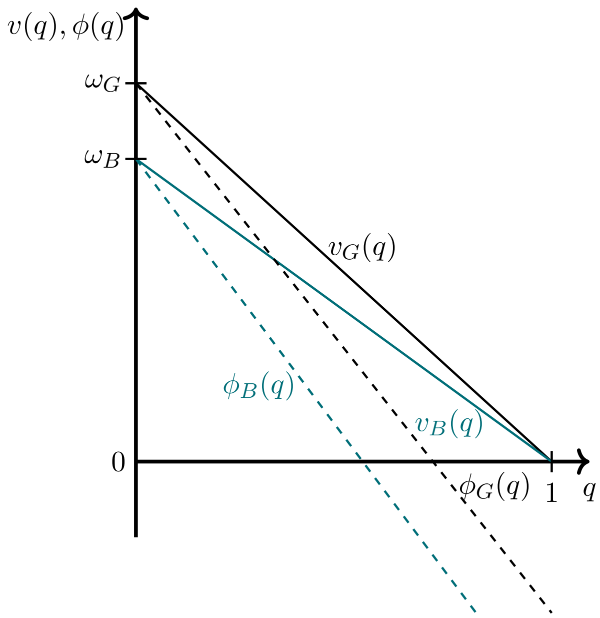
it allows the firm to secure the highest possible total effort which converges to nk as $\overline{\omega}$ diverges to infinity.

Next, we turn to the other individual value design problem, namely the one requiring the designed value distribution to be first order stochastically dominated by an arbitrary distribution G. As in the previous design problem, we begin by fixing the value distribution of worker A and characterising the optimal design for worker B subject to $G(v) \leq F_B(v)$ for all v. First order stochastic dominance implies a ranking of quantiles $q_G(v)$ and $q_B(v)$:

$$q_G(v) = 1 - G(v) \geq 1 - F_B(v) = q_B(v) \tag{16}$$

It follows that for any given quantile, the value under distribution G is higher than under any other distribution that is first order stochastically dominated by G, formally $v_G(q) \geq v_B(q)$ for any $q \in [0, 1]$. This observation is summarised in Figure 4. This observation turns out to be

Figure 4: G FOSD F_B



instrumental in proving the next result.

Proposition 3. *The optimal value design, F_B^*, among all designs satisfying $F_B(v) \geq G(v)$ for all $v \in [0, \overline{\omega}]$ is $F_B^* = G$.*

If the employer can only design distributions that are first order stochastically dominated by G, worker B will be assigned the upper-bound G in an optimal design. In this design problem, any other design might reduce information rents by making the worker's value more predictable, but this would come at the expense of reducing values and thus surplus. The result establishes that the gain in information rent never outweighs the loss in value.

To show this, we consider a distribution F_B that differs from G and is first order stochastically dominated by G. If so, it is possible to construct an alternative distribution for B, $\hat{F}_B$, which is first order stochastically dominated by G, but first order stochastically dominates F_B, with strict inequality for some v. The proof establishes that total effort is higher under $\hat{F}_B$. A change in distribution has two effects, it affects the virtual value for the worker whose value is being designed and it affects the probability of promotion for *both* workers. Our proof strategy fixes the allocation rule to the one that is optimal for value design (F_A, F_B) and shows that the total effort increases when B's distribution is transformed to $\hat{F}_B$ due to the adjustment of virtual values. Formally, the key insight of the proof establishes the second inequality in the following expression

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

The first inequality holds since total effort under the old allocation rule, associated with value design (F_A, F_B), and new virtual value, derived from $\hat{F}_B$, is a lower bound on total effort obtained from the optimal allocation rule for the design $(F_A, \hat{F}_B)$. This follows as an adjustment of the allocation rule cannot decrease total effort, or else the firm would select the old 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 $\phi(q) = \frac{\partial(v(q)q)}{\partial q}$. First order stochastic dominance implies that the value associated to each quantile under distribution $\hat{F}_B$ exceeds the value under distribution F_B, see also Figure 4. As such a distribution $\hat{F}_B$ can be constructed for any distribution F_B which is first order stochastically dominated by G, it follows that the optimal value design for worker B is in fact G.

The result also immediately implies that when both workers' distributions can be designed the employer will find the design (G, G) to be optimal.

Remark 3. *The optimal value design, (F_A^*, F_B^*), among all designs satisfying $F_i(v) \geq G(v)$ for all $v \in [0, \overline{\omega}]$ and all $i \in \{A, B\}$ is $(F_A^*, F_B^*) = (G, G)$.*

Results on first order stochastic dominance show that destroying value relative to some upper-bound G can never be optimal as losses in surplus are not compensated by gains in information rents.

Our contributions on individual value design establish that whereas destroying value is never optimal for the employer, spreading value always is. In particular, the employer will seek to create a bi-modal value distributions in which workers either value the promotion very highly or not at all. This is in line with the observed effort contributions in competitive environments such as consulting, law and finance where some workers value a promotion to partner or manager considerably and are willing to work incredibly hard for it, while the rest dislike this culture and opt out of the promotion race.

4.3 Jointly Optimal Value Design

Next we turn to settings where the employer designs the values of their workers jointly. As highlighted in the individual design setting, if the employer can implement a promotion policy that increases the value of all workers, then he will do so. We therefore focus here on designing values for promotion that benefit one worker at the cost of disadvantaging the other worker. Both design problems considered here are ones in which an increase in valuations for one worker must be compensated by a decrease in valuations for the other worker. That is, value is reallocated.

The first design problem posits that any design (F_A, F_B) is feasible as long as the sum of expected values does not exceed a given bound, $\mathbb{E}_{F_A}[v] + \mathbb{E}_{F_B}[v] \leq k < \overline{\omega}$. Under this constraint in any optimal design, the firm allocates all the value to only one worker, say A, and eliminates the information rent for that worker by choosing designs that either assign value k with certainty; or assign value only to 0 and to some $K \in (k, \overline{\omega}]$ and that have k as a mean.

Remark 4. *In any optimal value design, (F_A^*, F_B^*), among all designs satisfying $\mathbb{E}_{F_A}[v] + \mathbb{E}_{F_B}[v] \leq k$, the employer sets $F_i^*(0) = 1$ for some $i \in \{A, B\}$ and sets*

$$F_j^*(v) = \begin{cases} 1 - \frac{k}{K} & \text{if } v < K \\ 1 & \text{if } v \geq K \end{cases} \tag{18}$$

for $j \neq i$ and for some $K \geq k$.

The intuition follows from insights developed in the benchmark case and in the individually optimal design if the mean of the designed distributions is bounded. As in the benchmark case, the firm only needs one worker to have a positive valuation and therefore selects worker j to receive all expected value, while the other worker's expected value is reduced to zero. As in the individual design case, the firm can opt for designs with different volatility provided it can extract the full surplus in any of these. Consequently, an employer facing two potentially identical workers will induce differences in how they value the promotion, for instance by creating a workplace culture that is supportive of only one worker.

This result is stark and it raises questions as to whether discrimination between workers is generally optimal if values can be reallocated. To address this, we allow for any design (F_A, F_B) to be feasible as long as $F_A(v) + F_B(v) = H(v)$ for all v, where $H(v)$ is the cumulative associated to an arbitrary measure with mass *two*, with $v \in [\alpha, \omega]$. It turns out that in this setting discrimination remains optimal as long as the measure H assigns mass to more then one value – a symmetric design cannot be optimal. Rather, the employer maximises total effort by creating two maximally differentiated value distributions. To state the result, it is useful to define the median for the measure H, as the value v^M such that $\int_\alpha^{v^M} dH(v) = 1$, when such value exists. It may not exist if H is atomic, in which case the median is instead defined as the smallest value v^M such that $\int_\alpha^{v^M} dH(v) \geq 1$.

Proposition 4. *The optimal value design, (F_A^*, F_B^*), among all designs satisfying $F_A(v) +$*

$F_B(v) = H(v)$ *for all* $v \in [\alpha, \omega]$ *is one in which for* $i \neq j$

$$F_i^*(v) = H(v) - 1 \qquad \text{if } v \in [v^M, \omega], \tag{19}$$

$$F_j^*(v) = H(v) \qquad \text{if } v \in [\alpha, v^M] \tag{20}$$

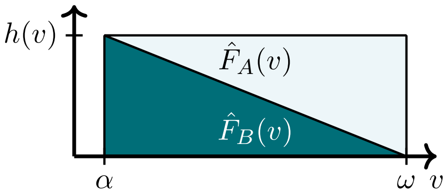
Maximal discrimination, with one worker's values below the median of H and the other worker's values above the median of H, yields a higher total effort than any other design. Such a design has two benefits. First, it maximises surplus, $\mathbb{E}[\max\{v_A, v_B\}]$, since it generates the highest possible expected value for one of the workers. This is better than any other design, which necessarily yields more similar expected values, due to the convexity of the maximum operator.

Second, this adjustment narrows the support of both workers' distributions, and thereby information rents. Such a design increases the probability of receiving the promotion for the worker with high values, inducing them to work harder, and decreases the probability of receiving the promotion for the low value worker, inducing them to work less. In spite of the decreased competition, the gains in effort provision from the high value worker more than compensate the losses from the low value worker. The key takeaway of this design problem is that when value redistribution is possible, the employer benefits from designs that treat workers differently and discriminates between them. This provides a novel rationale for regulations aiming to control the extent to which workers can receive differential compensations or working conditions on grounds of equity as well as efficiency. In particular, the optimality of asymmetric designs hints at why policy makers may want to regulate on-the-job discrimination even in contexts in which the employer has no discriminatory motive and simply aims to maximise workers' effort contributions.

To establish the optimality of maximal discrimination, the proof first posits that H is differentiable and compares maximal discrimination to any other design (F_A, F_B) in which the supports of distributions F_A and F_B match the support of H. The proof then compares maximal discrimination to general designs thereby allowing for disjoint supports and atomic measures H. These comparisons are illustrated in Figure 5: with an example in which the supports of the designed distributions coincide with the support of H illustrated in Figure 5a; and with an example in which the supports of the designed distributions are disjoint illustrated in Figure 5b. To establish the result, it suffices to show that any design belonging to these two classes of distributions does not lead to a higher total effort compared to maximal discrimination, because any other design simply amounts to linear combination of two designs belonging respectively to each of these two classes.

Next, we provide some intuition for the proof of the result for the case in which H is differentiable and the supports of the designed distributions match the support of H. To capture an arbitrary split of the density $h(v) = H'(v)$, consider any design $(\hat{F}_A, \hat{F}_B)$ which assigns a share $a_A(v) \in (0, 1)$ of the density $h(v)$ to $\hat{F}_A$, while allocating the remaining share $a_B(v) = 1 - a_A(v)$ to $\hat{F}_B$ and derive allocation probabilities, $\hat{x}_A(v)$ and $\hat{x}_B(v)$, for such a setting. Now, we compare the total effort generated by these arbitrary distributions that split the density and maximal

Figure 5: Alternative Distributions



(a) Same Support

(b) Disjoint Support

discrimination. This is challenging as we need to keep track of the differences in virtual values as well as allocation probabilities.

It turns out to be useful to rank the allocation probabilities $\hat{x}_A(v)$ and $\hat{x}_B(v)$ for each v according to magnitude. If the value is above the median, then the maximum over $\hat{x}_A(v)$ and $\hat{x}_B(v)$ is assigned to $\overline{x}_A(v)$ and the minimum becomes $\overline{x}_B(v)$. If the value is below the median, then the maximal allocation probability is absorbed by $\overline{x}_B(v)$, while the minimum is now $\overline{x}_A(v)$. This implies that we now know the ranking of the allocation probabilities.

Based on this, we can the show that total effort under maximal discrimination is higher than total effort under any alternative design by comparing

$$\begin{aligned}\underbrace{\mathbb{E}[\psi_A^*(v)x_A^*(v)] + \mathbb{E}[\psi_B^*(v)x_B^*(v)]}_{\text{Distribution:}F_A^*,F_B^*,\ \text{Allocation:}x_A^*,x_B^*} &\geq \underbrace{\mathbb{E}[\psi_A^*(v)\overline{x}_A(v)] + \mathbb{E}[\psi_B^*(v)\overline{x}_B(v)]}_{\text{Distribution:}F_A^*,F_B^*,\ \text{Allocation:}\overline{x}_A,\overline{x}_B} \\ &\geq \underbrace{\mathbb{E}[\overline{\psi}_A(v)\overline{x}_A(v)] + \mathbb{E}[\overline{\psi}_B(v)\overline{x}_B(v)]}_{\text{Distribution:}\overline{F}_A,\overline{F}_B,\ \text{Allocation:}\overline{x}_A,\overline{x}_B} \\ &= \underbrace{\mathbb{E}[\hat{\psi}_A(v)\hat{x}_A(v)] + \mathbb{E}[\hat{\psi}_B(v)\hat{x}_B(v)]}_{\text{Distribution:}\hat{F}_A,\hat{F}_B,\ \text{Allocation:}\hat{x}_A,\hat{x}_B},\end{aligned} \tag{21}$$

where $\overline{\psi}$ is the virtual value associated with $\bar{x}$. The first inequality follows again by replacing the optimal allocation probabilities under maximal discrimination with the constructed allocation probabilities, $\overline{x}$. The allocation probabilities x^* must be associated with a weakly higher total effort – otherwise they would not be optimal. However, it is not necessarily the case that $\overline{x}_A(v)$ and $\overline{x}_B(v)$ are feasible allocation probabilities under the new distributions. In Appendix C, we verify that these expected allocation probabilities are indeed interim feasible (Border (1991)) and can be be derived by taking expectations over a well-defined allocation rule, thus validating our approach.

The key to establish the result then remains the second inequality in expression (21) since the last equality holds by definition as we only relabelled allocation probabilities and associated virtual values. Rewriting the relevant inequality from (21), yields

$$\begin{aligned}&\int_{v^M}^{\omega} \psi_A^*(v)\overline{x}_A(v)h(v)dv + \int_{\alpha}^{v^M} \psi_B^*(v)\overline{x}_B(v)h(v)dv \\ &\quad > \int_{\alpha}^{\omega} \left(a_A(v)\overline{\psi}_A(v)\overline{x}_A(v) + a_B(v)\overline{\psi}_B(v)\overline{x}_B(v)\right) h(v)dv.\end{aligned} \tag{22}$$

Observe that the density for each v at each integrand is identical. This allows us to compare integrands pointwise for each v. Further, we assume for now that virtual values are positive, which yields

$$\begin{aligned}&\psi_i^*(v)\overline{x}_i(v) - a_A(v)\overline{\psi}_A(v)\overline{x}_A(v) - a_B(v)\overline{\psi}_B(v)\overline{x}_B(v)\\&\geq \left(\psi_i^*(v) - a_A(v)\overline{\psi}_A(v) - a_B(v))\overline{\psi}_B(v)\right)\overline{x}_i(v),\end{aligned} \tag{23}$$

where the inequality follows from the ranking of the allocation probabilities. For $v > v^M$, $\overline{x}_A(v) > \overline{x}_B(v)$ and so we can replace $\overline{x}_B(v)$ by $\overline{x}_A(v)$. Similarly, for $v < v^M$, $\overline{x}_A(v) < \overline{x}_B(v)$, allowing us to replace $\overline{x}_A(v)$ with $\overline{x}_B(v)$. This simplifies the problem to comparing virtual values under maximal discrimination to the average auxiliary virtual value. If the difference in virtual values is positive for both workers and all values v, total effort is higher under maximal discrimination compared the chosen design. 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 gains in marginal effort contributions occur for values below the median. Above the median, the sum of information rents paid to workers coincide for any design chosen by the employer. But for values below the median, the sum of information rents paid to workers is minimised by assigning all such value to a single worker as this maximises at once both $F(v)$ and $f(v)$ for the worker to whom such values are assigned.

If the virtual values are not positive, then more involved arguments are necessary, which are provided in Appendix C.

We follow a similar approach to establish that designs with disjoint supports that differ from maximal discrimination are sub-optimal. This follows because maximal discrimination minimises the information rent paid to workers with values below the median at the cost of increasing the information rent for workers above the median. For any other value design, the gain in information rent for higher values cannot make up for the loss at lower values – because the information rent is highest when values are low.

With disjoint supports, we also allow for atoms and as such we start out tackling the problem in the quantile space. This approach is particularly appealing as there we do not have to work with virtual values, but rather values as the mapping between the two is straightforward. We therefore specify total effort in the quantile-value space. However, given the nature of the constraint, the problem becomes simpler once we flip it into the value-quantile space. Using the same strategy as with split densities, we replace the allocation probabilities under maximal discrimination with those of some arbitrary designs, which creates a lower bound on total effort with maximal discrimination. We then compare this lower bound to different cases of how the measure can be split, proving for each of them that maximal discrimination is associated with a lower information rent, which makes it optimal.

Our proof documents that maximal discrimination increases virtual values as it minimizes information rent. This increase in virtual values does not necessarily translate into a higher effort. In fact, as we showed in the Example in Section 2, the effort for low value workers will decrease as the value distributions become more unequal. Effort depends on both virtual values as well as the allocation probabilities, and the latter have deteriorated for low values. This ultimately triggers the reduction in effort for low values. However, our proof highlights that the inequality in promotion probabilities can be ignored. Recall that the expected virtual surplus is given by

$$\mathbb{E}[\psi_A(v)x_A(\mathbf{v})] + \mathbb{E}[\psi_B(v)x_B(\mathbf{v})] \tag{24}$$

Unequal distributions which favor A at the expense of B imply that x_A increases, while x_B decreases. Crucially, this is beneficial, as A's virtual values are higher than B's, again due to the distribution favouring A. Therefore, maximal discrimination allocates higher allocation probabilities to higher values, while matching low virtual values and allocation probabilities. This is another force as to why maximal discrimination is optimal, and the reason why we can just fix the allocation probabilities: allowing them to adjust would only strengthen the result. In practice, this implies that the employer is better off by facing two workers, where one exerts high effort and the second low effort, instead of two medium effort workers.

5 Empirical Application

Our theoretical model yields various predictions that can be tested using the "After the JD" data set.[18] The American Bar Foundation surveyed a nationally representative sample of US lawyers first admitted to the bar in 2000. The first survey was conducted in 2002, with two follow up surveys in 2007 and 2012. The longitudinal nature of the study allows us to track the careers of the lawyers. The data contain information about job features as well as individual information about education, personal characteristics, experiences at the workplace as well as workplace information. We use the data to first document that our modelling approach matches the institutional environment. We then turn to testing the theoretical predictions our model generates.

5.1 A Lawyer's Work Environment

We characterise a lawyer's work environment and connect features of it to our modelling assumptions, establishing that our framework captures this sector well.

Lawyers in private firms generally start out as associates and aim to become partner subsequently. We measure *Associate* as a binary variable, with value one if the respondent held an associate positions in at least one of the three waves. Similarly, *Partner* takes value one if the survey participant was either a non-equity or equity partner in at least one of the waves. Our model assumes that there are fewer partner positions than associate positions.

Assumption 1. *There are fewer partners than associates.*

[18] http://www.americanbarfoundation.org/research/project/118

Table 1: Associate to Partner

Was Associate	Became Partner/Equity Partner					
	No		**Yes**		**Total**	
	No.	Col %	No.	Col %	No.	Col %
No	2769	62.9	314	38.9	3083	59.2
Yes	1633	37.1	494	61.1	2127	40.8
Total	4402	100.0	808	100.0	5210	100.0

Note: The sample comprises of all lawyers that at some point filled out the survey. It therefore represents the entire population of lawyers. Partner is measured as either non-equity or equity partner in any wave, Associate equals one if this individual is an associate in one of the waves.

This pattern indeed emerges for lawyers. Among the lawyers that hold an associate position a minority goes on to become a partner, see Table 1. The ratio of associate to partner positions is 2:1: there were 1633 associates and 808 partners. However, the probability of being promoted to partner is lower as there are numerous lawyers that become partners without having been an associate first: 314 out of 808 are partners without having held an associate position.[19]

In our model, we focus on two workers. These can be interpreted as two groups A and B. In this case, the values are distributed over workers, who are member of one of the groups. To obtain two groups of equal size, we consider men and women. Law is a gender balanced profession, with 48 % female respondents (see Table 8) ensuring similar numbers across groups. Moreover, in our theoretical framework there are no a priori differences between groups. We therefore would like for female and male lawyers to be similar at the beginning of their career.

Assumption 2. *Female and male lawyers are identical at the start of their career.*

We display summary statistics by gender in Table 2. The sample considered consists of all lawyers as we are interested in potential gender differences within the profession. Table 2 highlights that men and women are similar across a number of variables, which we classify as Race and Sexual Orientation, Personal Info, Education, Dedication, Work Culture, and Firm Characteristics. We provide summary statistics for the variables in Appendix D Table 8.

We first consider race. Survey participants are asked what their race/ethnic group is. Multiple mentions are possible. If a respondent identifies with a race then the variable takes value one, otherwise it is zero. Women are more likely to identify as Asian or Black relative to men. In contrast, men tend to be more white. We further consider sexual orientation, which is again a binary variable. It takes value one if the respondent identifies as gay, lesbian, transgender or bisexual, zero otherwise. We do not see any differences in sexual orientation between men and women.

[19]For example, it is possible to work in the public sector and then become partner in a private law firm. There is also a significant number of survey participants who declared they worked for a private firm, but did not specify their position. Additionally, some participants provide information in later waves, but not in Wave I. These respondents may well work for the private sector. We therefore choose to focus on the largest sample.

Table 2: Summary Statistics By Gender

	Men		Women		Difference	
	Mean	S.D.	Mean	S.D.	Men-Women	p-value
Race and Sexual Orientation						
Black	0.06	0.24	0.10	0.30	-0.04***	(0.00)
Hispanic	0.07	0.25	0.06	0.25	0.00	(0.60)
Native American	0.01	0.09	0.01	0.09	0.00	(0.98)
Asian	0.07	0.26	0.10	0.30	-0.02***	(0.01)
White	0.73	0.45	0.66	0.47	0.07***	(0.00)
Other Race	0.03	0.18	0.03	0.17	0.00	(0.54)
LGBT+	0.03	0.17	0.03	0.18	-0.00	(0.86)
Personal Info						
Age	31.88	5.65	31.39	5.75	0.49***	(0.01)
Married	0.60	0.49	0.52	0.50	0.08***	(0.00)
# Kids	0.53	0.94	0.32	0.73	0.21***	(0.00)
Education						
Law S Rank	2.42	1.09	2.47	1.12	-0.05	(0.29)
Law S Quality	3.76	1.28	3.81	1.20	-0.04	(0.26)
Law Review	1.32	0.67	1.32	0.68	-0.00	(0.97)
UGrad GPA	4.96	1.54	5.43	1.33	-0.48***	(0.00)
UGrad Rank	2.94	1.04	3.23	0.84	-0.29***	(0.00)
UGrad Quality	3.63	1.60	3.68	1.58	-0.05	(0.27)
Dedication						
Intend Practice Law	1.78	0.53	1.77	0.54	0.02	(0.26)
Consider Oth. Career	0.42	0.49	0.40	0.49	0.01	(0.65)
Influential	2.96	1.28	2.87	1.31	0.08	(0.10)
Loans	61323.30	41434.27	62823.85	41675.82	-1500.56	(0.25)
Loans []	4.75	2.43	4.85	2.45	-0.10	(0.17)
Firm Characteristics						
Firm Size	240.60	415.49	250.05	419.51	-9.45	(0.54)
Firm Size []	4.98	3.49	5.18	3.43	-0.20	(0.12)
Office Size	84.91	129.70	86.89	124.71	-1.98	(0.68)
Office Size []	3.90	2.64	4.03	2.59	-0.13	(0.18)
% Men	71.32	21.00	58.25	21.70	13.06***	(0.00)
Private Firm	0.55	0.50	0.47	0.50	0.08***	(0.00)
Work Culture						
Demeaning	0.07	0.26	0.22	0.42	-0.15***	(0.00)
Discrimination	0.07	0.25	0.17	0.37	-0.10***	(0.00)
Rec. Time Ass.	0.64	0.48	0.69	0.46	-0.04*	(0.05)
Rec. Time Partner	0.33	0.47	0.29	0.45	0.04*	(0.07)
Breakfast Partner	0.59	0.49	0.49	0.50	0.10***	(0.00)
Observations	2735		2475		5210	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Race variables are indicators if someone identifies with given race, multiple mentions possible. *LGTB+* indicates if someone is not straight. Age is calculated as 2002-birth year. *Married* indicates if someone is married or lives in a domestic partnership. *# Kids* specifies number of kids. All eduction variables are coded such that a higher value is better. *Law School Rank* denotes law school class rank upon completion, *Law School Quality* refers to the quality of the law school. *Law Review* equals zero if someone was not involved, one if involved as member and two if leader. *UGrad GPA* specifies the GPA achieved in undergrad. *UGrad Rank* and *UGrad Quality* are undergrad equivalent of *Law S Rank* and *Law S Quality*, respectively. *Intend Practice Law* refers to whether respondent planned to practice law when entering law school: zero is yes, one not sure, two no. *Consider Other Career* is response to "Did you consider other career in addition or instead of law?" with zero equal to no and one yes. *Influential* is a categorical variable measuring the response to "How important was becoming influential in powerful position in your decision to attend law school?": measured from one to five, with one irrelevant and five very important. *Loans* measures total amount of education-related debt. *Loans []* divides loans into categories, with higher values corresponding to higher loans. *Firm Size* and *Office Size* measure the number of lawyers in the entire private firm and office, respectively. Both sizes are bracketed as well with higher values indicating higher number of lawyers. *% Men* refers to the proportion of lawyers in the workplace that are men. *Private Firm* is an indicator equal to one if the respondent is employed by a private law firm. *Demeaning* is an indictor of value one if the respondent received demeaning comments, *Discrimination* an indicator if the respondent received discriminatory comments. *Recreational Time Partner* indicates if the respondent spent recreational time with partners, *Recreational Time Associates* if the respondent spent recreational time with other associates. *Breakfast Partner* takes value one if the respondent has breakfast with partners.

Male and female lawyers differ in terms of their personal characteristics, namely age, marital status and the number of kids. We measure age as the difference between 2002 and birth year and find that women tend to be younger than men. Marital status is measured as a binary variable, which takes value one if the respondent is married or in a domestic partnership. It is zero if the respondent is single, separated, divorced or widowed. Men are more likely to be married. They further have a higher number of children, as # *Kids* measures the number of children.

Importantly, the survey participants are identical in terms of their law school education. *Law School Rank* denotes the class rank within the law school (a higher value denotes a better rank). *Law School Quality* refers to the ranking of the law school among all law schools; again, a higher value refers to a better school. Male and female law students also do not differ in terms of their participation in the Law Review. This is a categorical variable, which takes value zero if someone did not participate, one if they were a member, and two if they were a leader. There are some discrepancies across gender in terms of undergrad: while both female and male lawyers attended similarly ranked colleges (measured by *UGrad Quality*), the average female student has a higher GPA and a higher class rank in college (again, a higher value corresponds to a better outcome).[20]

We capture dedication to the profession through the following variables: (i) the intent to practice law, (ii) whether they considered a different career during law school, and (iii) whether they would like to become influential in a powerful position, and (iv) students loans as a measure of willingness to pay to become a lawyer and thus, potentially an indicator of dedication. The intent to practice law is measured by the response to "When you entered law school, did you intend to practice law?". It takes value zero if the answer is no, one if the respondent was not sure and two if the answer was yes. The variable *Consider Other Career* is binary and takes value one if they did consider another career. The importance of being influential in a powerful position for the respondent (*Influential*) is measured by a categorical variable that takes value one if being influential is irrelevant. At the other end of the spectrum, five indicates that this is a very important consideration. Last, student loans are continuous. Given that this is a survey, respondents round numbers, so we use bracketed student loans.There are no statistically significant differences between men and women for any of these variables and in conclusion, regarding their dedication to the profession.

We further provide information of the type of firm that employs the respondents. We keep track of the size of the firm, which captures the number of lawyers in the firm. We further add the size of the office. As both measures display bunching (individuals are more likely to mention a number ending in 0), we bracket both firm and office size. A higher number corresponds to a larger firm/office. Female and male respondents do not select into differently sized firms or offices. Women, however, work in more female environments. This is captured by *% Men*, which

[20]However, including performance in undergrad does not affect results in our regression results once we control for law school performance. This indicates that the undergrad performance is per se irrelevant once further education is taken into account.

measures the percentage of male lawyers in the workplace. Moreover, men are more likely to work for a private firm: the variable *Private Firm* is binary with value one if the respondent works in a private law firm.

While we demonstrate the similarities across all lawyers, our model speaks mostly to the private sector. We therefore provide overall summary statistics as well as those by gender for those working for a private sector in the Supplementary Appendix Tables A.3 and A.4. The gender differences remain largely unchanged, with the exception of law school quality: the women sorting into a private law firm attended better law schools compared to male lawyers. Taken together, women and men are fairly similar, especially when it comes to the dedication to the job. If anything, women are somewhat younger, less likely to be married, with fewer children and slightly better educated.

Despite the pronounced similarities, the treatment female and male respondents receive at the workplace is markedly different. We consider whether these lawyers received demeaning or discriminatory comments. Both measures are once again binary, with value one if the respondent has indeed received such comments. Women are significantly more likely to receive these comments relative to men. In contrast, men are more likely to spend time with partners. The variable *Recreational Time Partners* indicates with a one if the respondent spends recreational time with partners, *Breakfast Partners* is a dummy taking value one if the lawyer goes for breakfast with partners. Women seem to be excluded from these social interactions. However, they are not less social: *Recreational Time Associates* takes value one if the survey participant spends recreational time with their associates, which occurs at a higher rate for women.[21]

Having identified disparities in treatment at work, we then ask whether these different experiences affect the values for promotion, another key assumption of our model.

Assumption 3. *Values for a promotion are affected by work culture.*

We proxy the value for a promotion by aspirations to become an equity partner. Respondents are asked how strongly they would like to become an equity partner. They report their aspiration on a scale from one to ten, with one indicating that they have no interest at all. Ten denotes a very strong aspiration to become equity partner. The aspirations are measured in Wave II, while information on work culture is collected in Wave I. This allows us to ensure that aspirations are not a cause of the treatment received. We report summary statistics of aspirations in Table 9 and separately by gender in Table 10. Aspirations are significantly higher for men than women. This discrepancy emerges despite observed gender equality in dedication in the first wave.

In order to assess the association between work culture and aspirations, we create a variable *Bad Experience*. This is an indicator equal to one if a respondent either receives a demeaning or discriminatory comment. Similarly, we create an indicator variable *Good Experience*, which is one if the respondent spends recreational time with either associates or partners, or attends breakfast with a partner. Further, we create a dummy variable *Time with Partner*, which takes

[21] We focus here on discrepancies in terms of treatment that is unrelated to the actual work as we are interested in the effect of work culture.

value one if the respondent spends time with the partner, that is either recreational time or through breakfast. We additionally create for each of these variables an alternative variable that takes into account intensities, where we add up demeaning and discriminatory comments as well as the time spent with partners and associates. We report the summary statistics for the constructed variables in Table A.1 by gender.

We then estimate the following model using OLS to evaluate the connection between experiences and aspirations:

$$a_i = C_i\rho + X_i\beta + \epsilon_i, \tag{25}$$

where a_i denotes aspirations of respondent i. C_i contains the workplace culture experience measures, while X_i refers to additional controls, including a constant.

The results for regression (25) are presented in Table 3. Having a bad experience is associated with a one point decrease in aspirations, while having a good experience does not affect aspirations. We therefore consider whether time spent with partner is associated with a raise in aspirations: this is indeed the case. It turns out that in particular spending recreational time with a partner relates positively to aspirations. A respondent who gets to spend time with their partner has a .8 higher aspiration. If we control for gender, then the effect of a bad experience vanishes, while the effect of recreational time spent with a partner persists. This is unsurprising given that women have disproportionately negative experiences, while the gender disparity in positive experiences is somewhat less pronounced. This pattern persists when we add additional controls in the last column of Table 3. There, we add indicators for race, age, marital status as well as a control for the percentage of men in the workplace. We further consider intensities, that is whether there is an effect of the number of positive and negative experiences. Our results are confirmed, see the Supplementary Appendix Table B.5.[22]

Relative to the overall number of respondents, less than half of them filled in their aspiration. In order to investigate, whether participants are selected in this sample, we create an indicator variable which equals one if the respondent shared his aspiration and zero if the observation is missing. The results for these regressions using a linear probability model are provided in Table 11. The aspiration measures are selected in three ways: first, older respondents are less likely to register their aspirations, although the effect is quantitatively negligible. Second, those who experienced discrimination are less likely to state aspirations. Third, those in private firms are more likely to report their aspirations. The last two effects are sizable and similar in magnitude. Our results therefore are more representative of the situation in private firms. Additionally, as those who experience discrimination are less likely to fill in aspirations, we suspect that our results on the association between bad experiences and aspirations provide a lower bound on the magnitude of the connection. We show that these results are robust to alternative estimation methods, see the Supplementary Appendix Tables B.6 and B.7. In sum, we suspect that the

[22] We further considered indicator variables for the intensities. Our results are robust to this specification, available upon request.

Table 3: Aspiration and Work Experiences

	Dependent Variable: Aspiration						
Bad Exp.	-1.077***	-0.603	-0.967**	-0.504	-1.034***	-0.554	-0.240
	(0.397)	(0.402)	(0.399)	(0.403)	(0.389)	(0.396)	(0.417)
Good Exp.	0.121	-0.0662					
	(0.559)	(0.545)					
Time with Partner			0.584*	0.488			
			(0.340)	(0.331)			
Rec. Time Partner					0.851***	0.801**	0.772**
					(0.329)	(0.321)	(0.329)
Female		-1.588***		-1.558***		-1.561***	-1.352***
		(0.319)		(0.319)		(0.319)	(0.335)
Observations	438	438	438	438	438	438	406
R^2	0.017	0.070	0.024	0.075	0.032	0.083	0.122
Additional Controls	No	No	No	No	No	No	Yes

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The dependent variable is *Aspiration* measured in Wave 2. Regressors are from Wave 1. *Bad Experience* is an indicator equal to one if the respondent either experienced discriminatory or demeaning comments. *Good Experience* is an indicator equal to one if the respondent spent recreational time with either partner or associates, or had breakfast with a partner. *Time With Partner* indicates whether the individual had breakfast with the partner or spent recreational time with them. *Recreational Time Partner* indicates if the respondent spent recreational time with a partner. The last column includes additional controls: race indicators, age, marital status in wave 2, number of kids in wave 2, the percentage of men at the workplace in wave 2.

negative effect of a bad experience could be exacerbated if the survey participant incurring it would have declared their aspirations.

We further provide a comparison of the difference in summary statistics for men and women across samples in the Supplementary Appendix, Table E.17. They highlight that discrepancies between samples are minimal, further affirming that our results are not driven by sample selection.

Taken together, we show that work place culture is related to aspirations. The negative effect of a bad experience is absorbed by gender as women face the majority of discriminatory and demeaning comments. Even though women are, if anything positively selected, they end up with lower aspirations. This is shaped by their experiences at the workplace and results in significantly different aspirations, or valuations, for becoming partner, the relevant promotion. In contrast, spending recreational time with a partner is associated with an increase in aspirations, an effect which is not solely determined by gender.

Having established that aspirations are influenced by workplace culture, we turn next to our predictions about aspirations, their relationship to effort and ultimately, to the promotion to partner.

5.2 Testing Theoretical Predictions

We begin by establishing the commonalities of our theoretical results. Our individually optimal designs translate into group optimal adjustments, while joint designs imply reallocation of values across groups.

Prediction 1. *The distribution of aspirations have the following features:*

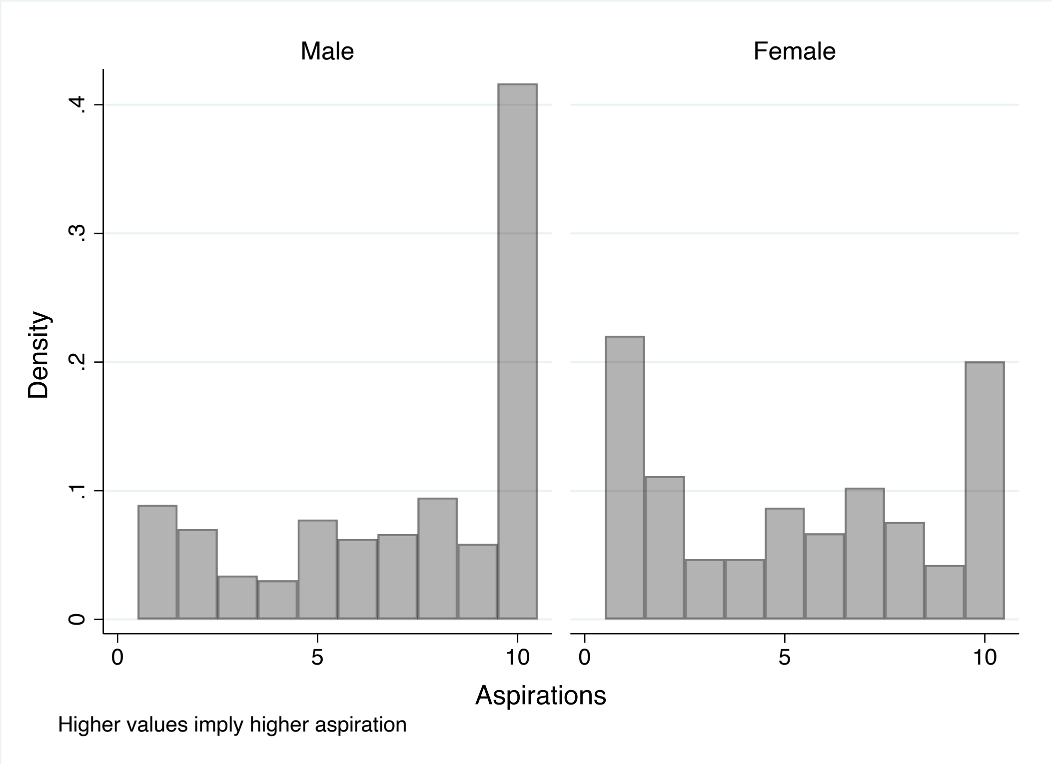
(a) If firms treat groups distinctly and separately, it is optimal to generate a bimodal value distribution for each of them (Proposition 2).

(b) If firms design aspirations jointly across groups, then one group receives higher values than the other (Remark 4, Proposition 4).

Our predictions hinge on the constraints the employer faces for his design of aspirations. If the employer designs distributions for each group individually, then a bimodal distribution for each group must emerge. If in addition, he is able to reallocate value across groups, then different groups must display different averages in terms of aspirations. The latter holds, as we already documented in Table 10. It indicates that the employer is indeed able to reallocate values for the promotion across groups.

We turn to the distribution of aspirations by gender, see Figure 6. The left panel provides the distribution for men, the right the aspirations for women. Women's aspirations have the highest density for the lowest value: more than 20% of women indicate no interest in becoming partner. The second highest density emerges for the highest aspiration with 20% of women reporting the strongest aspiration to become partner. Overall, we find a bimodal distribution for women, in line with our theoretical prediction. For men, the distribution is unimodal: 40%

Figure 6: Distribution of Aspirations by Gender



Note: Aspirations measured on scale from 1-10, Number of Observations: Men 528, Women 449, Total 977

of men state that they very strongly prefer to become partner. Such a result can emerge if the share of men with a low aspiration is very low, indistinguishable from noise in the data.

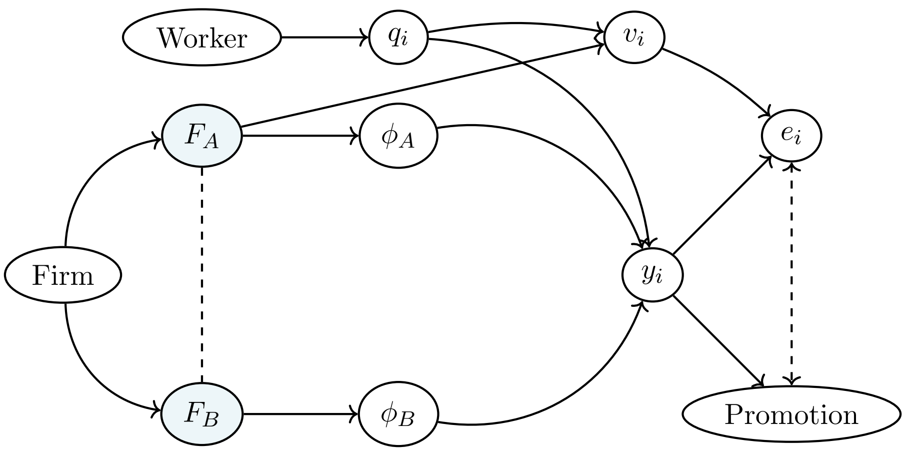
It is worth noting that the distribution of aspirations of men first order stochastically dominates the distribution of female lawyers. This may seem like a contradiction to Proposition 3 which establishes that a FOSD value destruction is not profitable. However, the results in Table 3 indicate values are not only destroyed through bad experiences. Rather, for one group values are diminished, while they rise mostly for the other group by allocation good experiences such as time spent with partners across groups. This is consistent with Remark 4 and Proposition 4. Both results imply that if a decrease in aspirations is at least matched by an increase in aspirations by the other group, then such a reallocation is strictly profitable. This implies by continuity that it remains profitable even if there is a sufficiently small loss in terms of aspirations.

Overall, the distributions confirm key predictions of our model and are in line with employers both re-allocating values across groups as well as redistributing values within groups. Even though our constraints are arguably abstract they seem to be able to generate meaningful predictions that are consistent with empirical facts.

A concern regarding our finding may be that the bimodal distribution was already present when lawyers start out in the profession. Our data does not contain a question regarding aspirations in the first wave and so we cannot map changes in aspirations. However, we can use one of our measures of dedication, namely how important it is to the respondent to become influential in a powerful position as it is recorded over a scale from 1 to 5.[23] We again separate the responses by gender and find that the responses are uni-modal, with most responses around the mean and smaller tails, see Figure 9. Therefore, it does not seem plausible that distributions have always been bi-modal.

Having established that the distributions of aspirations are consistent with our predictions, we turn to the connection between aspirations, effort and promotions. The relationships between them have already been described in Myerson (1981) and we build our following predictions on the results provided there. Our goal is to support our claim that observed effort and the allocation of promotions is consistent with the optimal mechanism. In order to clarify the connection between aspirations, effort and promotions, we present a causal graph in Figure 7. This graph yields a number of predictions, which we discuss in turn.

Figure 7: Value, Effort and Promotions



Note: The distribution of values is F. q denotes the quantile/type of the worker. The quantile and the distribution pin down the value, the aspiration. ϕ is the virtual value, the interim promotion probability is y, e captures effort. Every solid arrow between variables denotes a positive relationship, the dashed bijection indicates that there is an association between effort and the promotion: the worker with the highest marginal effort receives the promotion. The dashed line indicates that the distributions for the exact value. The blue nodes are unique to our model, the white nodes follow from Myerson (1981).

First, higher aspirations induce higher effort. The aspirations proxy values for the promotion, which depend on the an individual's quantile as well as the distributions induced by the employer, see Figure 7. The values then affect effort.

Prediction 2. *Higher aspirations induce higher effort within each group (Myerson (1981)).*

To measure effort, we use four different variables from Wave II: (i) *Hours Office*, which are the weekly hours spent at the office, (ii) *Total Hours*, which consists of the sum of hours spent at the office and hours worked outside the office, (iii) *Hours Billed* per week, and (iv) *New Clients*, an indicator variable equal to one if the respondent brought in a new client. All variables refer

[23] The other other measures are binary, except for *Loans*, which also yields a normal distribution, available upon request.

to 2006, the year before the second survey was conducted. Summary statistics are presented in Table 9, those by gender in Table 10.

We estimate the following model using OLS

$$e_i = a_i \rho + X_i' \beta + \epsilon_i, \tag{26}$$

Our dependent variable is effort e_i, captured by one of the measures described above. a_i is aspiration, between one and ten. We include a number of controls. In addition to the controls included in X_i above, we further take into account the quality of their law school, the percentage of men in their workplace, and indicators for firm and office size. We further include controls for expectations, that is the probability a respondent assigns to becoming equity partner.[24]

We provide the results for billable hours and new clients in Table 4 and relegate Table 12, which ties aspirations to hours worked to the Appendix. The first three columns in Table 4 relate to billable hours, while the final three columns consider as dependent variable whether the respondent brought in a new client. The first column highlights the correlation between aspiration and hours billed, for the largest sample. Column (2) replicates column (1) but for the smaller sample that emerges when we add controls in column (3). Therefore, the samples used to estimate columns (2) and (3) are identical. In column (1) a respondent, whose aspirations switch from 1 to 10 has 9h more billable hours per week. The effect is smaller for our restricted sample, highlighting that our results do not stem from sample selection. In column (3), the effect is intermediate and remains large.[25] A similar picture emerges for *New Clients*: higher aspirations make it more likely for a lawyer to bring in a new client. Turning to hours worked, we find that hours at the office are associated with an increase by almost 6h if aspirations switch from one to ten; for total hours the effect amounts to more than 7h. Remarkably, the expectation to become equity partner is insignificant and does not seem to affect the effort (with the exception of new clients). Independently of how likely a respondent thinks he is going to be partner, he does not increase or decrease his effort. This indicates that aspirations capture something other than beliefs and information, namely a value for the promotion.

Overall, aspirations affect effort significantly for each group: for women the effort is significantly lower relative to men.

We conduct two sets of robustness checks for our results. First, we provide in the Supplementary Appendix Table 4 the results of a Probit and Logit regression on *New Clients*, which confirm the findings in our linear probability model. Second, we provide a comparison of the gender differences across samples in the Supplementary Appendix, Tables E.17 and E.18 once more affirming that results are not due to sample selection.

Independently of how me measure effort, we find a positive relation between these measures

[24] We also considered a specification that controlled for undergrad measures as a robustness check given that these are variables that differ for men and women. Taking them into account does not have an effect, but reduced sample size substantially.

[25] The increase in billable between columns (2) and (3) is driven by a negative correlation between aspirations and size of the firm.

Table 4: Billable Hours, New Clients and Aspirations

	Hours Billed	Hours Billed	Hours Billed	New Clients	New Clients	New Clients
Aspiration	0.916***	0.803***	0.785***	0.0332***	0.0385***	0.0224***
	(0.124)	(0.144)	(0.178)	(0.00449)	(0.00556)	(0.00778)
Female			-3.588***			-0.0253
			(0.911)			(0.0411)
Age			0.00442			-0.000720
			(0.103)			(0.00453)
Married			-0.219			0.0260
			(1.080)			(0.0532)
# Kids			-1.062**			-0.00693
			(0.424)			(0.0190)
% Men			0.00484			0.00250**
			(0.0332)			(0.00121)
$[25, 50]$ % Partner			-0.929			0.0724
			(1.241)			(0.0615)
$[50, 75]$ % Partner			-1.121			0.132*
			(1.388)			(0.0688)
$[75, 100]$ % Partner			-1.547			0.194***
			(1.299)			(0.0648)
Observations	867	640	640	974	640	640
R^2	0.071	0.056	0.283	0.051	0.065	0.187
Additional Controls	No	No	Yes	No	No	Yes

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: *Hours Office* measures weekly time at office. *Total Hours* is the sum of *Hours Office* and hours worked outside the office in a week. *Aspiration* measures the level of aspirations on a scale from one to ten. *Female* is an indicator equal to one if the respondent is a woman. *Married* equals one if the respondent is married or lives in a domestic partnership in Wave II. *% Men* refers to the share of men at the workplace in Wave II. $[x, y]$*% Partner* is an indicator for the percentage the respondent assigns to becoming partner. It takes value one if the respondent's percentage lies between $[x, y]$. In the third and last column we include as additional controls: *Law S Quality* and from Wave II *Firm Size* [], *Office Size* [].

and aspirations.

We turn to the connection between aspirations, effort and attaining the promotion to partner. Our model does not directly tie in aspirations and promotion probabilities, but rather shows that a higher virtual value is associated with a higher probability of being promoted. We therefore calculate the virtual value, which is a function of the aspiration and the information rent. We compute the information rent by gender, in line with our model: the information rent for some aspiration level a simply amounts to the empirical cumulative probability of aspirations higher than a divided by the empirical probability of aspiration a. The information rent together with the aspiration mechanically determine the virtual value.

It is worth noting that there is an association between effort and attaining the promotion as effort is pinned down by the interim allocation probability, which depends on the the virtual values of both agents and is unobserved. This implies that when taking into account an agent's virtual value, we would still expect for effort to have an effect as it depends on the virtual values of both agents through the interim allocation probability. This motivates our final set of predictions.

Prediction 3. *Following Myerson (1981), for each group*

1. *a higher virtual value makes becoming partner more likely*
2. *there is a positive association between effort and attaining the promotion*

As the virtual value is pinned down by the value, our aspiration, and the information rent, we can also investigate the effect of aspiration and information rent separately. Given that employers aim to minimise information rent by designing value distributions, we expect that information rent has a negligible effect on attaining the promotion.

Prediction 4. *If the employer does not face distributional design constraints, the information rent has no effect on attaining the promotion in an optimal design, in line with Propositions 2 and 4.*

We test these predictions using the following empirical models,

$$P_i = \psi_i \rho + E_i \gamma + X_i' \beta + \epsilon_i, \tag{27}$$

$$P_i = a_i \rho_1 + r_i \rho_2 + E_i \gamma + X_i' \beta + \epsilon_i, \tag{27'}$$

The dependent variable P_i is an indicator equal to one if the respondent is either non-equity or equity partner in Wave 3. The virtual value is denoted by ψ_i. We include all of our effort measures, summarised in the matrix E_i. All additional controls, namely race, quality of law school, the percentage of men in their workplace, and indicators for firm and office size, are summarised by X''. We then replace the virtual value by its two determinant, aspiration and information rent, r_i.

As our LHS variable is binary, we estimate (27) and (27') using a linear probability model, probit as well as logit. We provide the results for the linear model (27) in Table 5, those for the

Table 5: Virtual Values and Becoming Partner

	Partner	Partner	Partner	Partner
Virtual Value	0.0106***	0.0127***	0.00969***	0.00866***
	(0.00175)	(0.00221)	(0.00226)	(0.00232)
New Clients			0.111***	0.109***
			(0.0364)	(0.0394)
Hours Billed			0.00579***	0.00510***
			(0.00147)	(0.00179)
Hours Office			0.00133	0.00212
			(0.00179)	(0.00188)
Total Hours			0.00140	0.000870
			(0.00178)	(0.00190)
Female				-0.0288
				(0.0393)
Age				-0.000122
				(0.00413)
Married				0.0132
				(0.0494)
# Kids				0.0103
				(0.0186)
% Men				0.000603
				(0.00111)
Observations	977	657	657	657
Additional Controls	No	No	No	Yes

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: *Partner* is an indicator equal to one if the respondent is either equity or non-equity partner in Wave III. *Virtual value* is calculated according to the aspiration distribution generated in Figure 6 for each group. *Hours Billed*, *Hours Office*, *Total Hours* refer to billable hours, hours at the office and total hours per week in Wave II. *New Clients* is an indicator equal to one if respondent brought in a new client in Wave II. *Female* equals one if the respondent is a woman. *Age* measures 2002- birth year. *Married* equals one if the respondent is married or lives in a domestic partnership in Wave II. *% Men* refers to the share of men at the workplace in Wave II. In the third column we include as additional controls *Law S Quality* and from Wave II *Firm Size* [], *Office Size* [].

linear model (27') in Table 6. The probit and logit regressions together with marginal effects are provided in the Supplementary Appendix, Tables D.9 to D.16.

Table 5 shows in the first column the correlation between virtual value and becoming partner. It is significantly positive: increasing the virtual value is connected to a .01 increase in the probability of becoming partner. The positive association persists if we control for effort measures. In line with our prediction, effort is positively related to becoming partner, in particular new clients and hours billed.[26] These effects continue to hold once we add additional controls, of which we report a subset of in column (3) of Table 5. Interestingly, gender no longer has an effect on becoming partner, highlighting that the discrepancies in outcomes do not appear to be driven by gender per se, but rather by different experiences, which affect aspirations, effort and ultimately, whether a respondent becomes partner. The same holds true for the number of children, which are no longer significant. Last, the share of men in the workplace remains irrelevant.

We replace the virtual value by aspirations and the constructed information rent, see Table 6.

Once again, effort is positively associated with becoming partner. The information rent does not have an effect, establishing that firms seem to be able to learn the aspirations of their employees.[27] Aspirations has a significantly positive effect: moving from having no aspiration to become partner to aiming strongly for it, is associated with an increase in the probability of becoming a partner of .3 points.

Taken together, we have established that workplace culture affects aspirations. A distribution of aspirations emerges, which is different across the two groups we consider, men and women. Aspirations determine effort. Aspirations have an impact on whether promotions are achieved. Interestingly, once aspirations are taken into account, gender differences in promotions vanish.

5.3 A Robustness Check: Racial Disparities

While our empirical focus has been on gender, a natural fit as it allowed us to split the population of lawyers into two equally sized groups, it is worth understanding whether our model also applies to other groups. We therefore turn to distribution of aspirations across racial groups.

The average aspirations by race (and by gender to provide a benchmark for the magnitude of differences) is provided in Table 7. We focus on Race Indicators for which we have larger samples (those above thirty). This leaves us with Black, Hispanic, Asian and White lawyers. Unsurprisingly, white lawyers have the highest average aspiration, followed by those who identify as Hispanic. The second lowest average aspiration is displayed by Asians and the lowest by Black lawyers.

[26] Hours in the office and total hours are highly correlated as total hours is the sum of hours in the office and hours outside the office.

[27] The significance of information rent depends on the estimation method used as well as sample selection, see the Supplementary Appendix. The effect is quantitatively small even when it becomes significant. Overall, it seems that employers learn realised aspirations reasonably well.

Table 6: Aspirations and Info Rent and Becoming Partner

	Partner	Partner	Partner	Partner
Aspiration	0.0378***	0.0461***	0.0356***	0.0388***
	(0.00433)	(0.00546)	(0.00619)	(0.00773)
Info Rent	0.00379	0.00387	0.00221	0.00529
	(0.00278)	(0.00350)	(0.00354)	(0.00419)
New Clients			0.0892**	0.0918**
			(0.0372)	(0.0398)
Hours Billed			0.00470***	0.00425**
			(0.00151)	(0.00180)
Hours Office			0.000590	0.00142
			(0.00178)	(0.00188)
Total Hours			0.00118	0.000901
			(0.00177)	(0.00189)
Female				0.0421
				(0.0450)
Age				0.000581
				(0.00404)
Married				0.0180
				(0.0493)
# Kids				0.00516
				(0.0186)
% Men				0.000749
				(0.00111)
Observations	977	657	657	657
Additional Controls	No	No	No	Yes

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: *Partner* is an indicator equal to one if the respondent is either equity or non-equity partner in Wave III. *Aspiration* measures the level of aspirations to become equity partner on a scale from one to ten. *Info Rent* is calculated according to the aspiration distribution generated in Figure 6 for each group. *Hours Billed*, *Hours Office*, *Total Hours* refer to billable hours, hours at the office and total hours per week in Wave II. *New Clients* is an indicator equal to one if respondent brought in a new client in Wave II. *Female* equals one if the respondent is a woman. *Age* measures 2002- birth year. *Married* equals one if the respondent is married or lives in a domestic partnership in Wave II. *% Men* refers to the share of men at the workplace in Wave II. In the last column we include as additional controls *Law S Quality* and from Wave II *Firm Size* [], *Office Size* [].

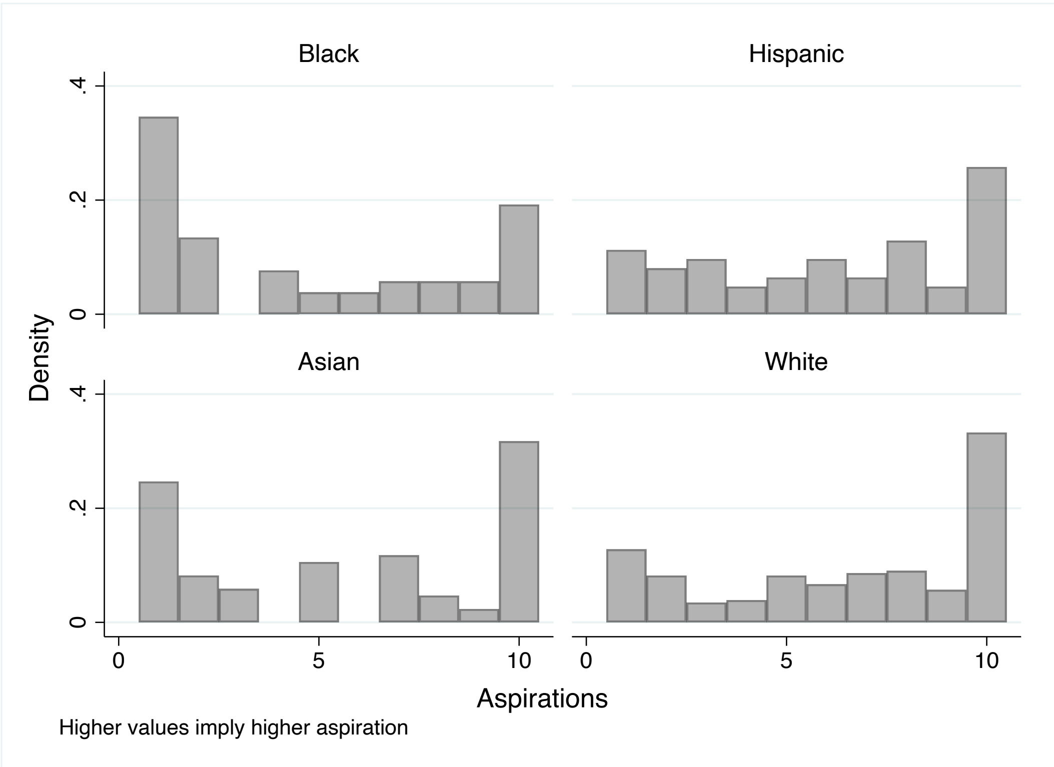
Table 7: Aspirations by Race and Gender

	Mean	S.D.	N
Black	4.65	0.51	52
Hispanic	6.16	0.41	62
Native American	6.83	1.14	6
Asian	5.71	0.40	85
White	6.54	0.12	718
Other	4.73	0.78	15
Mixed White	6.14	0.66	29
Mixed Non-White	4.67	1.58	6
Total Across Race	6.29	0.11	973
Male	7.13	0.14	528
Female	5.31	0.16	449
Total Across Gender	6.30	0.11	977

Note: Race allows for multiple mentions: those who identified as white and at least one other race were coded as *Mixed White*, those with multiple race identifier were coded as *Mixed Non-White*.

Notably, all aspiration distributions are bidmodal, with the exception of Hispanic lawyers. Their distribution is unimodal with the highest density of lawyers with the highest aspiration.

Figure 8: Distribution of Aspirations by Race



Note: *Aspirations* measure the aspiration of the respondent to become equity partner on a scale from one to ten.

This highlights that bi-modal distributions are pervasive in the context of lawyers, independently of which groups we consider. Moreover, these groups display distinct averages. As before, this suggests that employers may both reallocate values across groups as well as redistribute values within groups, consistent with the proposed mechanism and constraints.

5.4 Alternative Explanations

We have established that the predictions of our model are consistent with empirical patterns in law, implying that the data do not falsify the model. There may however also be other explanations for observed patterns.

One potential source of the discrepancies may be taste-based discrimination (Becker (1957)). However, if there was an underlying distaste toward certain types of employees, they would not even be hired. This contradicts that the law firms employ female lawyers, even if they are somewhat positively selected. It could also be that there is a distaste for female partners, while female associates are accepted. However, such a distaste would not explain why women's aspiration are diminished.

Alternatively, the discrepancies could arise due to statistical discrimination (Phelps (1972),Arrow (1973)). Arguably, there is less information about a worker's ability available at the hiring stage relative to the promotion stage. This 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 what we find here, but also more generally contradicts 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.[28] Similarly, Altonji and Pierret (2001) document that race matters more further down the career path, less so at the entry level, again in contrast to standard models of statistical discrimination.

It does not seem to be the case that family constraints are a major driver of becoming partner given that the marital status and gender do not have a significant effect attaining a partner position. Azmat et al. (2020) show that lower aspirations relate to a higher number of children, emphasising the importance of aspirations as a driver of both family and workplace decisions. They investigate alternative mechanisms to understand the discrepancies in promotions, and also arrive at aspirations as a key determinant.

Last, we provide correlations. We argue that for the purpose of our exercise this is sufficient. Our theory predicts that an employer has an incentive to introduce inequalities among workers based on some group marker. We select gender and race as these are observable in our data. However, firms could also introduce discrepancies based on an unobservable that ties in with the labels. We would then ultimately capture the effect of a label not present in the data–which is still consistent with our theory of firms generating profitable inequality.

[28] 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.

6 Conclusion

Our analysis of individual design problems highlights that an employer can induce higher effort provision by introducing cultures in which workers value the promotion very much or not at all. Results also establish that pure value destruction does not suffice to increase effort provision and that surplus growth is a key force in all value design problems. Additionally, the analysis of joint design problems shows that the employer always benefits by reallocating value, so that one worker displays a high value for the promotion, while the other one ends up with a lower valuation. We establish empirically that these discrepancies emerge for different groups of lawyers: their distributions of valuations for the promotion, measured by aspirations, are bimodal and there are significant disparities in average aspirations across different groups.

Our model does not make a prediction as to which group will be favoured or disadvantaged. Given that inequalities are profitable, employers have no incentive to tackle them and so historical patterns can persist, potentially driven by discrimination. It is then unsurprising that women and racial minorities tend to be disadvantaged, while white men display the highest aspirations.

However, we would not expect women and minorities to be as underrepresented at entry level jobs. If we posited that value design costs are related to the extent to which values need to be manipulated, at the hiring stage, the employer would prefer to enlist workers with high value distribution, independently of their gender or other characteristics. This follows as higher average values result in higher value extraction, for instance if the employer could freely design any value distribution with a mean bounded by the expected value of the initial distribution. Similarly, the employer would also prefer to recruit workers with distributions with low information rents. This translates into screening individuals to select those with high expected value and low variation. Such a hiring strategy is arguably followed by firms in law consulting and finance, where graduates from top programs are usually employed.

Our results imply that even conditional on such a hiring strategy, the employer has an incentive to create, through compensation design, work obligations, fringe benefits, and workplace culture, a certain value distribution for their entry level employees – a novel feature of our approach which contrasts with existing theories and explanations for gender gaps in promotions.

We discuss two possible extensions, namely, a cost to designing values and correlated values in Appendices A and B. Further extensions not considered in the analysis include having (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 optimal mechanisms. We therefore leave these extensions to future research.

A Cost of Value Design

Our approach focuses on design constraints rather than costs to highlight the effects of different designs on the employer's profit proxied by total effort. Nevertheless, our two main results are robust to costly design.

Value Dispersion Conceptually, constraint (7) posits that any distribution with a mean at or below k costs the same and we characterise the optimal distribution under this assumption. It may be sensible to posit that value distributions with lower means should have a lower design cost. In such settings, there would be a force pushing towards distributions that have a lower mean and with it a lower cost. The shape of the cost function would determine the mean set by the employer, which would be positive unless costs are prohibitively high. Setting a mean at zero would lead to zero valuation among workers and nobody would compete for the promotion. Therefore, it seems sensible to assume that the firm will allow for some positive k. Given a positive mean chosen by the employer, he would still choose distributions that display maximal dispersion. We also explicitly consider the case where the employer incurs a fixed cost from adjusting the distribution in our analysis, see Remark 1.

Value Reallocation We then turn to the case of reallocating values across workers. Although the assumption on redistribution implicit in this model may seem strong, similar insights would naturally arise in settings in which designing is costly and satisfies a weak linearity property requiring that any two value distributions having the same sum to cost the same. This assumption would be fulfilled by common design cost functions such as entropy, and would be met for instance by integrable cost functions satisfying

$$\int_0^{\overline{\omega}} c(v) d[F_A(v) + F_B(v)]. \tag{28}$$

for some function $c : [0, \overline{\omega}] \to \mathbb{R}_+$.

Thus, even in a model of costly design, we would get maximal discrimination and maximal dispersion as features of optimal designs.

B Interdependent Values

Maintaining the independence of value distributions across design problems is a natural assumption if one believes that the underlying skills and preferences of a worker are an independent trait that cannot be affected. It further leads to a more challenging problem, as it requires accounting for workers' information rents rather than focusing purely on surplus design. In addition, our results carry over to a setting with interdependent values.

Consider an interdependent value setting (Crémer and McLean (1985)) in which total effort coincides (generically) with surplus, $\mathbb{E}[\max\{v_A, v_B\}]$. To provide some insight while remaining close to the core of the analysis, we allow the employer to design marginal value distributions for both workers, but not the correlation structure between workers' values, which is fixed. This was also the case in the baseline analysis where the employer designed marginal value distributions for both workers, but was unable to affect the independence of workers' values.

To explain how we fix the correlation structure, take any joint distribution in the quantile space $Q : [0,1]^2 \to [0,1]$. The joint distribution $Q(q_A, q_B)$ has uniform marginal distributions by construction, but can display arbitrary correlation structures – such as independence, perfect positive correlation (*concordance*), and perfect negative correlation (*discordance*). Posit that Q determines the underlying correlation of skills across workers and cannot be affected by the employer, as was the case in the original setup where we always had that $Q(q_A, q_B) = q_A q_B$. As before, what the employer designs are the values associated to each quantile for both workers, $v_A(q_A)$ and $v_B(q_B)$. This is equivalent to designing the two marginal value distributions, $F_A(v_A)$ and $F_B(v_B)$. We consider this to be a suitable approach when quantiles reflect the underlying ability or intrinsic preferences of workers, since the correlation of these abilities across workers cannot be designed.

Individual Design Insights: When designing marginal distributions subject to the mean-bound constraint $\mathbb{E}_{F_i}[v_i] \leq k$ for all i, extreme bi-modal designs remain optimal. This follows because variance increases the expected value of the maximum of two random variables, $\mathbb{E}[\max\{v_A, v_B\}]$ by the convexity of the maximum operator. As an example, let quantiles be discordant, so that $q_A = 1 - q_B$ for any pair of quantiles (q_A, q_B) in the support of Q. Additionally assume that $\overline{\omega} = 2$, and that $k = 1$. If so, surplus, or equivalently total effort, is maximised by a value design (F_A, F_B) in which workers value the promotion at $v_i(q_i) = 2$ if $q_i \leq 1/2$ and at $v_i(q_i) = 0$ otherwise. Such design corresponds to maximal dispersion as described in Proposition 2. Under such a maximally spread two-atom distribution, surplus is exactly equal to 2, as at least one worker must value the promotion at 2. Thus, total effort also equals 2, and the employer secures such a surplus by awarding the promotion with certainty to one of the workers exerting effort 2. Clearly, no other value design can lead to higher surplus, as values never exceed 2. In such individual design settings, we conjecture that the design in which both distributions are maximally spread, discussed in Remark 3, will remain optimal when values are negatively correlated. With concordance, the design in which a worker always values the promotion at k while the other is maximally spread, see Remark 3, yields a surplus of 1.5, while the design in which both workers are maximally spread yields a surplus of 1. Based on this example, we conjecture, that with positively correlated values the employer is better off if one worker has all mass at the mean, whereas the other worker's values are maximally spread. While we have not solved this case in full generality, it is intuitive that if values are positively correlated, the designed distributions will be maximally distinct subject to the constraint. Therefore, allowing for such positive value correlation only strengthens our insight that the employer benefits from inequalities.

Joint Design Insights: When designing marginal distributions subject to the reallocation constraint $F_A(v) + F_B(v) = H(v)$ for all v, maximal discrimination remains optimal. This follows as maximal discrimination increases surplus, $\mathbb{E}[\max\{v_A, v_B\}]$, by minimizing the chance of having two workers with high values. As an example, let quantiles be concordant, so that $q_A = q_B$ for any pair of quantiles (q_A, q_B) in the support of Q. Additionally assume that $\overline{\omega} = 2$, and that $H(v) = v$ for all $v \in [0, 2]$. If so, surplus with maximal discrimination simply amounts to the expected value, $\mathbb{E}_{F_i^*}[v_i]$, of the worker i with the high marginal distribution $F_i^*(v_i) = H(v_i) - 1$, which is equal to 1.5. Thus, total effort also amounts to 1.5. Such an effort can be obtained by never promoting worker $j \neq i$ and awarding the promotion to worker i only when his effort coincides with his reported value, and the values reported by the two workers are associated to the same quantile, $q_A(v_A) = q_B(v_B)$. Without discrimination, $F_i(v_i) = H(v_i)/2$ for all i, surplus simply amounts to the expected value of one of the two workers, $\mathbb{E}_{F_i}[v_i]$, which is equal to 1. In such joint design settings, we conjecture that the maximally discriminating value designs, discussed in Proposition 4, will always remain optimal. But other designs may also be optimal for specific correlation structures – for instance, when quantiles are discordant and $q_A = 1 - q_B$. With discordance, no discrimination yields the same surplus as maximal discrimination in the example discussed above. Discordance generally benefits surplus without discrimination, as the employer is certain to face a worker with values in excess of the median in such settings. Nevertheless, discrimination remains optimal.

C Mathematical Proofs

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

$$\mathbb{E}\left[\phi_A(q)y_A(q)\right] + \mathbb{E}\left[\phi_B(q)y_B(q)\right] \leq \mathbb{E}\left[\overline{\omega}y_A(q)\right] + \mathbb{E}\left[\overline{\omega}y_B(q)\right] \leq \overline{\omega}, \tag{29}$$

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 $\mathbb{E}_{F_B}[v] = k$. In a quantile setting, for $\overline{q} = 1 - F^*(0)$, the virtual value for the distribution F^* amounts to

$$\phi_B^*(q) = \begin{cases} \overline{\omega} & \text{if } q < \overline{q} \\ 0 & \text{if } q > \overline{q} \end{cases}. \tag{30}$$

Total effort under distribution F^* is given by

$$\mathbb{E}\left[\phi_A(q)y_A^*(q)\right] + \mathbb{E}\left[\phi_B^*(q)y_B^*(q)\right], \tag{31}$$

while the effort under any alternative F_B amounts to

$$\mathbb{E}\left[\phi_A(q)y_A(q)\right] + \mathbb{E}\left[\phi_B(q)y_B(q)\right]. \tag{32}$$

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

$$\mathbb{E}\left[\phi_A(q)y_A^*(q)\right] + \mathbb{E}\left[\phi_B^*(q)y_B^*(q)\right] \geq \mathbb{E}\left[\phi_A(q)y_A(q)\right] + \mathbb{E}\left[\phi_B^*(q)y_B(q)\right]. \tag{33}$$

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

$$\mathbb{E}\left[\phi_B^*(q)y_B(q)\right] - \mathbb{E}\left[\phi_B(q)y_B(q)\right] \geq 0. \tag{34}$$

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

$$\mathbb{E}\left[\phi_B^*(q) - \max\{\phi_B(q), 0\}\right] \geq 0 \quad \Rightarrow \quad \mathbb{E}\left[(\phi_B^*(q) - \phi_B(q))y_B(q)\right] \geq 0. \tag{35}$$

Moreover, by construction, the expected virtual values satisfy

$$\mathbb{E}\left[\phi_B^*(q)\right] = \mathbb{E}\left[v_B^*(q)\right] = \overline{q\omega} = \mathbb{E}\left[v_A(q)\right] = \mathbb{E}\left[v_B(q)\right]. \tag{36}$$

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

$$\mathbb{E}\left[v_B(q)\right] \geq \mathbb{E}\left[\max\left\{\phi_B(q), 0\right\}\right], \tag{37}$$

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 when $F_A \neq F^*$, consider any other distribution F_B with mean k and a single atom on positive values at ω_B, meaning that

$$F_B(v) = \begin{cases} 1 & \text{if} \quad v = \omega_B \\ 1 - \frac{k}{\omega_B} & \text{if} \quad v < \omega_B \end{cases} . \tag{38}$$

Letting $p_B = k/\omega_B$ denote the probability of having value ω_B, total effort can be rewritten as

$$\mathbb{E}[\phi_A(q)y_A(q)] + \mathbb{E}[\phi_B(q)y_B(q)] \tag{39}$$

$$= (1-p_B)\int_{V_A} \max\{\psi_A(v_A), 0\}dF_A(v_A) + p_B \int_{V_A} \max\{\psi_A(v_A), \omega_B\}dF_A(v_A) \tag{40}$$

$$= (1-p_B)\int_{V_A} \max\{\psi_A(v_A), 0\}dF_A(v_A) + \int_{V_A} \max\{p_B\psi_A(v_A), k\}dF_A(v_A). \tag{41}$$

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 $v_B = \omega_B$.[29]

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

$$\frac{\partial \mathbb{E}[\phi_A(q)y_A(q) + \phi_B(q)y_B(q)]}{\partial p_B} = -\int_{V_+(0)} \psi_A(v_A)dF_A(v_A) + \int_{V_+(\omega_B)} \psi_A(v_A)dF_A(v_A) \leq 0, \tag{42}$$

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 uniqueness result when $F_A \neq F^*$. To prove tho optimality of designs F^K for $K \in [k, \overline{\omega}]$ when $F_A = F^*$, it suffices to show that any design F^K yields the same total effort as F^*, which is immediate and thus omitted. ■

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

$$\Delta(v) = G(v) - F^*(v) < 0, \tag{43}$$

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

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

$$\int_0^v \Delta(t)dt = \int_0^v G(t) - F^*(t)dt < 0. \tag{44}$$

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

$$\int_0^{\overline{\omega}} \Delta(t)dt = \left[t\left(G(t) - F^*(t)\right)\right]_0^{\overline{\omega}} - \int_0^{\overline{\omega}} td\left(G(t) - F^*(t)\right) = 0. \tag{45}$$

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 G is non-decreasing and F^* is constant up until $\overline{\omega}$. Therefore, G second order stochastically dominates F^*. The proof then follows because any distribution that is second order stochastically dominated by G must have a lower mean than G, and because F^* was optimal among all distributions with mean lower than G. ■

Proof of Proposition 3: First Order Stochastic Dominance Suppose by contradiction that the firm found it optimal to set $F_B \neq G$, which implies that $F_B(v) > G(v)$ 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 $\hat{F}_B(v) \geq G(v)$.

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[30]

$$\mathbb{E}[\phi_A(q)y_A(q)] + \mathbb{E}[\phi_B(q)y_B(q)] = \int_0^1 \phi_A(q)y_A(q)dq + \int_0^1 \phi_B(q)y_B(q)dq. \tag{46}$$

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

$$\mathbb{E}[\hat{\phi}_B(q)y_B(q)] \geq \mathbb{E}[\phi_B(q)y_B(q)] \quad \Leftrightarrow \quad \int_0^1 \left(\hat{\phi}_B(q) - \phi_B(q)\right) y_B(q)dq \geq 0. \tag{47}$$

Denote by $D \subseteq [0,1]$ the set of points at which y_B in non-differentiable, and by $C = [0,1]\backslash D$. 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

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

Riemann Stieltjes integrals implies that

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

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

$$\mathbb{E}[(\hat{\phi}_B(q) - \phi_B(q))y_B(q)] = \int_0^1 (\hat{\phi}_B(q) - \phi_B(q))y_B(q)dq \tag{49}$$

$$= (\hat{\alpha}_B - \alpha_B)y_B(1) + \sum_{q\in D} q(\hat{v}_B(q) - v_B(q))\left(-J(q)\right) + \int_C (q\hat{v}_B(q) - qv_B(q))\left(-y_B'(q)\right)dq. \tag{50}$$

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

$$\mathbb{E}[\phi_A(q)y_A(q)] + \mathbb{E}[\hat{\phi}_B(q)y_B(q)] \geq \mathbb{E}[\phi_A(q)y_A(q)] + \mathbb{E}[\phi_B(q)y_B(q)]. \tag{51}$$

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

$$\mathbb{E}[\phi_A(q)\hat{y}_A(q)] + \mathbb{E}[\hat{\phi}_B(q)\hat{y}_B(q)] \geq \mathbb{E}[\phi_A(q)y_A(q)] + \mathbb{E}[\hat{\phi}_B(q)y_B(q)], \tag{52}$$

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

$$v_B^-(q)y_B^-(q) - v_B^+(q)y_B^+(q) \leq \hat{v}_B^-(q)y_B^-(q) - \hat{v}_B^+(q)y_B^+(q), \tag{53}$$

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, $\hat{v}_B^-(q)y_B^-(q) - \hat{v}_B^+(q)y_B^+(q) = \hat{v}_B^-(q)\left(-J(q)\right)$ and as values are decreasing in q, $\hat{v}_B^-(q)\left(-J(q)\right) > v_B^-(q)y_B^-(q) - v_B^+(q)y_B^+(q)$. Thus, there exists once again a deviation that leads to a higher total effort.

Last, suppose the support of the distribution G 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 (53) is violated. We therefore require a more sophisticated approach. In this case, we need to amend expression (48) to

$$\mathbb{E}[\phi_B(q)y_B(q)] = \alpha_B y_B(1) + \sum_{q\in D} q\left[v_B^-(q)y_B^-(q) - v_B^+(q)y_B^+(q)\right] - \int_C q v_B(q) dy_B(q). \tag{54}$$

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

$$\alpha_B x_B(1) + \sum_{v\in D} v q_B(v)\left[x_B^+(v) - x_B^-(v)\right] + \int_{V_A\setminus D} q_B(v) v x_B'(v) dv, \tag{55}$$

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

$$\int_{v^+}^{\hat{v}^+} \hat{q}_B(v) v \hat{x}_B'(v) dv \geq 0, \tag{56}$$

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

$$F_B^*(v) = H(v) \qquad \text{if } v \in [\alpha, v^M) \tag{57}$$

$$F_A^*(v) = H(v) - 1 \qquad \text{if } v \in [v^M, \omega]. \tag{58}$$

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

$$\int_{v^M}^{\omega} \psi_A(v) x_A(v) h(v) dv + \int_{\alpha}^{v^M} \psi_B(v) x_B(v) h(v) dv \tag{59}$$

$$\geq \int_{\alpha}^{\omega} \hat{\psi}_A(v) \hat{x}_A(v) a_A(v) h(v) dv + \int_{\alpha}^{\omega} \hat{\psi}_B(v) \hat{x}_B(v) a_B(v) h(v) dv. \tag{60}$$

For convenience, define

$$\overline{x}_A(v) = \begin{cases} \max\{\hat{x}_A(v), \hat{x}_B(v)\} & \text{if } v \geq v^M \\ \min\{\hat{x}_A(v), \hat{x}_B(v)\} & \text{if } v < v^M \end{cases}, \tag{61}$$

$$\overline{x}_B(v) = \{\hat{x}_A(v), \hat{x}_B(v)\} \setminus \overline{x}_A(v). \tag{62}$$

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

$$\overline{\psi}_i(v) = \begin{cases} \hat{\psi}_i(v) & \text{if } \hat{x}_i(v) = \overline{x}_i(v) \\ \hat{\psi}_j(v) & \text{if } \hat{x}_i(v) \neq \overline{x}_i(v) \end{cases}, \tag{63}$$

$$\overline{a}_i(v) = \begin{cases} a_i(v) & \text{if } \hat{x}_i(v) = \overline{x}_i(v) \\ a_j(v) & \text{if } \hat{x}_i(v) \neq \overline{x}_i(v) \end{cases}. \tag{64}$$

These definitions immediately imply that

$$\int_{\alpha}^{\omega} \hat{\psi}_A(v) \hat{x}_A(v) a_A(v) h(v) dv + \int_{\alpha}^{\omega} \hat{\psi}_B(v) \hat{x}_B(v) a_B(v) h(v) dv \tag{65}$$

$$= \int_{\alpha}^{\omega} \overline{\psi}_A(v) \overline{x}_A(v) \overline{a}_A(v) h(v) dv + \int_{\alpha}^{\omega} \overline{\psi}_B(v) \overline{x}_B(v) \overline{a}_B(v) h(v) dv. \tag{66}$$

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

$$\int_{\max\{\overline{v}, v^M\}}^{\omega} \overline{x}_A(v) h(v) dv + \int_{\min\{\overline{v}, v^M\}}^{v^M} \overline{x}_B(v) h(v) dv \leq 1 - F_A(\overline{v}) F_B(\overline{v}), \tag{67}$$

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

$$\int_{\overline{v}}^{\omega} \overline{x}_A(v) h(v) dv \leq 2 - H(v), \tag{68}$$

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

$$\int_{v^m}^{\omega} \overline{x}_A(v) h(v) dv + \int_{\overline{v}}^{v^M} \overline{x}_B(v) h(v) dv \leq 1 \tag{69}$$

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

$$\int_{v^m}^{\omega} \overline{x}_A(v)h(v)dv + \int_{\alpha}^{v^M} \overline{x}_B(v)h(v)dv \leq 1 \tag{70}$$

Inequality (70) corresponds to the following inequality in the quantile-probability space:

$$\int_q \overline{y}_A(q)dq + \int_q \overline{y}_B(q)dq \leq 1 \tag{71}$$

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

$$\overline{y}_A(1) + \int_{\overline{y}_A} q_A(y)dy + \overline{y}_B(1) + \int_{\overline{y}_B} q_B(y)dy \leq 1 \tag{72}$$

The allocation probability for A, $\overline{y}_A$, lies between 1 and the allocation probability at the median value $\overline{y}_A(1)$. Inequality (72) 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

$$y^- \equiv \lim_{\epsilon \to 0} y(q - \epsilon) \tag{73}$$

$$y^+ \equiv \lim_{\epsilon \to 0} y(q + \epsilon) \tag{74}$$

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:

$$\overline{y}_A(1) + \int_{\overline{y}_A} (\hat{q}_A(y) + \hat{q}_B(y))\, dy + \overline{y}_B(1) + \int_{\overline{y}_B} (\hat{q}_A(y) + \hat{q}_B(y) - 1)\, dy \leq 1 \tag{75}$$

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

$$\overline{y}_A(1) - \overline{y}_B(0) + 2\overline{y}_B(1) + \int_{\overline{y}_A} (\hat{q}_A(y) + \hat{q}_B(y))\, dy + \int_{\overline{y}_B} (\hat{q}_A(y) + \hat{q}_B(y))\, dy \leq 1 \tag{76}$$

If $\overline{y}_A(1) = \overline{y}_B(0)$, that is there is no jump at the median value, then we can rewrite inequality (76) as

$$2\overline{y}_B(1) + \int_{\overline{y}_B(1)}^{\overline{y}_A(0)} \hat{q}_A(y)dy + \int_{\overline{y}_B(1)}^{\overline{y}_A(0)} \hat{q}_B(y)dy \leq 1 \tag{77}$$

$$\Leftrightarrow \quad \int_q \hat{y}_A(q)dq + \int_y \hat{y}_B(q)dq \leq 1, \tag{78}$$

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 (76) can be expressed as

$$\overline{y}_A(1) - \overline{y}_B(0) + 2\overline{y}_B(1) + \int_{\overline{y}_B(1)}^{\overline{y}_B(0)} \hat{q}_A(y)dy + \int_{\overline{y}_A(1)}^{\overline{y}_A(0)} \hat{q}_A(y)dy$$
$$+ \int_{\overline{y}_B(1)}^{\overline{y}_B(0)} \hat{q}_B(y)dy + \int_{\overline{y}_A(1)}^{\overline{y}_A(0)} \hat{q}_B(y)dy \leq 1 \tag{79}$$

Note that $\overline{y}_A(1) - \overline{y}_B(0) = \int_{\overline{y}_B(0)}^{\overline{y}_A(1)} qdy$, 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 (79) to

$$2\overline{y}_B(1) + \int_{\overline{y}_B(1)}^{\overline{y}_A(0)} \hat{q}_A(y)dy + \int_{\overline{y}_B(1)}^{\overline{y}_A(0)} \hat{q}_B(y)dy \leq 1 \tag{80}$$
$$\Leftrightarrow \qquad \int_q \hat{y}_A(q)dq + \int_y \hat{y}_B(q)dq \leq 1, \tag{81}$$

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

$$\int_{v^M}^{\omega} \psi_A(v)\overline{x}_A(v)h(v)dv + \int_{\alpha}^{v^M} \psi_B(v)\overline{x}_B(v)h(v)dv \tag{82}$$
$$\geq \int_{\alpha}^{\omega} \overline{\psi}_A(v)\overline{x}_A(v)\overline{a}_A(v)h(v)dv + \int_{\alpha}^{\omega} \overline{\psi}_B(v)\overline{x}_B(v)\overline{a}_B(v)h(v)dv. \tag{83}$$

This is equivalent to establishing that

$$\underbrace{\int_{v^M}^{\omega} \left(\left(\psi_A(v) - \overline{\psi}_A(v)\overline{a}_A(v)\right)\overline{x}_A(v) - \overline{\psi}_B(v)\overline{x}_B(v)\overline{a}_B(v)\right)h(v)dv}_{\text{Part 1}} \tag{84}$$
$$+ \underbrace{\int_{\alpha}^{v^M} \left(\left(\psi_B(v) - \overline{\psi}_B(v)\overline{a}_B(v)\right)\overline{x}_B(v) - \overline{\psi}_A(v)\overline{x}_A(v)\overline{a}_A(v)\right)h(v)dv}_{\text{Part 2}} \geq 0. \tag{85}$$

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

$$\int_{v^M}^{\omega} \left(\left(\psi_A(v) - \overline{\psi}_A(v)\overline{a}_A(v)\right)\overline{x}_A(v) - \overline{\psi}_B(v)\overline{x}_B(v)\overline{a}_B(v)\right)h(v)dv \tag{86}$$
$$\geq \int_{v^M}^{\omega} \left(\psi_A(v) - \overline{\psi}_A(v)\overline{a}_A(v) - \overline{\psi}_B(v)\overline{a}_B(v)\right)\overline{x}_A(v)h(v)dv = 0. \tag{87}$$

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

$$\psi_A(v) - \overline{\psi}_A(v)\overline{a}_A(v) - \overline{\psi}_B(v)\overline{a}_B(v) = \psi_A(v) - \hat{\psi}_A(v)\hat{a}_A(v) - \hat{\psi}_B(v)\hat{a}_B(v) \tag{88}$$

$$= v - \frac{1 - F_A(v)}{h(v)} - \hat{a}_A(v)\left(v - \frac{1 - \hat{F}_A(v)}{\hat{a}_A(v)h(v)}\right) - \hat{a}_B(v)\left(v - \frac{1 - \hat{F}_B(v)}{\hat{a}_B(v)h(v)}\right) \tag{89}$$

$$= \frac{1}{h(v)}\left(1 + F_A(v) - \hat{F}_A(v) - \hat{F}_B(v)\right) = 0, \tag{90}$$

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

$$\int_{\alpha}^{v^M} \left(\left(\psi_B(v) - \overline{\psi}_B(v)\overline{a}_B(v)\right)\overline{x}_B(v) - \overline{\psi}_A(v)\overline{x}_A(v)\overline{a}_A(v)\right) h(v)dv \tag{91}$$

$$\geq \int_{\alpha}^{v^M} \left(\psi_B(v) - \overline{\psi}_A(v)\overline{a}_A(v) - \overline{\psi}_B(v)\overline{a}_B(v)\right)\overline{x}_B(v)h(v)dv = \int_{\alpha}^{v^M} \overline{x}_B(v)dv \geq 0. \tag{92}$$

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

$$\psi_B(v) - \overline{\psi}_A(v)\overline{a}_A(v) - \overline{\psi}_B(v)\overline{a}_B(v) = \frac{1}{h(v)}\left(1 + F_B(v) - \hat{F}_A(v) - \hat{F}_B(v)\right) = \frac{1}{h(v)}, \tag{93}$$

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

$$\left(\psi_B(v) - \overline{\psi}_B(v)\overline{a}_B(v)\right)\overline{x}_B(v) - \overline{\psi}_A(v)\overline{x}_A(v)\overline{a}_A(v) > 0, \tag{94}$$

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, $\overline{\psi}_A(v)>0>\overline{\psi}_B(v)$

$$\left(\psi_B(v)-\overline{\psi}_B(v)\overline{a}_B(v)\right)\overline{x}_B(v)-\overline{\psi}_A(v)\overline{x}_A(v)\overline{a}_A(v) \tag{95}$$

$$>\left(\psi_B(v)-\overline{\psi}_B(v)\overline{a}_B(v)-\overline{\psi}_A(v)\overline{a}_A(v)\right)\overline{x}_B(v)>0, \tag{96}$$

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

$$\psi_B(v)\overline{x}_A(v)-\overline{\psi}_B(v)\overline{a}_B(v)\overline{x}_B(v)-\overline{\psi}_A(v)\overline{x}_A(v)\overline{a}_A(v) \tag{97}$$

$$\geq\psi_B(v)\overline{x}_A(v)-\overline{\psi}_B(v)\overline{a}_B(v)\overline{x}_A(v)-\overline{\psi}_A(v)\overline{x}_A(v)\overline{a}_A(v), \tag{98}$$

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

$$\tilde{F}_A(v)=\begin{cases}\overline{F}_A(v) & \forall \quad v\leq\underline{v}\\ \overline{F}_A(\underline{v}) & \forall \quad \underline{v}<v\leq\overline{v}\\ H(v)-\overline{F}_A(\underline{v}) & \forall \quad \overline{v}<v,\end{cases} \tag{99}$$

$$\tilde{F}_B(v)=\begin{cases}\overline{F}_B(v) & \forall \quad v\leq\underline{v}\\ H(v)-\overline{F}_B(\underline{v}) & \forall \quad \underline{v}<v\leq\overline{v}\\ 1 & \forall \quad \overline{v}<v,\end{cases} \tag{100}$$

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

$$1 + \tilde{F}_A(v) - \hat{F}_A(v) - \hat{F}_B(v) = 1 + H(v) - \overline{F}_A(\underline{v}) - H(v) > 0. \tag{101}$$

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

$$\psi_A(v)\overline{x}_B(v) - \overline{\psi}_A(v)\overline{a}_A(v)\overline{x}_A(v) - \overline{\psi}_B(v)\overline{x}_B(v)\overline{a}_B(v) \tag{102}$$

$$> \left(\psi_A(v) - \overline{\psi}_A(v)\overline{a}_A(v) - \overline{\psi}_B(v)\overline{a}_B(v)\right)\overline{x}_B(v) = 0, \tag{103}$$

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

$$\int_q \phi_A(q)y_A(q)dq + \phi_B(q)y_B(q)dq \tag{104}$$

Integrating each component by parts yields

$$v_i(q)qy_i(q)\Big|_0^1 + \int_q v_i(q)\left(-y_i'(q)\right)qdq \tag{105}$$

$$= v_i(1)y_i(1) + \int_C v_i(q)\left(-y_i'(q)\right)qdq + \sum_{q \in D} q\left[v^-(q)y^-(q) - v^+(q)y^+(q)\right] \tag{106}$$

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

$$\alpha_A x_A(\alpha_A) + \int_{V_A \setminus D} vq_A(v)x_A'(v)dv \tag{107}$$

$$+\alpha_B x_B(\alpha_B) + \int_{V_B \setminus D} vq_B(v)x_B'(v)dv + \sum_{v \in D} q_{A\setminus B}(v)v\left[x^+_{A\setminus B}(v) - x^-_{A\setminus B}(v)\right], \tag{108}$$

which allows us once more to directly account for gaps in the support. We use $A\backslash B$ 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

$$v^M \tilde{x}_A(v^M) + \int_{v^M}^{\omega} q_A(v)\tilde{x}'_A(v)vdv + \int_{\alpha}^{v^M} q_B(v)\tilde{x}'_B(v)vdv + \sum_{v \in D} q_{A\setminus B}(v)v\left[x^+_{A\setminus B}(v) - x^-_{A\setminus B}(v)\right], \tag{109}$$

$$> t_1\tilde{x}_B(t_1) + \int_{v^M}^{\omega} \tilde{q}_A(v)\tilde{x}'_A(v)vdv + \int_{\alpha}^{v^M} \tilde{q}_B(v)\tilde{x}'_B(v)vdv + \sum_{v \in D} \tilde{q}(v)v\left[\tilde{x}^+_{A\setminus B}(v) - \tilde{x}^-_{A\setminus B}(v)\right]. \tag{110}$$

Disjoint Support for $\hat{F}_A, \hat{F}_B$: $t_2 \leq v^M$ In this case inequality (109) can be amended to

$$v^M \tilde{x}_A(v^M) + \int_{v^M}^{t_3} q_A(v)\tilde{x}'_A(v)vdv + \int_{t_1}^{v^M} q_B(v)\tilde{x}'_B(v)vdv + \sum_{v \in D} q_{A\setminus B}(v)v\left[x^+_{A\setminus B}(v) - x^-_{A\setminus B}(v)\right], \tag{111}$$

$$> t_1\tilde{x}_B(t_1) + \int_{v^M}^{t_3} \tilde{q}_A(v)\tilde{x}'_A(v)vdv + \int_{t_1}^{v^M} \tilde{q}_B(v)\tilde{x}'_B(v)vdv + \sum_{v \in D} \tilde{q}_{A\setminus B}(v)v\left[\tilde{x}^+_{A\setminus B}(v) - \tilde{x}^-_{A\setminus B}(v)\right], \tag{112}$$

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

$$\int_{v^M}^{t_3} \left(q_A(v) - \hat{q}_B(v)\right)\tilde{x}'_A(v)vdv \tag{113}$$

where

$$q_A(v) = 2 - H(v) \tag{114}$$

while

$$\hat{q}_B(v) = 1 - \hat{F}_B = 1 - (H(v) - H(t_2) + H(t_1)) \tag{115}$$

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

$$q_A(v) - \hat{q}_B(v) = 1 - H(v) + (H(v) - H(t_2) + H(t_1)) = 1 - H(t_2) + H(t_1) > 0 \tag{116}$$

implying that the difference is constant. Expression (113) then becomes:

$$(1 - H(t_2) + H(t_1))\int_{v^M}^{t_3} \tilde{x}'_A(v)vdv \tag{117}$$

We then turn to

$$\int_{t_1}^{v^M} \left(q_B(v) - \tilde{q}_B(v)\right) \tilde{x}'_B(v) v dv \tag{118}$$

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

$$q_B(v) - \hat{q}_B(v) = 1 - H(v) - (1 - (H(v) - H(t_2) + H(t_1))) \tag{119}$$

$$= 1 - H(v) - 1 + (H(v) - H(t_2) + H(t_1)) = -\left(H(t_2) - H(t_1)\right) < 0 \tag{120}$$

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

$$q_B(v) - \hat{q}_A(v) = 1 - H(v) - (1 - (H(v) - H(t_1))) \tag{121}$$

$$= -H(v) + (H(v) - H(t_1)) = -H(t_1) \tag{122}$$

Expression (118) is then equivalent to

$$\int_{t_1}^{v^M} \left(q_B(v) - \tilde{q}_B(v)\right) \tilde{x}'_B(v) v dv = -\left(H(t_2) - H(t_1)\right) \int_{t_2}^{v^M} \tilde{x}'_B(v) v dv - H(t_1) \int_{t_1}^{t_2} \tilde{x}'_B(v) v dv \tag{123}$$

Collecting all terms leads to the following comparison

$$v^M \tilde{x}_A(v^M) + (1 - H(t_2) + H(t_1)) \int_{v^M}^{t_3} \tilde{x}'_A(v) v dv \tag{124}$$

$$-\left(H(t_2) - H(t_1)\right) \int_{t_2}^{v^M} \tilde{x}'_B(v) v dv - H(t_1) \int_{t_1}^{t_2} \tilde{x}'_B(v) v dv \tag{125}$$

$$+ (1 - H(t_2) + H(t_1)) \sum_{v \in D \cap [v^M, t_3]} v \left[\tilde{x}^+_A(v) - \tilde{x}^-_A(v)\right] \tag{126}$$

$$-\left(H(t_2) - H(t_1)\right) \sum_{v \in D \cap [t_2, v^M]} v \left[\tilde{x}^+_B(v) - \tilde{x}^-_B(v)\right] \tag{127}$$

$$-H(t_1) \sum_{v \in D \cap [t_1, t_2]} v \left[\tilde{x}^+_B(v) - \tilde{x}^-_B(v)\right] > t_1 \tilde{x}_B(t_1) \tag{128}$$

Rearranging yields

$$v^M \tilde{x}_A(v^M) + \int_{v^M}^{t_3} \tilde{x}'_A(v) v dv - \left(H(t_2) - H(t_1)\right) \int_{t_2}^{t_3} \tilde{x}'_{A \setminus B}(v) v dv - H(t_1) \int_{t_1}^{t_2} \tilde{x}'_B(v) v dv \tag{129}$$

$$+ \sum_{v \in D \cap [v^M, t_3]} v \left[\tilde{x}^+_A(v) - \tilde{x}^-_A(v)\right] - \left(H(t_2) - H(t_1)\right) \sum_{v \in D \cap [t_2, t_3]} v \left[\tilde{x}^+_{A \setminus B}(v) - \tilde{x}^-_{A \setminus B}(v)\right] \tag{130}$$

$$-H(t_1) \sum_{v \in D \cap [t_1, t_2]} v \left[\tilde{x}^+_B(v) - \tilde{x}^-_B(v)\right] > t_1 \tilde{x}_B(t_1) \tag{131}$$

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

$$t_3\tilde{x}_A(t_3) - \int_{v^M}^{t_3} \tilde{x}_A(v)dv - (t_3\tilde{x}_A(t_3) - t_1\tilde{x}_B(t_1)) + \int_{t_1}^{t_3} \tilde{x}(v)dv > \tilde{x}_B(t_1)t_1 \tag{132}$$

$$\Leftrightarrow \qquad \int_{t_1}^{v^M} \tilde{x}_B(v)dv > 0, \tag{133}$$

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

$$v^M\tilde{x}_A(v^M) + \int_{v^M}^{t_3} q_A(v)\tilde{x}'_A(v)vdv + \int_{t_1}^{v^M} q_B(v)\tilde{x}'_B(v)vdv \tag{134}$$

$$> t_1\tilde{x}_B(t_1) + \int_{v^M}^{t_3} \tilde{q}_A(v)\tilde{x}'_A(v)vdv + \int_{t_1}^{v^M} \tilde{q}_B(v)\tilde{x}'_B(v)vdv \tag{135}$$

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

$$\int_{t_2}^{t_3} q_A(v) - \hat{q}_B(v)\tilde{x}'_A(v)vdv = (1 - H(t_2) + H(t_1)) \int_{t_2}^{t_3} \tilde{x}'_A(v)vdv \tag{136}$$

We turn to

$$\int_{v^M}^{t_2} (q_A(v) - \hat{q}_A(v))\, \tilde{x}'_A(v)vdv \tag{137}$$

$$q_A(v) - \hat{q}_A(v) = 2 - H(v) - (1 - (H(v) - H(t_1))) \tag{138}$$

$$= 1 - H(v) + (H(v) - H(t_1)) = 1 - H(t_1) \tag{139}$$

Then,

$$\int_{v^M}^{t_2} (q_A(v) - \hat{q}_A(v))\, \tilde{x}'_A(v)vdv = (1 - H(t_1)) \int_{v^M}^{t_2} \tilde{x}'_A(v)vdv \tag{140}$$

Last,

$$\int_{t_1}^{v^M} (q_B(v) - \hat{q}_A(v))\, \tilde{x}'_A(v)vdv = -H(t_1) \int_{t_1}^{v^M} \tilde{x}'_B(v)vdv \tag{141}$$

Collecting terms yields

$$v^M \tilde{x}_A(v^M) + (1 - H(t_2) + H(t_1)) \int_{t_2}^{t_3} \tilde{x}'_A(v) v dv + (1 - H(t_1)) \int_{v^M}^{t_2} \tilde{x}'_A(v) v dv \tag{142}$$

$$- H(t_1) \int_{t_1}^{v^M} \tilde{x}'_B(v) v dv > t_1 \tilde{x}_B(t_1) \tag{143}$$

As before it is sufficient to show that

$$t_3 \tilde{x}_A(t_3) - \int_{v^M}^{t_3} \tilde{x}_A(v) dv - (t_3 \tilde{x}_A(t_3) - t_1 \tilde{x}_B(t_1)) + \int_{t_1}^{t_3} \tilde{x}_{A \setminus B}(v) dv > \tilde{x}_B(t_1) t_1 \tag{144}$$

$$\Leftrightarrow \quad \int_{t_1}^{v^M} \tilde{x}_B(v) dv > 0 \tag{145}$$

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

$$v^M \tilde{x}_A(v^M) + \int_{v^M}^{t_2} q_A(v) \tilde{x}'_A(v) v dv + \int_{t_1}^{v^M} q_B(v) \tilde{x}'_B(v) v dv \tag{146}$$

$$> t_1 \tilde{x}_B(t_1) + \int_{v^M}^{t_2} \tilde{q}_A(v) \tilde{x}'_A(v) v dv + \int_{t_1}^{v^M} \tilde{q}_B(v) \tilde{x}'_B(v) v dv \tag{147}$$

As before,

$$\int_{v^M}^{t_2} q_A(v) - \hat{q}_A(v) \tilde{x}'_A(v) v dv = (1 - H(t_1)) \int_{v^M}^{t_2} \tilde{x}'_A(v) v dv \tag{148}$$

$$\int_{t_1}^{v^M} q_B(v) - \hat{q}_A(v) \tilde{x}'_A(v) v dv = -H(t_1) \int_{t_1}^{v^M} \tilde{x}'_A(v) v dv \tag{149}$$

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.■

D Additional Empirical Results

Table 8: Summary Statistics

	Mean	S.D.	Min	Max	N
Gender, Race and Sexual Orientation					
Female	0.48	0.50	0	1	5210
Black	0.08	0.27	0	1	4219
Hispanic	0.07	0.25	0	1	4219
Native American	0.01	0.09	0	1	4219
Asian	0.09	0.28	0	1	4219
White	0.70	0.46	0	1	4219
Other Race	0.03	0.18	0	1	4219
LGBT+	0.03	0.17	0	1	4231
Personal Info					
Age	31.65	5.70	23	77	4239
Married	0.56	0.50	0	1	4185
# Kids	0.43	0.86	0	6	4206
Education					
Law S Rank	2.45	1.11	0	4	1852
Law S Quality	3.78	1.24	0	6	4260
Law Review	1.32	0.67	1	3	4149
UGrad GPA	5.20	1.46	0	7	2453
UGrad Rank	3.08	0.96	0	4	1281
UGrad Quality	3.65	1.59	1	6	4039
Dedication					
Intend Practice Law	1.78	0.53	0	2	4300
Consider Oth. Career	0.41	0.49	0	1	1773
Influential	2.91	1.29	1	5	2513
Loans	62029.51	41549.81	0	213000	4103
Loans []	4.80	2.44	1	8	4103
Firm Characteristics					
Firm Size	244.66	417.18	0	5000	2992
Firm Size []	5.07	3.47	1	11	2992
Office Size	85.77	127.55	0	850	2935
Office Size []	3.96	2.62	1	10	2935
% Men	64.75	22.33	0	100	1999
Private Firm	0.51	0.50	0	1	5210
Work Culture					
Demeaning	0.14	0.35	0	1	4126
Discrimination	0.11	0.32	0	1	4103
Rec. Time Partner	0.31	0.46	0	1	1763
Rec. Time Ass.	0.67	0.47	0	1	1763
Breakfast Partner	0.54	0.50	0	1	1763
Observations	5210				

Note: Race variables are indicators if someone identifies with given race, multiple mentions possible. *LGTB+* indicates if someone is not straight. Age is calculated as 2002-birth year. *Married* indicates if someone is married or lives in a domestic partnership. *# Kids* specifies number of kids. All eduction variables are coded such that a higher value is better. *Law School Rank* denotes law school class rank upon completion, *Law School Quality* refers to the quality of the law school. *Law Review* equals zero if someone was not involved, one if involved as member and two if leader. *UGrad GPA* specifies the GPA achieved in undergrad. *UGrad Rank* and *UGrad Quality* are undergrad equivalent of *Law S Rank* and *Law S Quality*, respectively. *Intend Practice Law* refers to whether respondent planned to practice law when entering law school: zero is yes, one not sure, two no. *Consider Other Career* is response to "Did you consider other career in addition or instead of law?" with zero equal to no and one yes. *Influential* is a categorical variable measuring the response to "How important was becoming influential in powerful position in your decision to attend law school?": measured from one to five, with one irrelevant and five very important. *Loans* measures total amount of education-related debt. *Loans []* divides loans into categories, with higher values corresponding to higher loans. *Firm Size* and *Office Size* measure the number of lawyers in the entire private firm and office, respectively. Both sizes are bracketed as well with higher values indicating higher number of lawyers. *% Men* refers to the proportion of lawyers in the workplace that are men. *Private Firm* is an indicator equal to one if the respondent is employed by a private law firm. *Demeaning* is an indictor of value one if the respondent received demeaning comments, *Discrimination* an indicator if the respondent received discriminatory comments. *Recreational Time Partner* indicates if the respondent spent recreational time with partners, *Recreational Time Associates* if the respondent spent recreational time with other associates. *Breakfast Partner* takes value one if the respondent has breakfast with partners.

Table 9: Summary Statistics: Waves 2 and 3

	Mean	S.D.	Min	Max	N
Wave 2 Variables					
Married	0.75	0.44	0	1	3587
# Kids	0.98	1.11	0	6	3536
Firm Size	262.62	921.33	0	30000	2783
Firm Size []	4.45	3.39	1	11	2783
Office Size	56.20	180.00	0	7200	2810
Office Size []	3.06	2.31	1	11	2810
% Men	62.73	24.54	0	100	2175
Expectation Partner	53.81	37.02	0	100	1366
Expectation []	2.53	1.27	1	4	1366
Aspiration	6.30	3.41	1	10	977
Hours Billed	29.89	14.83	0	67	1577
New Clients	0.60	0.49	0	1	1892
Hours Office	38.60	15.78	0	96	3630
Total Hours	46.58	17.15	0	135	3630
Wave 3 Variable					
Partner	0.12	0.32	0	1	5210
Observations	5210				

Note: *Married* indicates if someone is married or lives in a domestic partnership in 2007. *# Kids* specifies number of kids in 2007. *Firm Size* and *Office Size* measure the number of lawyers in the entire private firm and office the respondent works in in 2007, respectively. Both sizes are bracketed as well with higher values indicating higher number of lawyers. *% Men* refers to the proportion of lawyers in the workplace of 2007 that are men. *Expectation Partner* measures the probability a respondent assigns to becoming equity partner in 2007. This is also bracketed into 4 categories, with a higher value indicating a higher probability. *Aspiration* measures the aspiration to become equity partner, with one indicating no aspiration and ten very high aspirations. *Hours Billed* are the hours billed in a week, *Hours Office* are the weekly hours spent at the office, *Total Hours* is the sum of *Hours Office* and hours worked outside the office. *New Clients* indicates with one if the respondent brought in a new client in the past year. *Partner* indicates whether an individual is either an equity or non-equity partner in wave 3.

Table 10: Summary Statistics By Gender: Wave 2 and 3

	Men Mean	S.D.	Women Mean	S.D.	Difference Men-Women	p-value
Wave 2 Variables						
Married	0.78	0.41	0.71	0.45	0.07***	(0.00)
# Kids	1.12	1.20	0.84	0.98	0.28***	(0.00)
Firm Size	261.88	734.53	263.51	1105.60	-1.63	(0.96)
Firm Size []	4.45	3.44	4.45	3.33	-0.00	(0.98)
Office Size	52.56	122.38	60.46	229.66	-7.90	(0.27)
Office Size []	3.04	2.27	3.08	2.35	-0.04	(0.64)
% Men	71.16	21.63	52.92	24.09	18.24***	(0.00)
Expectation Partner	58.78	35.86	47.02	37.52	11.75***	(0.00)
Expectation []	2.69	1.24	2.31	1.26	0.38***	(0.00)
Aspiration	7.13	3.20	5.31	3.40	1.82***	(0.00)
Hours Billed	31.01	14.89	28.18	14.59	2.83***	(0.00)
New Clients	0.66	0.47	0.53	0.50	0.13***	(0.00)
Hours Office	42.15	13.43	34.71	17.19	7.45***	(0.00)
Total Hours	50.78	14.31	41.96	18.76	8.81***	(0.00)
Wave 3 Variable						
Partner	0.14	0.35	0.09	0.29	0.05***	(0.00)
Observations	2735		2475		5210	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: *Married* indicates if someone is married or lives in a domestic partnership in 2007. *# Kids* specifies number of kids in 2007. *Firm Size* and *Office Size* measure the number of lawyers in the entire private firm and office the respondent works in in 2007, respectively. Both sizes are bracketed as well with higher values indicating higher number of lawyers. *% Men* refers to the proportion of lawyers in the workplace of 2007 that are men. *Expectation Partner* measures the probability a respondent assigns to becoming equity partner in 2007. This is also bracketed into 4 categories, with a higher value indicating a higher probability. *Aspiration* measures the aspiration to become equity partner, with one indicating no aspiration and ten very high aspirations. *Hours Billed* are the hours billed in a week, *Hours Office* are the weekly hours spent at the office, *Total Hours* is the sum of *Hours Office* and hours worked outside the office. *New Clients* indicates with one if the respondent brought in a new client in the past year. *Partner* indicates whether an individual is either an equity or non-equity partner in wave 3.

Table 11: Aspirations: Are those who don't declare aspirations different?

	Dependent Variable: Aspiration Indicator		
Female	0.00696	0.0158	0.00730
	(0.0312)	(0.0331)	(0.0350)
LGBT+	-0.0150	-0.00916	-0.00283
	(0.106)	(0.105)	(0.110)
Age	-0.00646**	-0.00638**	-0.00549*
	(0.00307)	(0.00312)	(0.00312)
Married	0.00924	0.00907	0.000440
	(0.0338)	(0.0339)	(0.0347)
# Kids	0.0316	0.0327	0.0331
	(0.0202)	(0.0205)	(0.0204)
Demeaning		0.0760	0.0766
		(0.0494)	(0.0492)
Discrimination		-0.145***	-0.150***
		(0.0526)	(0.0528)
Rec. Time Partner		0.0220	0.0395
		(0.0360)	(0.0374)
Rec. Time Ass.		0.0392	-0.00801
		(0.0347)	(0.0386)
Breakfast Partner		0.0282	0.0235
		(0.0336)	(0.0338)
Private Firm			0.161***
			(0.0579)
% Men			-0.000108
			(0.000751)
Observations	985	985	985
Race Indicators	Yes	Yes	Yes
Education Controls	Yes	Yes	Yes
Dedication Controls	Yes	Yes	Yes
Firm Controls	No	No	Yes

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The dependent variable is an indicator equal to one if the respondent entered an aspiration level and zero if this is missing. Controls as defined in Table 8. Estimations include controls for Race, Education (*Law School Rank* and *UGrad Rank*), Dedication (*Intent Practice Law*, *Influential*, *Loans* []), and Firm Characteristics (*Firm Size* [], *Office Size* []). Estimated using OLS.

Table 12: Hours and Aspirations

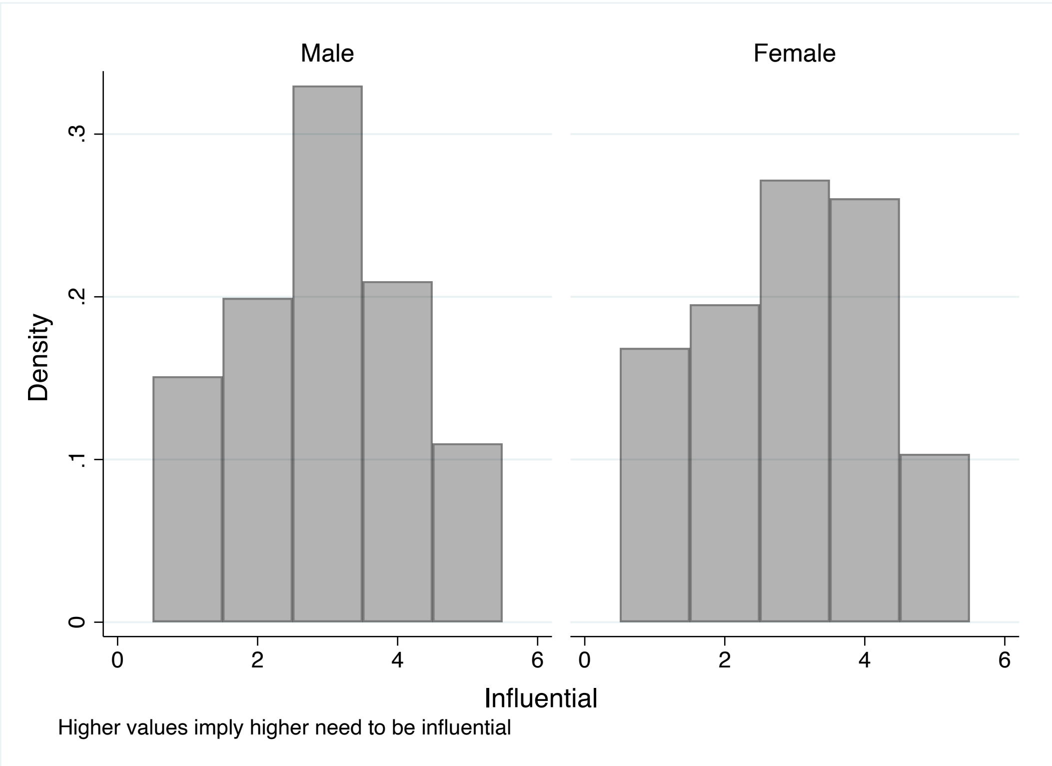
	Hours Office	Hours Office	Hours Office	Total Hours	Total Hours	Total Hours
Aspiration	0.847***	0.642***	0.597***	0.964***	0.732***	0.673***
	(0.125)	(0.147)	(0.175)	(0.130)	(0.153)	(0.182)
Female			-4.385***			-5.018***
			(0.951)			(0.967)
Age			0.0495			-0.00540
			(0.0977)			(0.104)
Married			1.699			-0.802
			(1.324)			(1.309)
# Kids			-2.063***			-1.683***
			(0.441)			(0.459)
% Men			0.0176			0.0138
			(0.0284)			(0.0299)
$[25, 50]$ % Partner			1.231			1.085
			(1.364)			(1.210)
$[50, 75]$ % Partner			1.971			0.782
			(1.576)			(1.565)
$[75, 100]$ % Partner			-0.328			-0.938
			(1.437)			(1.474)
Observations	977	640	640	977	640	640
R^2	0.051	0.033	0.161	0.063	0.040	0.180
Additional Controls	No	No	Yes	No	No	Yes

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: *Hours Office* measures weekly time at office. *Total Hours* is the sum of *Hours Office* and hours worked outside the office in a week. *Aspiration* measures the level of aspirations on a scale from one to ten. *Female* is an indicator equal to one if the respondent is a woman. *Married* equals one if the respondent is married or lives in a domestic partnership in Wave II. *% Men* refers to the share of men at the workplace in Wave II. $[x, y]$*% Partner* is an indicator for the percentage the respondent assigns to becoming partner. It takes value one if the respondent's percentage lies between $[x, y]$. In the third and last column we include as additional controls: *Bad Experience*, *Recreational Time Partner*, *Law S Quality* and from Wave II *Firm Size* [], *Office Size* [].

Figure 9: Distribution of Wanting to be Influential by Gender



Note: Influential measured on scale from 1-5, same sample as Aspiration

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.

Azmat, G., V. Cuñat, and E. Henry (2020). Gender promotion gaps: Career aspirations and workplace discrimination.

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.

Bulow, J. and J. Roberts (1989). The simple economics of optimal auctions. *The Journal of Political Economy 97, No. 5*, 1060–1090.

Cantillon, E. (2008). The effect of bidders' asymmetries on expected revenue in auctions. *Games and Economic Behavior 62*(1), 1–25.

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.

Deb, R. and M. M. Pai (2017). Discrimination via symmetric auctions. *American Economic Journal: Microeconomics 9*(1), 275–314.

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

Empson, L. (2017). *Leading professionals: Power, politics, and prima donnas.* Oxford University Press.

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

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.

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.

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.

Pérez-Castrillo, D. and D. Wettstein (2016). Discrimination in a model of contests with incomplete information about ability. *International economic review 57*(3), 881–914.

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.

Profitable Inequality: Supplementary Appendix

Anja Prummer* Francesco Nava†

October 28, 2021

Contents

*Johannes-Kepler-University Linz and Queen Mary University of London, anja.prummer@jku.at
†London School of Economics, f.nava@lse.ac.uk

A Summary Statistics: Other Samples

We provide summary statistics for workplace experiences across gender for the entire sample in Table A.1. Note that even though the number of observations differs from the Table 8 these are the same samples. For the selected sample, which we use to assess the connection between aspirations and work experiences, the summary statistics by gender are provided in A.2. It is evident that this is not a selected sample.

Table A.3 provides summary statistics for those working in private firms. while Table A.4 distinguishes by gender.

Table A.1: Experiences By Gender

	Men Mean	S.D.	Women Mean	S.D.	Difference Men-Women	p-value
Bad Exp.	0.10	0.30	0.28	0.45	-0.17***	(0.00)
Bad Exp. Intensity	0.14	0.43	0.39	0.68	-0.25***	(0.00)
Good Exp.	0.88	0.33	0.86	0.35	0.02	(0.33)
Good Exp. Intensity	1.56	0.96	1.47	0.94	0.10**	(0.03)
Time with Partner	0.66	0.47	0.56	0.50	0.10***	(0.00)
Time w/ Partner Intensity	0.92	0.77	0.78	0.78	0.14***	(0.00)
Rec. Time Partner	0.33	0.47	0.29	0.45	0.04*	(0.07)
Observations	2243		1908		4151	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: *Bad Experience* is an indicator equal one if the respondent received a demeaning or discriminatory comment. *Good Experience* is an indicator equal to one if the respondent spent recreational time with either partner or associates, or had breakfast with a partner. *Time With Partner* indicates whether the individual had breakfast with the partner or spent recreational time with them. *Recreational Time Partner* indicates if the respondent spent recreational time with a partner.*Bad Experience Intensity* equals one if the respondent either experienced discriminatory or demeaning comments and two if they experienced both. *Good Experience Intensity* equals one if the respondent spent recreational time with either partner or associates, or had breakfast with a partner, two if the respondent did two of the previous things and three if he had breakfast with the partner and spent recreational time both with associates and partners. *Time With Partner Intensity* equals one if the individual had breakfast with the partner or spent recreational time with them and two if both happened.

Table A.2: Experiences By Gender Restricted Sample

	Men Mean	S.D.	Women Mean	S.D.	Difference Men-Women	p-value
Bad Exp.	0.10	0.30	0.30	0.46	-0.20***	(0.00)
Bad Exp. Intensity	0.13	0.42	0.41	0.68	-0.27***	(0.00)
Good Exp.	0.94	0.24	0.88	0.33	0.06**	(0.04)
Good Exp. Intensity	1.77	0.94	1.59	0.97	0.18**	(0.05)
Time with Partner	0.71	0.45	0.62	0.49	0.09**	(0.04)
Time w/ Partner Intensity	1.02	0.77	0.89	0.80	0.12*	(0.10)
Rec. Time Partner	0.37	0.48	0.33	0.47	0.04	(0.36)
Observations	223		215		438	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: *Bad Experience* is an indicator equal one if the respondent received a demeaning or discriminatory comment. *Good Experience* is an indicator equal to one if the respondent spent recreational time with either partner or associates, or had breakfast with a partner. *Time With Partner* indicates whether the individual had breakfast with the partner or spent recreational time with them. *Recreational Time Partner* indicates if the respondent spent recreational time with a partner.*Bad Experience Intensity* equals one if the respondent either experienced discriminatory or demeaning comments and two if they experienced both. *Good Experience Intensity* equals one if the respondent spent recreational time with either partner or associates, or had breakfast with a partner, two if the respondent did two of the previous things and three if he had breakfast with the partner and spent recreational time both with associates and partners. *Time With Partner Intensity* equals one if the individual had breakfast with the partner or spent recreational time with them and two if both happened.

Table A.3: Summary Statistics: Private Firms

	Mean	S.D.	Min	Max	N
Gender, Race and Sexual Orientation					
Female	0.44	0.50	0	1	2672
Black	0.06	0.24	0	1	2592
Hispanic	0.06	0.24	0	1	2592
Native American	0.01	0.08	0	1	2592
Asian	0.08	0.28	0	1	2592
White	0.72	0.45	0	1	2592
Other Race	0.03	0.18	0	1	2592
LGBT+	0.02	0.15	0	1	2597
Personal Info					
Age	30.91	4.84	24	77	2608
Married	0.57	0.50	0	1	2571
# Kids	0.41	0.85	0	6	2589
Education					
Law S Rank	2.64	1.06	0	4	1122
Law S Quality	3.92	1.22	0	6	2612
Law Review	1.40	0.74	1	3	2556
UGrad GPA	5.32	1.44	0	7	1401
UGrad Rank	3.13	0.94	0	4	774
UGrad Quality	3.48	1.59	1	6	2496
Dedication					
Intend Practice Law	1.83	0.48	0	2	2640
Consider Oth. Career	0.38	0.49	0	1	1199
Influential	2.93	1.27	1	5	1437
Loans	63788.67	41099.97	0	200000	2516
Loans []	4.91	2.41	1	8	2516
Firm Characteristics					
Firm Size	262.10	411.55	1	4500	2648
Firm Size []	5.39	3.43	1	11	2648
Office Size	94.02	131.74	1	850	2609
Office Size []	4.20	2.61	1	10	2609
% Men	69.66	19.07	0	100	1363
Workplace Culture					
Demeaning	0.13	0.34	0	1	2638
Discrimination	0.10	0.30	0	1	2622
Rec. Time Partner	0.32	0.47	0	1	1209
Rec. Time Ass.	0.67	0.47	0	1	1209
Breakfast Partner	0.59	0.49	0	1	1209
Observations	2672				

Note: Race variables are indicators if someone identifies with given race, multiple mentions possible. *LGTB+* indicates if someone is not straight. Age is calculated as 2002-birth year. *Married* indicates if someone is married or lives in a domestic partnership. *# Kids* specifies number of kids. All eduction variables are coded such that a higher value is better. *Law School Rank* denotes law school class rank upon completion, *Law School Quality* refers to the quality of the law school. *Law Review* equals zero if someone was not involved, one if involved as member and two if leader. *UGrad GPA* specifies the GPA achieved in undergrad. *UGrad Rank* and *UGrad Quality* are undergrad equivalent of *Law S Rank* and *Law S Quality*, respectively. *Intend Practice Law* refers to whether respondent planned to practice law when entering law school: zero is yes, one not sure, two no. *Consider Other Career* is response to "Did you consider other career in addition or instead of law?" with zero equal to no and one yes. *Influential* is a categorical variable measuring the response to "How important was becoming influential in powerful position in your decision to attend law school?": measured from one to five, with one irrelevant and five very important. *Loans* measures total amount of education-related debt. *Loans* [] divides loans into categories, with higher values corresponding to higher loans. *Firm Size* and *Office Size* measure the number of lawyers in the entire private firm and office, respectively. Both sizes are bracketed as well with higher values indicating higher number of lawyers. *% Men* refers to the proportion of lawyers in the workplace that are men. *Private Firm* is an indicator equal to one if the respondent is employed by a private law firm. *Demeaning* is an indictor of value one if the respondent received demeaning comments, *Discrimination* an indicator if the respondent received discriminatory comments. *Recreational Time Partner* indicates if the respondent spent recreational time with partners, *Recreational Time Associates* if the respondent spent recreational time with other associates. *Breakfast Partner* takes value one if the respondent has breakfast with partners.

Table A.4: Summary Statistics By Gender: Private Firms

	Men Mean	S.D.	Women Mean	S.D.	Difference Men-Women	p-value
Race and Sexual Orientation						
Black	0.05	0.21	0.08	0.28	-0.04***	(0.00)
Hispanic	0.06	0.24	0.06	0.24	0.00	(0.81)
Native American	0.00	0.07	0.01	0.08	-0.00	(0.47)
Asian	0.07	0.26	0.10	0.30	-0.03***	(0.01)
White	0.75	0.43	0.68	0.46	0.07***	(0.00)
Other Race	0.03	0.18	0.03	0.17	0.00	(0.76)
LGBT+	0.02	0.16	0.02	0.15	0.00	(0.66)
Personal Info						
Age	30.98	4.69	30.83	5.04	0.14	(0.45)
Married	0.61	0.49	0.51	0.50	0.10***	(0.00)
# Kids	0.49	0.92	0.30	0.74	0.19***	(0.00)
Education						
Law S Rank	2.59	1.05	2.71	1.08	-0.12*	(0.06)
Law S Quality	3.93	1.23	3.91	1.21	0.02	(0.69)
Law Review	1.39	0.73	1.41	0.75	-0.01	(0.63)
UGrad GPA	5.11	1.51	5.55	1.33	-0.44***	(0.00)
UGrad Rank	3.00	1.01	3.29	0.83	-0.29***	(0.00)
UGrad Quality	3.44	1.59	3.54	1.59	-0.10	(0.12)
Dedication						
Intend Practice Law	1.84	0.46	1.81	0.49	0.02	(0.20)
Consider Oth. Career	0.38	0.49	0.39	0.49	-0.01	(0.62)
Influential	2.97	1.24	2.88	1.31	0.08	(0.21)
Loans	63256.50	40635.34	64473.73	41698.99	-1217.23	(0.46)
Loans []	4.87	2.40	4.95	2.44	-0.07	(0.46)
Firm Characteristics						
Firm Size	255.91	395.53	270.18	431.60	-14.28	(0.38)
Firm Size []	5.33	3.45	5.46	3.41	-0.14	(0.30)
Office Size	93.39	134.07	94.84	128.68	-1.44	(0.78)
Office Size []	4.16	2.64	4.26	2.59	-0.10	(0.32)
% Men	73.63	19.24	65.36	17.93	8.26***	(0.00)
Work Culture						
Demeaning	0.06	0.24	0.22	0.42	-0.16***	(0.00)
Discrimination	0.05	0.22	0.16	0.37	-0.11***	(0.00)
Rec. Time Partner	0.35	0.48	0.30	0.46	0.05*	(0.08)
Rec. Time Ass.	0.66	0.48	0.69	0.46	-0.04	(0.17)
Breakfast Partner	0.64	0.48	0.54	0.50	0.11***	(0.00)
Observations	1505		1167		2672	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Race variables are indicators if someone identifies with given race, multiple mentions possible. *LGTB+* indicates if someone is not straight. Age is calculated as 2002-birth year. *Married* indicates if someone is married or lives in a domestic partnership. *# Kids* specifies number of kids. All eduction variables are coded such that a higher value is better. *Law School Rank* denotes law school class rank upon completion, *Law School Quality* refers to the quality of the law school. *Law Review* equals zero if someone was not involved, one if involved as member and two if leader. *UGrad GPA* specifies the GPA achieved in undergrad. *UGrad Rank* and *UGrad Quality* are undergrad equivalent of *Law S Rank* and *Law S Quality*, respectively. *Intend Practice Law* refers to whether respondent planned to practice law when entering law school: zero is yes, one not sure, two no. *Consider Other Career* is response to “Did you consider other career in addition or instead of law?” with zero equal to no and one yes. *Influential* is a categorical variable measuring the response to “How important was becoming influential in powerful position in your decision to attend law school?”: measured from one to five, with one irrelevant and five very important. *Loans* measures total amount of education-related debt. *Loans []* divides loans into categories, with higher values corresponding to higher loans. *Firm Size* and *Office Size* measure the number of lawyers in the entire private firm and office, respectively. Both sizes are bracketed as well with higher values indicating higher number of lawyers. *% Men* refers to the proportion of lawyers in the workplace that are men. *Private Firm* is an indicator equal to one if the respondent is employed by a private law firm. *Demeaning* is an indictor of value one if the respondent received demeaning comments, *Discrimination* an indicator if the respondent received discriminatory comments. *Recreational Time Partner* indicates if the respondent spent recreational time with partners, *Recreational Time Associates* if the respondent spent recreational time with other associates. *Breakfast Partner* takes value one if the respondent has breakfast with partners.

B Aspirations and Work Experiences

In Table B.5 we consider the effect of the intensity of good and bad experiences at the workplace. Tables B.6 and B.7 confirm the findings of the linear probability model presented in Table 11.

Table B.5: Aspiration and Work Experiences: Intensity

	Dependent Variable: Aspiration						
Bad Exp. Intensity	-0.507*	-0.190	-0.505*	-0.183	-0.538*	-0.208	-0.0417
	(0.276)	(0.274)	(0.278)	(0.276)	(0.276)	(0.274)	(0.299)
Good Exp. Intensity	0.408**	0.348**					
	(0.170)	(0.165)					
Time w/ Partner Intensity			0.452**	0.397**			
			(0.202)	(0.195)			
Rec. Time Partner					0.866***	0.811**	0.782**
					(0.329)	(0.321)	(0.328)
Female		-1.590***		-1.605***		-1.613***	-1.390***
		(0.319)		(0.317)		(0.317)	(0.333)
Observations	438	438	438	438	438	438	406
R^2	0.023	0.076	0.021	0.075	0.025	0.080	0.122
Controls	No	No	No	No	No	No	Yes

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The dependent variable is *Aspiration* measured in Wave 2. Regressors are from Wave 1. *Bad Experience Intensity* equals one if the respondent either experienced discriminatory or demeaning comments and two if they experienced both. *Good Experience Intensity* equals one if the respondent spent recreational time with either partner or associates, or had breakfast with a partner, two if the respondent did two of the previous things and three if he had breakfast with the partner and spent recreational time both with associates and partners. *Time With Partner Intensity* equals one if the individual had breakfast with the partner or spent recreational time with them and two if both happened. *Recreational Time Partner* indicates if the respondent spent recreational time with a partner. The last column includes additional controls: race indicators, age, marital status in wave 2, number of kids in wave 2, the percentage of men at the workplace in wave 2.

Table B.6: Aspirations Indicator: Probit

	Dependent Variable: Aspiration Indicator		
Female	0.0239	0.0478	0.0260
	(0.0857)	(0.0904)	(0.0952)
LGBT+	-0.0420	-0.0273	-0.0136
	(0.310)	(0.307)	(0.325)
Age	-0.0199**	-0.0196**	-0.0172*
	(0.00961)	(0.00974)	(0.0103)
Married	0.0256	0.0246	0.00190
	(0.0929)	(0.0937)	(0.0967)
# Kids	0.0920*	0.0954*	0.101*
	(0.0554)	(0.0561)	(0.0580)
Demeaning		0.219	0.238*
		(0.137)	(0.140)
Discrimination		-0.429***	-0.455***
		(0.161)	(0.165)
Rec. Time Partner		0.0568	0.108
		(0.0980)	(0.104)
Rec. Time Ass.		0.117	-0.0120
		(0.0976)	(0.110)
Breakfast Partner		0.0880	0.0670
		(0.0930)	(0.0953)
Private Firm			0.723**
			(0.289)
% Men			-0.000329
			(0.00230)
Observations	985	985	985
Race Indicators	Yes	Yes	Yes
Education Controls	Yes	Yes	Yes
Dedication Controls	Yes	Yes	Yes
Firm Controls	No	No	Yes
Estimation Method	Probit	Probit	Probit

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The dependent variable is an indicator equal to one if the respondent entered an aspiration level and zero if this is missing. Controls as defined in Table 8. Estimations include controls for Race, Education (*Law School Rank* and *UGrad Rank*), Dedication (*Intent Practice Law*, *Influential*, *Loans* []), and Firm Characteristics (*Firm Size* [], *Office Size* []).

Table B.7: Aspirations Indicator: Logit

	Dependent Variable: Aspiration Indicator		
Female	0.0340	0.0717	0.0393
	(0.141)	(0.150)	(0.159)
LGBT+	-0.0752	-0.0442	-0.0525
	(0.518)	(0.510)	(0.547)
Age	-0.0324**	-0.0321*	-0.0292*
	(0.0162)	(0.0165)	(0.0177)
Married	0.0408	0.0430	0.00772
	(0.153)	(0.155)	(0.162)
# Kids	0.149	0.153*	0.164*
	(0.0905)	(0.0922)	(0.0968)
Demeaning		0.356	0.391*
		(0.226)	(0.230)
Discrimination		-0.708***	-0.774***
		(0.272)	(0.279)
Rec. Time Partner		0.0990	0.186
		(0.161)	(0.173)
Rec. Time Ass.		0.191	-0.0287
		(0.162)	(0.184)
Breakfast Partner		0.131	0.100
		(0.154)	(0.159)
Private Firm			1.308**
			(0.566)
% Men			-0.000424
			(0.00384)
Observations	985	985	985
Race Indicators	Yes	Yes	Yes
Education Controls	Yes	Yes	Yes
Dedication Controls	Yes	Yes	Yes
Firm Controls	No	No	Yes
Estimation Method	Logit	Logit	Logit

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The dependent variable is an indicator equal to one if the respondent entered an aspiration level and zero if this is missing. Controls as defined in Table 8. Estimations include controls for Race, Education (*Law School Rank* and *UGrad Rank*), Dedication (*Intent Practice Law*, *Influential*, *Loans* []), and Firm Characteristics (*Firm Size* [], *Office Size* []).

C Aspirations and Effort

We then connect whether a lawyer managed to bring a new client and aspirations using a logit and probit model. These different estimation techniques confirm our findings in Table 4, column (4)-(6).

Table C.8: New Clients and Aspirations: Probit and Logit

	New Clients	New Clients	New Clients	New Clients	New Clients	New Clients
New Clients						
Aspiration	0.137***	0.161***	0.107***	0.0854***	0.100***	0.0456**
	(0.0196)	(0.0251)	(0.0363)	(0.0121)	(0.0154)	(0.0227)
Female			-0.136			-0.0412
			(0.195)			(0.121)
Age			-0.00251			-0.00835
			(0.0212)			(0.0137)
Married			0.137			0.0283
			(0.251)			(0.160)
# Kids			-0.0365			-0.00517
			(0.0895)			(0.0589)
% Men			0.0116**			0.00826**
			(0.00572)			(0.00356)
$[25, 50]$ % Partner			0.339			0.277
			(0.287)			(0.180)
$[50, 75]$ % Partner			0.602*			0.481**
			(0.316)			(0.197)
$[75, 100]$ % Partner			0.885***			0.585***
			(0.298)			(0.187)
Observations	974	640	640	974	640	576
Additional Controls	No	No	Yes	No	No	Yes
Estimation Method	Logit	Logit	Logit	Probit	Probit	Probit

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: *New Clients* is an indicator equal to one if the respondent brought in a new client in the year before the second survey. *Aspiration* measures the level of aspirations on a scale from one to ten. *Female* is an indicator equal to one if the respondent is a woman. *Married* equals one if the respondent is married or lives in a domestic partnership in Wave II. *% Men* refers to the share of men at the workplace in Wave II. $[x, y]$*% Partner* is an indicator for the percentage the respondent assigns to becoming partner. It takes value one if the respondent's percentage lies between $[x, y]$. In the third and last column we include as additional controls: *Bad Experience*, *Recreational Time Partner*, *Law S Quality* and from Wave II *Firm Size* [], *Office Size* [].

D Aspirations, Effort and Promotions

Table D.9: Virtual Values and Becoming Partner: Probit

	Partner	Partner	Partner	Partner
Partner Virtual Value	0.0321***	0.0366***	0.0283***	0.0265***
	(0.00585)	(0.00719)	(0.00717)	(0.00727)
New Clients			0.327***	0.336***
			(0.106)	(0.114)
Hours Billed			0.0193***	0.0181***
			(0.00560)	(0.00655)
Hours Office			0.00472	0.00787
			(0.00548)	(0.00597)
Total Hours			0.00344	0.00164
			(0.00551)	(0.00604)
Female				-0.0660
				(0.117)
Age				-0.00157
				(0.0129)
Married				0.0441
				(0.150)
# Kids				0.0360
				(0.0553)
% Men				0.00257
				(0.00359)
Observations	977	657	657	657
Additional Controls	No	No	No	Yes

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: *Partner* is an indicator equal to one if the respondent is either equity or non-equity partner in Wave III. *Virtual value* is calculated according to the aspiration distribution generated in Figure 6 for each group. *Hours Billed*, *Hours Office*, *Total Hours* refer to billable hours, hours at the office and total hours per week in Wave II. *New Clients* is an indicator equal to one if respondent brought in a new client in Wave II. *Female* equals one if the respondent is a woman. *Age* measures 2002- birth year. *Married* equals one if the respondent is married or lives in a domestic partnership in Wave II. *% Men* refers to the share of men at the workplace in Wave II. In the third column we include as additional controls *Law S Quality* and from Wave II *Firm Size* [], *Office Size* [].

Table D.10: Aspirations, Info Rent and Becoming Partner: Probit

	Partner	Partner	Partner	Partner
Partner Aspiration	0.126***	0.150***	0.122***	0.141***
	(0.0175)	(0.0225)	(0.0236)	(0.0284)
Info Rent	0.0177*	0.0191	0.0150	0.0264*
	(0.00961)	(0.0117)	(0.0119)	(0.0135)
New Clients			0.276**	0.298**
			(0.108)	(0.116)
Hours Billed			0.0163***	0.0160**
			(0.00577)	(0.00678)
Hours Office			0.00265	0.00616
			(0.00567)	(0.00627)
Total Hours			0.00298	0.00170
			(0.00569)	(0.00635)
Female				0.171
				(0.134)
Age				0.000477
				(0.0133)
Married				0.0682
				(0.152)
# Kids				0.0174
				(0.0567)
% Men				0.00336
				(0.00362)
Observations	977	657	657	657
Additional Controls	No	No	No	Yes

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: *Partner* is an indicator equal to one if the respondent is either equity or non-equity partner in Wave III. *Aspiration* measures the level of aspirations to become equity partner on a scale from one to ten. *Info Rent* is calculated according to the aspiration distribution generated in Figure 6 for each group. *Hours Billed*, *Hours Office*, *Total Hours* refer to billable hours, hours at the office and total hours per week in Wave II. *New Clients* is an indicator equal to one if respondent brought in a new client in Wave II. *Female* equals one if the respondent is a woman. *Age* measures 2002- birth year. *Married* equals one if the respondent is married or lives in a domestic partnership in Wave II. *% Men* refers to the share of men at the workplace in Wave II. In the third column we include as additional controls *Law S Quality* and from Wave II *Firm Size* [], *Office Size* [].

Table D.11: Virtual Values and Becoming Partner: Probit Margins

Virtual Value	0.0108***	0.0129***	0.00954***	0.00854***
	(0.00188)	(0.00235)	(0.00232)	(0.00228)
New Clients			0.111***	0.108***
			(0.0350)	(0.0364)
Hours Billed			0.00651***	0.00584***
			(0.00185)	(0.00208)
Hours Office			0.00159	0.00254
			(0.00185)	(0.00193)
Total Hours			0.00116	0.000531
			(0.00186)	(0.00195)
Female				-0.0213
				(0.0378)
Age				-0.000506
				(0.00418)
Married				0.0143
				(0.0483)
# Kids				0.0116
				(0.0178)
% Men				0.000829
				(0.00116)
Observations	977	657	657	657
Additional Controls	No	No	No	Yes

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: *Partner* is an indicator equal to one if the respondent is either equity or non-equity partner in Wave III. *Virtual value* is calculated according to the aspiration distribution generated in Figure 6 for each group. *Hours Billed*, *Hours Office*, *Total Hours* refer to billable hours, hours at the office and total hours per week in Wave II. *New Clients* is an indicator equal to one if respondent brought in a new client in Wave II. *Female* equals one if the respondent is a woman. *Age* measures 2002- birth year. *Married* equals one if the respondent is married or lives in a domestic partnership in Wave II. *% Men* refers to the share of men at the workplace in Wave II. In the third column we include as additional controls *Law S Quality* and from Wave II *Firm Size* [], *Office Size* [].

Table D.12: Aspirations, Info Rent and Becoming Partner: Probit Margins

Aspiration	0.0413***	0.0508***	0.0404***	0.0445***
	(0.00535)	(0.00695)	(0.00742)	(0.00852)
Info Rent	0.00580*	0.00645	0.00495	0.00833**
	(0.00314)	(0.00396)	(0.00391)	(0.00424)
New Clients			0.0909***	0.0940***
			(0.0353)	(0.0362)
Hours Billed			0.00539***	0.00506**
			(0.00187)	(0.00211)
Hours Office			0.000873	0.00194
			(0.00187)	(0.00198)
Total Hours			0.000984	0.000537
			(0.00187)	(0.00200)
Female				0.0538
				(0.0421)
Age				0.000150
				(0.00418)
Married				0.0215
				(0.0480)
# Kids				0.00549
				(0.0179)
% Men				0.00106
				(0.00114)
Observations	977	657	657	657
Additional Controls	No	No	No	Yes

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: *Partner* is an indicator equal to one if the respondent is either equity or non-equity partner in Wave III. *Aspiration* measures the level of aspirations to become equity partner on a scale from one to ten. *Info Rent* is calculated according to the aspiration distribution generated in Figure 6 for each group. *Hours Billed*, *Hours Office*, *Total Hours* refer to billable hours, hours at the office and total hours per week in Wave II. *New Clients* is an indicator equal to one if respondent brought in a new client in Wave II. *Female* equals one if the respondent is a woman. *Age* measures 2002- birth year. *Married* equals one if the respondent is married or lives in a domestic partnership in Wave II. *% Men* refers to the share of men at the workplace in Wave II. In the third column we include as additional controls *Law S Quality* and from Wave II *Firm Size* [], *Office Size* [].

Table D.13: Virtual Values and Becoming Partner: Logit

	Partner	Partner	Partner	Partner
Partner				
Virtual Value	0.0555***	0.0637***	0.0484***	0.0449***
	(0.0103)	(0.0128)	(0.0127)	(0.0129)
New Clients			0.538***	0.538***
			(0.175)	(0.192)
Hours Billed			0.0328***	0.0320***
			(0.00998)	(0.0123)
Hours Office			0.00682	0.0123
			(0.00877)	(0.00991)
Total Hours			0.00578	0.00320
			(0.00887)	(0.00992)
Female				-0.0785
				(0.198)
Age				0.000368
				(0.0216)
Married				0.0651
				(0.250)
# Kids				0.0600
				(0.0930)
% Men				0.00402
				(0.00618)
Observations	977	657	657	657
Additional Controls	No	No	No	Yes

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: *Partner* is an indicator equal to one if the respondent is either equity or non-equity partner in Wave III. *Virtual value* is calculated according to the aspiration distribution generated in Figure 6 for each group. *Hours Billed*, *Hours Office*, *Total Hours* refer to billable hours, hours at the office and total hours per week in Wave II. *New Clients* is an indicator equal to one if respondent brought in a new client in Wave II. *Female* equals one if the respondent is a woman. *Age* measures 2002- birth year. *Married* equals one if the respondent is married or lives in a domestic partnership in Wave II. *% Men* refers to the share of men at the workplace in Wave II. In the third column we include as additional controls *Law S Quality* and from Wave II *Firm Size* [], *Office Size* [].

Table D.14: Aspirations, Info Rent and Becoming Partner: Logit

	Partner	Partner	Partner	Partner
Partner				
Aspiration	0.215***	0.254***	0.207***	0.243***
	(0.0310)	(0.0399)	(0.0416)	(0.0510)
Info Rent	0.0312*	0.0333*	0.0265	0.0466**
	(0.0161)	(0.0198)	(0.0201)	(0.0232)
New Clients			0.452**	0.472**
			(0.180)	(0.194)
Hours Billed			0.0284***	0.0285**
			(0.0102)	(0.0127)
Hours Office			0.00360	0.00988
			(0.00925)	(0.0106)
Total Hours			0.00535	0.00323
			(0.00930)	(0.0106)
Female				0.322
				(0.228)
Age				0.00468
				(0.0225)
Married				0.106
				(0.255)
# Kids				0.0302
				(0.0958)
% Men				0.00496
				(0.00623)
Observations	977	657	657	657
Additional Controls	No	No	No	Yes

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: *Partner* is an indicator equal to one if the respondent is either equity or non-equity partner in Wave III. *Aspiration* measures the level of aspirations to become equity partner on a scale from one to ten. *Info Rent* is calculated according to the aspiration distribution generated in Figure 6 for each group. *Hours Billed*, *Hours Office*, *Total Hours* refer to billable hours, hours at the office and total hours per week in Wave II. *New Clients* is an indicator equal to one if respondent brought in a new client in Wave II. *Female* equals one if the respondent is a woman. *Age* measures 2002- birth year. *Married* equals one if the respondent is married or lives in a domestic partnership in Wave II. *% Men* refers to the share of men at the workplace in Wave II. In the third column we include as additional controls *Law S Quality* and from Wave II *Firm Size* [], *Office Size* [].

Table D.15: Virtual Values and Becoming Partner: Logit Margins

Virtual Value	0.0113***	0.0136***	0.00986***	0.00871***
	(0.00196)	(0.00247)	(0.00245)	(0.00239)
New Clients			0.110***	0.104***
			(0.0348)	(0.0366)
Hours Billed			0.00668***	0.00621***
			(0.00198)	(0.00233)
Hours Office			0.00139	0.00239
			(0.00179)	(0.00192)
Total Hours			0.00118	0.000621
			(0.00180)	(0.00192)
Female				-0.0152
				(0.0385)
Age				0.0000714
				(0.00420)
Married				0.0126
				(0.0486)
# Kids				0.0117
				(0.0180)
% Men				0.000781
				(0.00120)
Observations	977	657	657	657
Additional Controls	No	No	No	Yes

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: *Partner* is an indicator equal to one if the respondent is either equity or non-equity partner in Wave III. *Virtual value* is calculated according to the aspiration distribution generated in Figure 6 for each group. *Hours Billed*, *Hours Office*, *Total Hours* refer to billable hours, hours at the office and total hours per week in Wave II. *New Clients* is an indicator equal to one if respondent brought in a new client in Wave II. *Female* equals one if the respondent is a woman. *Age* measures 2002- birth year. *Married* equals one if the respondent is married or lives in a domestic partnership in Wave II. *% Men* refers to the share of men at the workplace in Wave II. In the third column we include as additional controls *Law S Quality* and from Wave II *Firm Size* [], *Office Size* [].

Table D.16: Aspirations, Info Rent and Becoming Partner: Logit Margins

Aspiration	0.0423***	0.0520***	0.0413***	0.0460***
	(0.00568)	(0.00743)	(0.00787)	(0.00910)
Info Rent	0.00614*	0.00682*	0.00527	0.00881**
	(0.00317)	(0.00406)	(0.00401)	(0.00436)
New Clients			0.0901**	0.0892**
			(0.0353)	(0.0365)
Hours Billed			0.00567***	0.00539**
			(0.00198)	(0.00237)
Hours Office			0.000716	0.00187
			(0.00184)	(0.00201)
Total Hours			0.00107	0.000611
			(0.00185)	(0.00200)
Female				0.0608
				(0.0427)
Age				0.000886
				(0.00425)
Married				0.0200
				(0.0482)
# Kids				0.00571
				(0.0181)
% Men				0.000937
				(0.00118)
Observations	977	657	657	657
Additional Controls	No	No	No	Yes

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: *Partner* is an indicator equal to one if the respondent is either equity or non-equity partner in Wave III. *Aspiration* measures the level of aspirations to become equity partner on a scale from one to ten. *Info Rent* is calculated according to the aspiration distribution generated in Figure 6 for each group. *Hours Billed*, *Hours Office*, *Total Hours* refer to billable hours, hours at the office and total hours per week in Wave II. *New Clients* is an indicator equal to one if respondent brought in a new client in Wave II. *Female* equals one if the respondent is a woman. *Age* measures 2002- birth year. *Married* equals one if the respondent is married or lives in a domestic partnership in Wave II. *% Men* refers to the share of men at the workplace in Wave II. In the third column we include as additional controls *Law S Quality* and from Wave II *Firm Size* [], *Office Size* [].

E Comparison Samples

We provide the differences for the variables for each of the samples we consider in order to highlight that the samples are not systematically selected to generate our results. Wave I variables are presented in Table E.17, Wave II variables in Table E.18.

Table E.17: Summary Statistics By Gender: Comparison Samples

	All	Aspiration	Aspiration Control	Hours/Clients	Partner
		Difference b/w Men-Women Across Samples			
Race and Sexual Orientation					
Black	-0.04***	-0.06***	-0.06***	-0.05***	-0.06***
Hispanic	0.00	-0.01	-0.01	-0.02	-0.02
Native American	0.00	0.01	0.01	0.00	0.00
Asian	-0.02***	-0.02	-0.02	-0.01	-0.01
White	0.07***	0.06	0.05	0.09**	0.09***
Other Race	0.00	0.01	0.01	0.00	0.00
LGBT+	-0.00	-0.00	-0.01	-0.01	-0.01
Personal Info					
Age	0.49***	0.57	0.58	0.60*	0.62*
Married	0.08***	0.11**	0.11**	0.06	0.07*
# Kids	0.21***	0.25***	0.24***	0.29***	0.30***
Education					
Law S Rank	-0.05	-0.05	-0.02	-0.04	-0.06
Law S Quality	-0.04	0.01	0.04	-0.03	-0.04
Law Review	-0.00	-0.04	0.01	0.02	0.03
UGrad GPA	-0.48***	-0.37***	-0.39***	-0.54***	-0.53***
UGrad Rank	-0.29***	-0.18	-0.22*	-0.18	-0.19
UGrad Quality	-0.05	-0.19	-0.26*	-0.06	-0.05
Dedication					
Intend Practice Law	0.02	-0.03	-0.05	-0.01	-0.01
Consider Oth. Career	0.01				
Influential	0.08	-0.03	-0.02	-0.05	-0.02
Loans	-1500.56	-4693.43	-5584.33	-3867.63	-4227.24
Loans []	-0.10	-0.31	-0.38	-0.24	-0.26
Firm Characteristics					
Firm Size	-9.45	-33.33	-33.29	-27.40	-25.92
Firm Size []	-0.20	0.18	0.24	-0.02	-0.02
Office Size	-1.98	5.41	9.28	4.34	4.33
Office Size []	-0.13	0.10	0.18	0.10	0.08
% Men	13.06***	6.32***	5.92***	4.20**	4.49**
Private Firm	0.08***	0.01	0.01	0.02	0.02
Work Culture					
Demeaning	-0.15***	-0.17***	-0.19***	-0.18***	-0.18***
Discrimination	-0.10***	-0.11***	-0.11***	-0.11***	-0.11***
Rec. Time Partner	0.04*	0.04	0.05	0.06	0.05
Rec. Time Ass.	-0.04*	0.06	0.07	0.05	0.05
Breakfast Partner	0.10***	0.08*	0.07	0.09	0.09*
Observations	5210	438	406	640	657

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: These are differences between the different samples for the regressions. The header refers to the dependent variable of the regression. Race variables are indicators if someone identifies with given race, multiple mentions possible. *LGTB+* indicates if someone is not straight. Age is calculated as 2002-birth year. *Married* indicates if someone is married or lives in a domestic partnership. *# Kids* specifies number of kids. All eduction variables are coded such that a higher value is better. *Law School Rank* denotes law school class rank upon completion, *Law School Quality* refers to the quality of the law school. *Law Review* equals zero if someone was not involved, one if involved as member and two if leader. *UGrad GPA* specifies the GPA achieved in undergrad. *UGrad Rank* and *UGrad Quality* are undergrad equivalent of *Law S Rank* and *Law S Quality*, respectively. *Intend Practice Law* refers to whether respondent planned to practice law when entering law school: zero is no, one not sure, two yes. *Influential* is a categorical variable measuring the response to "How important was becoming influential in powerful position in your decision to attend law school?": measured from one to five, with one irrelevant and five very important. *Loans* measures total amount of education-related debt. *Loans* [] divides loans into categories, with higher values corresponding to higher loans. *Firm Size* and *Office Size* measure the number of lawyers in the entire private firm and office, respectively. Both sizes are bracketed as well with higher values indicating higher number of lawyers. *% Men* refers to the proportion of lawyers in the workplace that are men. *Private Firm* is an indicator equal to one if the respondent is employed by a private law firm. *Demeaning* is an indictor of value one if the respondent received demeaning comments, *Discrimination* an indicator if the respondent received discriminatory comments. *Recreational Time Partner* indicates if the respondent spent recreational time with partners, *Recreational Time Associates* if the respondent spent recreational time with other associates. *Breakfast Partner* takes value one if the respondent has breakfast with partners.

Table E.18: Summary Statistics By Gender: Comparison Samples Wave 2 and 3 Variables

	All	Aspiration	Aspiration Control	Hours/Clients	Partner
	Difference b/w Men-Women Across Samples				
Wave 2 Variables					
Married	0.07***	0.07*	0.06**	0.04	0.05
# Kids	0.28***	0.32***	0.38***	0.42***	0.43***
Firm Size	-1.63	-24.60	-43.70	-86.34	-85.80
Firm Size []	-0.00	0.13	0.27	0.09	0.05
Office Size	-7.90	-5.34	-5.53	-13.25	-13.22
Office Size []	-0.04	-0.08	-0.03	-0.19	-0.20
% Men	18.24***	1.81	3.08**	1.39	1.75
Expectation Partner	11.75***	12.90***	9.16***	8.88***	8.88***
Expectation []	0.38***	0.46***	0.31***	0.29***	0.29***
Aspiration	1.82***	1.70***	1.71***	1.79***	1.82***
Hours Billed	2.83***	2.73**	4.01***	4.27***	3.96***
New Clients	0.13***	0.11**	0.07*	0.07*	0.08**
Hours Office	7.45***	6.02***	5.02***	4.66***	4.70***
Total Hours	8.81***	7.17***	5.69***	5.54***	5.37***
Wave 3 Variable					
Partner	0.05***	0.05	0.08**	0.07*	0.07**
Observations	5210	438	754	640	657

Note: These are differences between the different samples for the regressions. The header refers to the dependent variable of the regression. *Married* indicates if someone is married or lives in a domestic partnership in 2007. *# Kids* specifies number of kids in 2007. *Firm Size* and *Office Size* measure the number of lawyers in the entire private firm and office the respondent works in in 2007, respectively. Both sizes are bracketed as well with higher values indicating higher number of lawyers. *% Men* refers to the proportion of lawyers in the workplace of 2007 that are men. *Expectation Partner* measures the probability a respondent assigns to becoming equity partner in 2007. This is also bracketed into 4 categories, with a higher value indicating a higher probability. *Aspiration* measures the aspiration to become equity partner, with one indicating no aspiration and ten very high aspirations. *Hours Billed* are the hours billed in a week, *Hours Office* are the weekly hours spent at the office, *Total Hours* is the sum of *Hours Office* and hours worked outside the office. *New Clients* indicates with one if the respondent brought in a new client in the past year. *Partner* indicates whether an individual is either an equity or non-equity partner in wave 3.