Incentives and Selection*

Henrique Castro-Pires† and George Georgiadis‡

February 25, 2026

Abstract

Performance pay shapes both effort and workforce composition. Conventional wisdom holds that stronger incentives attract disproportionately more high-skilled workers. We show this need not hold. In a model where workers differ in skill and outside options, steeper incentives improve selection only if high-skill applicants increase proportionally more than low-skill applicants; under plausible outside-option distributions, they may instead worsen selection. We derive a sufficient-statistics test: selection improves if and only if the share of applicants passing a screening test rises. Experimental variation in incentives characterizes selection effects, identifies the profit-maximizing local change, and, with additional structure, the optimal contract.

Keywords: Incentive contracts, performance pay, selection, moral hazard

*We are grateful to Eduardo Azevedo, Arjada Bardhi, Rahul Deb, Evgenii Fadeev, Michael Grubb, Johannes Horner, Jan Knoepfle, Erik Madsen, Lucas Maestri, Humberto Moreira, Canice Prendergast, Doron Ravid, Luis Rayo, Julia Salmi, João Thereze, as well as participants of the QMUL-City Theory Workshop, the Game Theory & Information Economics Conference at EPGE/FGV-Rio de Janeiro, the NBER Organizational Economics meeting, and seminar participants at Arizona State University, Boston College, Harvard Business School, University of Chicago, and University of Miami.

†Harvard University, Harvard Business School; hcastropires@hbs.edu

‡Northwestern University, Kellogg School of Management;g-georgiadis@kellogg.northwestern.edu

1 Introduction

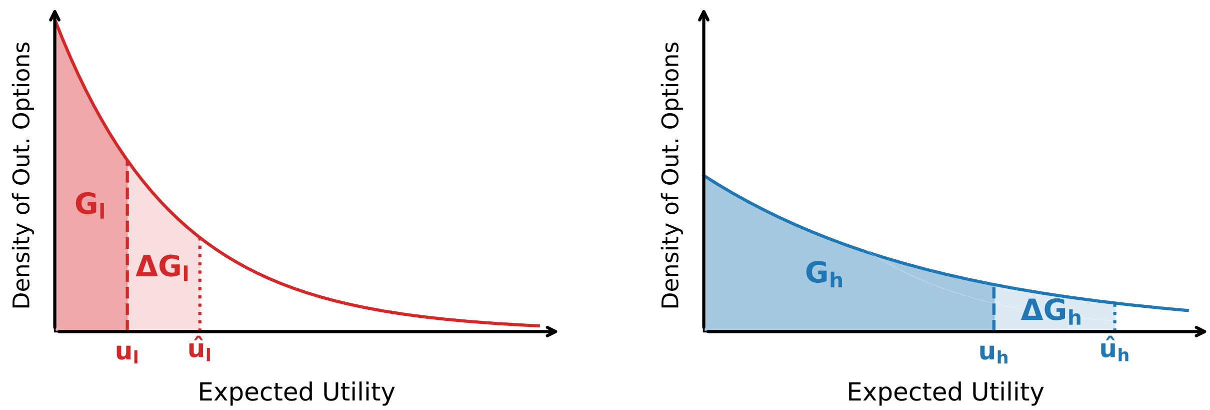
Designing the right incentive scheme is central to a firm's success. When a firm changes its compensation plan, it naturally affects how much workers produce (the "incentive effect"). Less obvious, but no less critical, is that the new compensation structure also shapes who applies for the job in the first place (the "selection effect"). Whether this new applicant pool is more or less skilled overall is not guaranteed, and understanding precisely when it improves the quality of the workforce and when it backfires is the central question of this paper.

A well-known empirical illustration of the role of incentives in labor markets is Lazear's (2000) classic study of Safelite, a windshield-repair firm that switched from hourly wages to piece rates. That switch spurred a 44% increase in productivity, with roughly half of this gain attributed to the firm attracting more productive windshield installers, and the other half attributed to existing installers working harder. In that setting, steeper incentives naturally favored higher-skilled workers, who could more profitably convert effort into output. Yet, this success story need not always hold: a higher-powered contract might improve everyone's payoffs enough that low-skilled workers flood in disproportionately, leading to negative selection. The question is: *how can we tell whether a shift toward higher-powered incentives will worsen or improve the skill composition of the workforce?*

To see how incentives can shape workforce composition, consider the following simplified setup. You manage a firm and compensate workers through an incentive contract: each worker exerts costly effort to generate output, and workers differ in skill. High-skilled workers face lower absolute and marginal costs of effort compared to low-skilled ones, and thus benefit differently from incentive changes. Prospective workers, whose outside options are drawn from type-dependent distributions, apply if your contract's expected payoff exceeds their outside option. You then administer a screening test, which high-skilled applicants pass with higher probability, and hire randomly from those who pass. Suppose you are now thinking of modifying the contract in a particular way. How will this affect *selection*; i.e., the share of high-skilled workers among those hired?

Figure 1 illustrates how modifying the contract influences each type's decision to apply, using an example where the modified contract increases both types' expected payoffs.[1]

[1]Here, (u_l, u_h) and $(\hat{u}_l, \hat{u}_h)$ denote low and high types' expected payoffs under the original and modified contracts, respectively.



(a) Low types (b) High types

Figure 1: The horizontal axis plots each type's expected payoff (u_t: original contract; $\hat{u}_t$: modified), while the vertical axis shows the density of outside options. The lightly shaded areas indicate the mass of new applicants drawn in by the modified contract.

Here, G_l and G_h mark the probability that a low-skilled or high-skilled worker, respectively, applies under the original contract, while ΔG_l and ΔG_h denote the additional mass of applicants drawn in by the new contract.

A key insight is that the modification improves selection if and only if

$$\frac{\Delta G_h}{G_h} > \frac{\Delta G_l}{G_l}. \tag{1}$$

That is, the *percentage* increase in high-skilled applicants must exceed that of low-skilled applicants. Importantly, even if the contract change benefits high-skilled workers more than low-skilled ones, it may still attract a larger proportional increase in low-skilled applicants. Ultimately, whether selection improves or worsens depends solely on how much each type benefits and how their outside options are distributed. Neither the fraction of high-skilled workers in the overall population nor the accuracy of the firm's screening test plays a role in determining this outcome.

This insight transcends our specific principal-agent model. The core mechanism described above applies whenever compensation adjustments affect relative rents, influencing the selection dynamics among different agent types. Crucially, this reasoning holds even if effort is contractible or motivation is driven by non-monetary factors. Ultimately, what governs selection outcomes are shifts in the relative attractiveness of each group resulting from compensation changes—not considerations like the ease of clearing a screening hurdle or the supply of highly skilled individuals.

If a firm can adjust incentives in a way that raises payoffs for high-skilled workers while lowering payoffs for low-skilled workers so that $\Delta G_h > 0 > \Delta G_l$, then the improvement in selection is straightforward (Theorem 1, part (*i*)). This clear-cut scenario, however, is theoretically less interesting. Moreover, designing precise contract "rotations" that affect payoffs in opposite directions requires detailed knowledge of worker preferences and production technologies that firms rarely have. In practice, incentive changes typically move the payoffs of both worker types in the same direction; indeed, most empirical evidence, including Lazear's (2000) study of Safelite, arises from scenarios where pay improved across the board. At Safelite, for example, the shift to piece rates included a wage floor that guaranteed no worker would be worse off, likely reflecting resistance by workers to reforms that threaten to reduce their earnings.

The second part of our first main result (Theorem 1, part (*ii*)) reveals a sharp negative possibility: *for any given baseline contract* and *any* adjustment that changes the utility of high and low-skill workers in the same direction, there always exist outside-option distributions that make selection worse. In other words, whenever a new scheme raises the payoffs of both high- and low-skilled workers, there is always a scenario where low-skilled applicants come in disproportionately, hurting workforce quality.

A practical hurdle to applying this theorem is that firms rarely observe detailed data on outside-option distributions and many other aspects of the environment. Hence, we propose a sufficient-statistics style test: namely, Remark 2 shows that a contract tweak improves selection if and only if the proportion of applicants who pass the firm's screening test increases. Intuitively, because high-skilled individuals pass the screen more often, a higher approval rate means that the contract adjustment attracted proportionally more high-skilled workers than low-skilled ones. This result does not require knowledge of screening-test accuracy, the overall fraction of high-skilled workers in the market, or even the shape of their outside-option distributions. We then show (Theorem 2) that with just one experiment—a single directional change in the contract—and observing how the total and rejected masses of applicants move, the firm can extrapolate how *any* local adjustment would affect selection.

After clarifying how local adjustments affect selection, we ask what is the *optimal* contract adjustment. Any adjustment to the contract affects profits through three channels: workers adjust their effort (the "incentive effect"), the workforce composition changes (the "selection effect"), and any direct change in wage costs (the "direct effect"). The direct effect does not depend on agents' responses and, hence, can be easily computed by the firm. However, to solve for the optimal local adjustment, the firm must also compute the other two effects. Theorem 3 demonstrates that, under an assumption about how effort translates into output, firms can identify the optimal adjustment.

In practice, firms are primarily interested in substantial contract modifications rather than only local adjustments. In Section 5, we formulate the problem of finding the optimal contract that incorporates selection effects. We specify a series of assumptions regarding the principal's informational environment and the functional forms of the workers' effort costs and their outside option distributions, and we show that, when coupled with empirical data from an experiment, they enable the firm to compute the optimal contract.

This paper contributes to the extensive literature on principal-agent problems under moral hazard. Since the seminal work of Holmström (1979) and Mirrlees (1999), the canonical framework has been extended to address a wide range of real-world complexities, including how moral hazard interacts with private information, see Georgiadis (2024) for a comprehensive review. A hallmark of this literature is that it treats the distribution of worker types as fixed, even when examining whether and how firms screen private information (Foarta and Sugaya, 2021; Castro-Pires, Chade and Swinkels, 2024) or choose not to screen (Ollier and Thomas, 2013; Castro-Pires and Moreira, 2021; Gottlieb and Moreira, 2022). In contrast, our paper endogenizes the composition of the applicant pool by studying how adjustments to an existing incentive scheme influence the distribution of workers who choose to apply.

Since Lazear (1986), the literature has recognized that stronger incentives can influence the composition of the workforce. However, most existing studies consider only the possibility that steeper incentives improve selection—see, for example, Moen and Åsa Rosén (2005), Gielen, Kerkhofs and van Ours (2010), Cornelissen, Heywood and Jirjahn (2011), and Wu (2017). We demonstrate that increasing the power of incentives can, in fact, worsen selection.

Several papers document productivity gains following an increase in incentives; e.g., Booth and Frank, 1999, Lazear, 2000, 2018, Paarsch and Shearer, 2000, Shearer, 2004, Bandiera, Barankay and Rasul, 2005, and Friebel et al., 2017. Yet, others find monetary incentives to be ineffective or even counterproductive; e.g., Leuven, Oosterbeek and van der Klaauw, 2010; Fryer, 2011, 2013; Bryson, Forth and Stokes, 2017, and Alfitian, Sliwka and Vogelsang, 2024. Existing explanations for why incentives might backfire include the crowding out of intrinsic motivation (e.g., Frey and Oberholzer-Gee, 1997; Kreps, 1997; Bénabou and Tirole, 2003; Casadesus-Masanell, 2004; Bénabou and Tirole, 2006; Conti et al., 2023), interactions with social norms (e.g., Gneezy and Rustichini, 2000; Sliwka, 2007), and social preferences or peer pressure (e.g., Hamilton, Nickerson and Owan, 2003; Ashraf and Bandiera, 2018). We show that negative effects on firm performance can also stem from worsening the firm's workforce composition even when agents are fully rational and motivated solely by pecuniary rewards.

This paper also relates to recent work on how monetary incentives affect employee selection. As in the productivity literature, empirical findings are mixed. For instance, Lo, Ghosh and Lafontaine (2011) and Dal Bó, Finan and Rossi (2013) report improved selection in sales and civil service settings, while Guiteras and Jack (2018) and Deserranno (2019) find no improvement—or even negative selection—in informal labor markets. We contribute to this literature in three ways. First, we identify a novel channel for negative selection that does not rely on intrinsic or pro-social motivation: selection can worsen due to the shape of outside option distributions. Second, we propose a simple test (Remark 2) to assess whether selection improves, offering a practical tool for empirical settings with limited data. Third, we develop a framework that allows researchers and firms to quantify the selection effect of any directional contract change using only the observed response to a single contract variation (Theorem 2). This approach enables counterfactual analysis of alternative incentive schemes without requiring multiple policy experiments, making it particularly useful in environments where contract variation is limited.

Finally, on methodological grounds, our work is connected to the literature using sufficient-statistics approaches via envelope conditions, which characterize behavioral responses and optimal policies with a few key parameters (typically elasticities); see Chetty (2009) for a review. This approach has a long history dating back to Harberger (1964) measuring deadweight losses of commodity taxes, and has been applied in income-taxation (Saez, 2001), corruption policy design (Ortner and Chassang, 2018), and welfare program evaluation (Finkelstein and Notowidigdo, 2019) among others. The closest paper to ours is

Georgiadis and Powell (2022), who bring these tools to analyze moral hazard problems. We extend this methodology by incorporating worker selection into the analysis, and deriving sufficient statistics for assessing when selection improves or worsens.

2 Model

There is a principal (also referred to as the firm) and a unit mass of agents (also referred to as the workers). The principal seeks to hire a fixed number of agents and motivate them to exert unobservable effort.

Events unfold in the following order:

i. The principal posts a mass $V > 0$ of identical job openings and a wage contract $w(\cdot)$, which is a bounded and upper-semicontinuous function mapping output x to monetary compensation.

ii. Each agent possesses a private type $t \in \{l, h\}$ with the proportion of high types ($t = h$) denoted by $p \in (0, 1)$. Conditional on his type, each agent draws his outside option $\overline{u}$ from a type-dependent distribution $G_t(\cdot)$, and then decides whether to apply for an opening.[2]

iii. The principal utilizes an imperfect screening mechanism, in which a type-t applicant successfully passes the screen with probability r_t. We assume that $r_h > r_l$, indicating that high types pass with strictly higher probability than low types. Subsequently, the principal randomly selects applicants who have passed the screening to fill the job vacancies.

iv. Each selected applicant then chooses an effort level $a \in [0, \overline{a}] \subset \mathbb{R}_+$. Individual output x is generated according to the probability density function $f(\cdot|a)$ that has support on some set $X \subseteq \mathbb{R}$ and payoffs are realized.

For a type-t agent who is hired, is paid y, and exerts effort a, the resulting payoff is $v(y) - c_t(a)$. Here, v is a strictly increasing, weakly concave, and twice continuously differentiable function, while c_t is strictly increasing, strictly convex, and twice continuously differentiable. Thus, a type-t agent applies whenever his expected payoff from

[2] In Appendix A we endogenize the outside option distributions using a microfoundation based on the skill-weights approach of Lazear (2009).

employment meets or exceeds his outside option:

$$u_t(w) := \max_a \left\{ \int (v \circ w)(x) f(x|a) dx - c_t(a) \right\} \geq \bar{u}.$$

From Bayes' rule, the share of high-types among hired agents, given the contract w, is

$$q(w) := \frac{r_h p (G_h \circ u_h)(w)}{r_h p (G_h \circ u_h)(w) + r_l (1-p)(G_l \circ u_l)(w)}. \tag{2}$$

The numerator represents the probability that a high-type agent applies and successfully passes the screening process, while the denominator captures the probability that a randomly selected agent applies and passes the screen.

Accordingly, the principal's expected profit per job position under contract w is given by

$$\pi(w) = \int \left[x - w(x)\right] \left[q(w) f\big(x|a_h(w)\big) + (1 - q(w)) f(x|a_l(w))\right] dx.$$

where $a_t(w)$ denotes the optimal effort chosen by a type-t agent.

We will study how the proportion of high-type agents among those hired, q, and the principal's payoff π respond to modifications of a given baseline contract w. A contractual adjustment is said to induce *positive selection* if it increases the share of high-type hires, as measured by $q(\cdot)$; conversely, it induces *negative selection* if this share decreases.

A pertinent analytical tool is the reverse hazard rate function,

$$\rho_t(u) := \frac{g_t(u)}{G_t(u)},$$

which quantifies the proportional increase in type-t applicants per additional util.

Finally, we impose the following assumptions:

i. High types possess weakly superior outside options; formally, G_h weakly first-order stochastically dominates G_l; that is, $G_h(u) \leq G_l(u)$ for all u. Both distributions are assumed to be continuously differentiable.

ii. High types incur strictly smaller absolute and marginal effort costs than low types;

that is, $c_l(a) > c_h(a)$ and $c_l'(a) > c_h'(a)$ for all $a > 0$.

iii. The density function $f(\cdot \mid a)$ has full support over the output space, a finite first moment, and is differentiable with respect to a. Moreover, higher effort strictly shifts the output distribution upward in the sense of first-order stochastic dominance. Without loss of generality, we normalize effort so that $a = \mathbb{E}[x \mid a]$, allowing effort to be interpreted as expected output.

iv. Under the baseline contract, strictly positive masses of both types apply and enough candidates are approved to fill all vacancies; formally, $(G_h \circ u_h)(w), (G_l \circ u_l)(w) > 0$ and $pr_h(G_h \circ u_h)(w) + (1-p)r_l(G_l \circ u_l)(w) > V$.

3 Steeper Incentives, Worse Selection?

Conventional wisdom suggests that higher-powered incentives increase the share of higher-ability workers because they will gain more under the new scheme than lower-ability workers. For example, when Safelite, a windshield repair firm, switched from hourly wages to piece rates, productivity soared by 44%, half of which was attributable to increased motivation and the other half to a larger share of more productive workers joining the firm (Lazear, 2000); i.e., improved selection. Here, we demonstrate that this conclusion need not always be true—higher-powered incentives may, in fact, harm selection.

Towards this goal, by dividing both the numerator and the denominator of (2) by $(G_l \circ u_l)(w)$, we make the following observation:

Remark 1. A change in the contract w leads to positive selection if and only if it increases

$$\frac{(G_h \circ u_h)(w)}{(G_l \circ u_l)(w)}. \tag{3}$$

This remark highlights that the fraction of high types in the population, as well as the precision of the principal's screen (i.e., r_l and r_h) are immaterial for whether a change in the contract improves or harms selection. On the other hand, the outside option distributions play a central role.

3.1 An Example

We illustrate the selection mechanism using the CARA–normal environment with affine contracts familiar from Holmstrom and Milgrom (1987). Agent payoffs are

$$u_t(w) = \max_a \left\{ -\int \exp\Big(-\xi \big[w(x) - c_t(a) \big] \Big)\, dF(x|a) \right\},$$

where $\xi > 0$ is the coefficient of absolute risk aversion,

$$c_t(a) = \frac{a^2}{2\theta_t}, \qquad w(x) = b + \alpha x, \quad \text{and} \quad x \sim \mathcal{N}(a, \sigma^2)$$

for some $\theta_l < \theta_h$. The optimal effort of a type-t agent is $a_t(w) = \alpha\theta_t$, which yields the certainty-equivalent (CE) utility

$$\tilde{u}_t(w) := b + \frac{\theta_t}{2}\alpha^2 - \frac{\xi}{2}\alpha^2\sigma^2. \tag{4}$$

Notice that this example doesn't possess the additive separability property of the model; however, separability is restored after the certainty-equivalent transformation. Therefore, we will define here the outside option distribution in terms of CE utilities, which is without loss of generality. For simplicity, we will assume $G_l \equiv G_h =: G$; that is, the outside option distributions are type-independent.

Raising the slope. Consider a marginal increase in the slope of the contract, α. This raises the CE utility of both types, thereby incentivizing more high types as well as low types to apply. Notice that the marginal impact $d\tilde{u}_t/d\alpha$ increases in θ_t; that is, high types benefit more from the increased slope than the low types. Of course, this alone does not guarantee improved selection. Indeed, according to Remark 1, selection *worsens* precisely when $(G \circ \tilde{u}_h)(w)/(G \circ \tilde{u}_l)(w)$ decreases in α. We have

$$\frac{d}{d\alpha}\frac{(G \circ \tilde{u}_h)(w)}{(G \circ \tilde{u}_l)(w)} =_s \frac{g(\tilde{u}_h)}{G(\tilde{u}_h)}\frac{d\tilde{u}_h}{d\alpha} - \frac{g(\tilde{u}_l)}{G(\tilde{u}_l)}\frac{d\tilde{u}_l}{d\alpha} =_s (\tilde{u}_h - b)\rho(\tilde{u}_h) - (\tilde{u}_l - b)\rho(\tilde{u}_l)$$

where $\rho(u) := g(u)/G(u)$ is the reverse hazard rate of G. Consequently, selection worsens if $(u - b)\rho(u)$ *decreases* in u.[3]

[3]If the wage level $b = 0$, this condition reduces to ρ having elasticity less than -1. On the other hand, selection is improved if $(u - b)\rho(u)$ *increases* in u. Note that "$=_s$" indicates that both sides have strictly the

Several distributions exhibit this property. For instance, for distributions with bounded support and decreasing reverse hazard rates (e.g., the uniform), $(u-b)\rho(u)$ is strictly decreasing whenever b lies below the lower bound of support. The exponential distribution also shares this property for all $u \geq b + 1/\lambda$ (and for all $u > 0$ when $b = 0$), as does the Pareto distribution, provided u is larger than some threshold. Finally, for log-concave distributions such as the Gaussian, the right tail of the Gaussian distribution demonstrates similar behavior, where $(u-b)\rho(u)$ decreases for sufficiently large u.

Raising the level. Next, consider a marginal increase in the base pay b. This adjustment raises the expected utility of both types by the same increment, so

$$\frac{d}{db}\frac{(G \circ \tilde{u}_h)(w)}{(G \circ \tilde{u}_l)(w)} =_s \rho(\tilde{u}_h) - \rho(\tilde{u}_l),$$

Hence, selection worsens if the reverse hazard rate ρ decreases in u, or equivalently, if the outside option distribution G is log-concave. Many common distributions satisfy this condition, including Gaussian, exponential, Pareto, and the extreme-value distribution with shape parameters greater than -1.

3.2 General Framework

In this section, we broaden the earlier insight to encompass a more general contracting scenario and pinpoint the conditions under which a marginal modification to any given contract might improve or harm selection.

To perform this analysis, it is necessary to explain how the share of high types among hired workers shifts as we locally modify the baseline contract w. Given a contract w and a function $h(w)$, we define the Gateaux differential of h in the direction ℓ by

$$\mathcal{D}h(w,\ell) := \lim_{\varepsilon \downarrow 0} \frac{h(w+\varepsilon\ell) - h(w)}{\varepsilon},$$

where the adjustment direction $\ell : X \to \mathbb{R}$ represents the difference between a modified contract $\hat{w}$ and the baseline contract w. Our primary focus is on how $q(w)$ responds to changes in the baseline contract w along direction ℓ, specifically whether $\mathcal{D}q(w,\ell)$ is positive.

same sign.

Definition. Adjusting w in direction ℓ *improves* selection if $\mathcal{D}q(w,\ell) > 0$. Conversely, this adjustment *harms* selection if $\mathcal{D}q(w,\ell) < 0$.

The following lemma provides a necessary and sufficient criterion for an adjustment to improve or harm selection.

Lemma 1. *An adjustment of w in direction ℓ improves (harms) selection if and only if*

$$(\rho_h \circ u_h)(w) \times \mathcal{D}u_h(w,\ell) \underset{(<)}{>} (\rho_l \circ u_l)(w) \times \mathcal{D}u_l(w,\ell), \tag{5}$$

where $\mathcal{D}u_t(w,\ell) = \int (v' \circ w)(x)\ell(x) f(x|a_t(w))dx$.

The interpretation is similar to (1): On the margin, adjusting the baseline contract in the direction of ℓ increases the probability that a type-t agent applies by $(\rho_t \circ u_t)(w)$ percentage points per util, and their utility by $\mathcal{D}u_t(w,\ell)$ utils. Thus, selection improves if (and only if) the percentage change of high-type applicants exceeds that of low-type applicants.

Any adjustment that impacts the payoffs of high types and low types in opposite directions has an unambiguous effect on selection. On the other hand, for any adjustment that impacts payoffs in the same direction, the following theorem shows that there exist outside option distributions for which selection is harmed.

Theorem 1. *Consider an arbitrary adjustment of the baseline contract w along the direction of ℓ.*

(i) If $\mathcal{D}u_h(w,\ell) > 0 > \mathcal{D}u_l(w,\ell)$, then this adjustment improves selection, and conversely, it harms selection if the inequalities are reversed.

(ii) If $\mathcal{D}u_h(w,\ell) \times \mathcal{D}u_l(w,\ell) > 0$, then there exist distributions G_h and G_l such that $G_h \succ_{fosd} G_l$ for which a local adjustment in direction ℓ worsens selection.

Part (i) stems directly from Lemma 1: If the adjustment raises the expected utility of high types while decreasing that of low types, the firm will attract more high-type applicants and fewer low-type ones, thereby improving selection.

Part (ii) shows that for any such adjustment that changes the utility of both types in the same direction, one can always construct outside option distributions—specifically type-I generalized extreme value ones—such that the adjusted contract attracts proportionally more low types than high types, thereby worsening selection. In particular, note that this statement holds even if high types benefit more than low types, meaning that their gain in utility is larger than the gain for the low types.

Theorem 1 implies that, absent any information about outside option distributions, the only contractual changes that can guarantee an improvement in selection are those that increase the expected utility of high types while reducing that of low types. An intuitively appealing candidate is a *rotation* of the contract, whereby the firm raises the slope of the contract while commensurately lowering base pay. We examine below whether this intuition is in fact sufficient to ensure improved selection.

Rotating the contract. We now revisit the CARA–Normal example from Section 3.1 to examine whether rotations improve selection.[4] Consider the perturbed contract

$$w_\varepsilon(x) := (b - \delta\varepsilon) + (\alpha + \varepsilon)x,$$

where δ governs the rate at which the wage is reduced for every unit that the slope is increased. Differentiating the CE utility of a type-t agent given in (4) with respect to ε and evaluating the derivative at $\varepsilon = 0$ yields the marginal effect

$$\left.\frac{d\tilde{u}_t(w_\varepsilon)}{d\varepsilon}\right|_{\varepsilon=0} = a_t(w) - \delta - \xi\alpha\sigma^2.$$

Intuitively, if δ is too small, both types benefit from the rotation; if it is too large, both types are harmed. In both cases, per Theorem 1(ii), selection may worsen. Improving selection (for all outside option distributions) requires choosing the wage adjustment so that low types are worse off while high types are better off. Formally, this is the case if and only if

$$\left.\frac{d\tilde{u}_l(w_\varepsilon)}{d\varepsilon}\right|_{\varepsilon=0} \leq 0 \leq \left.\frac{d\tilde{u}_h(w_\varepsilon)}{d\varepsilon}\right|_{\varepsilon=0}, \text{ or equivalently, } \quad a_l(w) \leq \delta + \xi\alpha\sigma^2 \leq a_h(w). \tag{6}$$

Observe that if agents are risk neutral (i.e., $\xi = 0$) and we rotate the contract around the average output (i.e., $\delta = q(w)a_h(w) + (1 - q(w))a_l(w)$), then selection necessarily improves

[4]We are grateful to Dan Barron and Jeroen Swinkels for suggesting this exercise.

(since $a_h(w) > a_l(w)$). Intuitively, this rotation ensures that—absent effort responses—the average payment across types remains unchanged. However, because the rotation increases payments for high output realizations and decreases payments for low ones, it must strictly benefit high types and strictly harm low types.

With risk-averse agents, however, increasing the slope also raises exposure to risk. Therefore, if agents are sufficiently risk averse, output is sufficiently volatile, or the effort gap between high and low types is small, then (6) fails.

Taken together, this analysis highlights two caveats about affecting selection through steeper incentives. First, even contract rotations do not, in general, guarantee improved selection. Whether they succeed depends on how incentive changes interact with agents' effort responses, risk exposure, and outside option distributions. Second, in practice, compensation changes often move expected utility in the same direction for all workers, rather than deliberately disadvantaging some types. Fearing pushback from employees, firms may avoid reforms that render a subgroup worse off. For example, when Safelite switched from hourly wages to piece rates, the initial plan was to reduce the wage floor by 30%. However, concerned about a surge in turnover and unionization threats, the firm ultimately kept the wage floor at the previous level (Hall, Lazear and Madigan, 2000). These considerations motivate a natural question: given a baseline contract, what must the firm observe in order to determine which adjustments improve or harm selection?

4 What Must Be Observed to Assess Selection

This section provides a theoretical characterization of the information that must be observable to assess how adjustments to the baseline contract affect selection. In Section 4.1, we specify the minimal data necessary to empirically determine whether a modification to the baseline contract improves or harms selection. Section 4.2 delineates the data required to predict the selection effect of *any* potential adjustment to the contract. Finally, Section 4.3 completes the empirical characterization by identifying the data needed to pinpoint the optimal adjustment.

4.1 A Sufficient Statistic for Improved Selection

The following observation pinpoints observable conditions to assess whether a particular adjustment to the baseline contract improves or harms selection. For this purpose, let us represent by $A(w)$ the share of applicants who successfully pass the screening test.[5] The following remark shows that $A(w)$ is a sufficient statistic for determining whether a change to the baseline contract improved or harmed selection.

Remark 2. The share of high types among hired workers is

$$q(w) = \frac{r_h}{r_h - r_l} \times \frac{A(w) - r_l}{A(w)}.$$

Therefore, a modification to the baseline contract has improved (worsened) selection if it caused $A(w)$ to increase (decrease).[6]

Intuitively, if a modification to the baseline contract leads to an increase in the passing rate, it must have attracted more high types than low types as the former have a greater chance of passing the screen. Note that having this knowledge prior to training the new hires and putting them to work is valuable, as it gives the firm an opportunity to take corrective action before bearing these sizable costs.

An implication of this observation is that to assess whether selection has improved does not require knowledge of many of the model's primitives, including the accuracy of the screening technology, the outside option distributions, the prevalence of each type in the market, the workers' utility function, or the effort-related costs specific to each type. Moreover, to quantify the change in selection, all that the principal needs to know in addition to the share of applicants who pass the screen is the accuracy of the screening technology.

4.2 Selection Effect of *every* Adjustment

This section characterizes the data and knowledge required to evaluate the selection impact of any adjustment to the baseline contract; formally, to compute $\mathcal{D}q(w, \ell)$ for an arbitrary direction ℓ. By Lemma 1, this requires identifying both the utility effect $\mathcal{D}u_t(w, \ell)$

[5] Formally, $A(w) := S(w)/T(w)$, where $T(w) := p(G_h \circ u_h)(w) + (1-p)(G_l \circ u_l)(w)$ denotes the *total* mass of applicants and $S(w) := r_h p(G_h \circ u_h)(w) + r_l(1-p)(G_l \circ u_l)(w)$ denotes the mass of passing applicants.

[6] Note that $A(w) \in [r_l, r_h]$, and it equals the lower (upper) bound if all applicants are low (high) types.

and the induced change in applicant masses $(\rho_t \circ u_t)(w)$.

We will assume that the firm observes outcome data under two contracts: the baseline contract w, as well as a local adjustment of w.

Assumption 1. *The firm observes, under the baseline contract w, the output distributions generated by each worker-type, $f(\cdot|a_l(w))$ and $f(\cdot|a_h(w))$, the total mass of applicants, $T(w)$, as well as the mass who successfully passes the screening, $S(w)$.*

This implicitly assumes that the firm knows the proportion of high types in its workforce; e.g., by observing longitudinal outcome data for each employee and then using a clustering algorithm to group them into high and low types. Under the adjustment, we assume that the firm observes only aggregate outcome data. The distribution and the expectation of aggregate output are, respectively, defined as

$$\bar{f}(x|w) := q(w)f\big(x|a_h(w)\big) + (1-q(w))f(x|a_l(w)), \text{ and}$$
$$\bar{a}(w) := q(w)a_h(w) + (1-q(w))a_l(w).$$

Experiment 1. *The firm observes under an adjustment of w along the direction of ℓ_1, the change in the distribution of overall output $\mathcal{D}\bar{f}(x|w,\ell_1)$, and hence $\mathcal{D}\bar{a}(w,\ell_1)$, as well as the changes in the total mass of applicants, $\mathcal{D}T(w,\ell_1)$ and the mass of those passing the screening, $\mathcal{D}S(w,\ell_1)$.*

Next, we impose the following assumption about the firm's knowledge, which enables it to infer $\mathcal{D}u_t(w,\ell)$ for every adjustment direction ℓ, as well as the necessary applicant-mass terms given the data at its disposal.

Assumption 2. *The firm knows the workers' monetary utility function, $v : \mathbb{R} \to \mathbb{R}$, and the precision of its screening technology, r_l and r_h.*

The next theorem shows that, given Assumptions 1 and 2 and data from Experiment 1, it is possible to determine the selection consequences of *any* potential adjustment.

Theorem 2. *Suppose Assumptions 1 and 2 hold, the firm observes the outcome data from Experiment 1, and $\mathcal{D}u_t(w,\ell_1) \neq 0$ for each $t \in \{l,h\}$. Then, for any adjustment direction ℓ, the change* *in selection is given by:*

$$\mathcal{D}q(w,\ell) = K \times \int (v' \circ w)(x)\ell(x)\big[\mathcal{E}_h f(x|a_h(w)) - \mathcal{E}_l f(x|a_l(w))\big]dx,$$

where the coefficients $K > 0$, $\mathcal{E}_h := (\rho_h \circ u_h)(w)$, and $\mathcal{E}_l := (\rho_l \circ u_l)(w)$ are identifiable from observable data.[7]

Theorem 2 implies that data from a single local experiment is sufficient to identify the selection effect of *any* local contract adjustment. The integrand comprises three components. First, $(v' \circ w)(x)\ell(x)$ translates the proposed payment adjustment at each output level x into a marginal change in utility. Second, the elasticities $\mathcal{E}_h$ and $\mathcal{E}_l$ convert this utility change into proportional adjustments in application rates for high and low types, respectively. Third, the densities $f(x|a_t(w))$ reflect the probability of each type producing output x. By subtracting the low-type contribution from the high-type at each x and integrating, we obtain the net change in the applicant composition. Thus, $\mathcal{D}q(w,\ell)$ is positive exactly when the adjustment raises high-type applications proportionally more than low-type applications, thereby improving selection.

While Theorem 2 provides a complete characterization of the selection effect of an arbitrary adjustment, it also highlights that the impact of a given direction ℓ depends on underlying primitives—most notably the marginal utility term $(v' \circ w)(x)$, which captures workers' risk attitudes. Although Assumption 2 treats v as known, the representation can be implemented quantitatively even when v is specified only up to a parametric family. For any candidate utility function, the observable objects from Experiment 1 enable the firm or researcher to compute the elasticities $\mathcal{E}_t$ and evaluate $\mathcal{D}q(w,\ell)$. One can therefore restrict attention to economically relevant preference classes—such as CRRA, CARA, or, more generally, HARA—and examine how the predicted selection effect varies across parameter values. This approach makes it possible to report ranges of implied selection effects or to identify the parameter regions under which a given adjustment improves selection.

[7] The expressions for these coefficients in terms of observable data are provided in the appendix.

5 Optimal Adjustment

Thus far, we have focused on how adjustments to the baseline contract influence worker selection, and we have specified the necessary information for this assessment. In this section, we first turn our attention to the profit-maximizing local adjustment, and later to optimal nonlocal contract modifications.

5.1 Local Adjustments

Suppose the firm adjusts the baseline contract w in some direction ℓ. The resulting impact on its profits can be broken down into three components:

$$
\begin{aligned}
&\mathcal{D}\pi(w,\ell) = \\
&\quad \underbrace{\int [x - w(x)][q(w) f_a(x|a_h(w)) \mathcal{D}a_h(w,\ell) + (1-q(w)) f_a(x|a_l(w)) \mathcal{D}a_l(w,\ell)]dx}_{\text{Incentive effect}} \\
&\quad + \underbrace{\left[\int [x - w(x)][f(x|a_h(w)) - f(x|a_l(w))]dx\right] \mathcal{D}q(w,\ell)}_{\text{Selection effect}} \\
&\quad - \underbrace{\int \ell(x)[q(w) f(x|a_h(w)) + (1-q(w)) f(x|a_l(w))]dx}_{\text{Direct effect}}.
\end{aligned}
$$

The incentive effect captures how profits change when workers adjust their efforts holding the composition of the workforce constant. The selection effect reflects the profit change due to shifts in workforce composition holding efforts fixed. Finally, the direct effect quantifies the immediate cost associated with adjusting payments along the direction of ℓ. We shall explore the information needed to solve

$$\max_{\ell} \mathcal{D}\pi(w,\ell) \text{ subject to } \|\ell\|_2 \leq 1. \tag{Local-P}$$

To solve this problem, the firm must be able to evaluate all three effects for any ℓ. The direct effect can be evaluated using Assumption 1. Theorem 2 has shown that Assumptions 1-2 coupled with Experiment 1 suffice to evaluate the selection effect. What remains is to determine the incentive effect for each ℓ.

To accomplish this, we introduce an additional assumption, stipulating that the output distribution varies linearly with effort, ensuring that marginal incentives are unaffected by the effort level.

Assumption 3. *The output distribution $f(x|a)$ is affine in a, that is, $f(x|a) = f_0(x) + af_a(x)$ for some $f_a(x)$ and $f_0(x)$ satisfying $\int f_a(x)dx = 0$ and $\int f_0(x)dx = 1$.*

This assumption serves multiple purposes. First, it enables the inference that

$$f_a(x) = \frac{\mathcal{D}\bar{f}(x|w, \ell_1)}{\mathcal{D}\bar{a}(w, \ell_1)}$$

by observing only how the aggregate output distribution shifts when the baseline contract is adjusted along the direction of ℓ_1. Second, it ensures that each agent's effort choice can be fully characterized by its first-order condition, and that $a_t(w)$ is Gateaux-differentiable.

And third, it implies that evaluating the incentive effect requires knowledge only of the average effort response for a given $q(w)$, rather than type-specific effort responses. To see this, observe that

$$\mathcal{D}\bar{a}(w, \ell) = \Big[q(w)\mathcal{D}a_h(w, \ell) + [1 - q(w)]\mathcal{D}a_l(w, \ell)\Big] + \Big[a_h(w) - a_l(w)\Big]\mathcal{D}q(w, \ell). \qquad (7)$$

Experiment 1 provides the means to determine $\mathcal{D}\bar{a}(w, \ell_1)$, and by Theorem 2, the second term on the right-hand side of (7). Thus, by Assumption 1, we can identify the first term for $\ell = \ell_1$. Next, we demonstrate how Assumption 2 enables the firm to evaluate the first term—which determines the incentive effect—for every ℓ.

When the baseline contract is adjusted in the direction of ℓ, the change in effort for a type-t agent's effort is given by

$$\mathcal{D}a_t(w, \ell) = \frac{\int \ell(x)(v' \circ w)(x) f_a(x)dx}{c_t''(a_t(w))}.$$

Therefore, we can express

$$\frac{q(w)\mathcal{D}a_h(w, \ell) + [1 - q(w)]\mathcal{D}a_l(w, \ell)}{q(w)\mathcal{D}a_h(w, \ell_1) + [1 - q(w)]\mathcal{D}a_l(w, \ell_1)} = \frac{\int \ell(x)(v' \circ w)(x) f_a(x)dx}{\int \ell_1(x)(v' \circ w)(x) f_a(x)dx}. \qquad (8)$$

As previously explained, Experiment 1 coupled with Assumptions 1-3 provides the denominator on the left-hand side and the necessary information to compute the right-hand side for any ℓ. Consequently, we can determine the numerator on the left-hand side, and thus, the incentive effect for each ℓ.

The following theorem shows that Experiment 1 coupled with Assumptions 1-3 suffice to solve (Local-P) to obtain the profit-maximizing adjustment to the baseline contract.

Theorem 3. *Let Assumptions 1, 2 and 3 hold, and assume* $\int \ell_1(x)(v' \circ w)(x) f_a(x) dx \neq 0$. *The information obtained from Experiment 1 suffices to solve* (Local-P). *Moreover, the optimal adjustment*

$$\ell^*(x) \propto \Big(\lambda^* \big[\mathcal{E}_h f(x|a_h) - \mathcal{E}_l f(x|a_l)\big] + \mu^* f_a(x)\Big) v'\big(w(x)\big) - \bar{f}(x|a(w)),$$

where $\lambda^*, \mu^* \geq 0$ *depend only on observables.*

The optimal local adjustment resembles the one in Georgiadis and Powell (2022), where selection is not considered. The main difference here is the term multiplying λ^*, which captures how local changes affect the firm's workforce composition. Specifically, since $\mathcal{E}_h = (\rho_h \circ u_h)(w)$ and $\mathcal{E}_l = (\rho_l \circ u_l)(w)$, the impact of the adjustment on worker selection depends on the reverse hazard rates evaluated under the baseline contract. The adjustment itself is akin to a modified Holmström-Mirlees-type contract (Mirrlees, 1999 and Holmström, 1979) with the added consideration of how payment changes for specific outputs affect worker selection. The optimal trade-off between these forces is governed by the coefficients λ^* and μ^*, detailed in the Appendix.

5.2 Nonlocal Contract Modifications

So far, we have focused on local adjustments to the baseline contract. In practice, however, most firms are interested in broader, more substantial alterations that can significantly boost profits. Here, we explore the informational and data prerequisites necessary for identifying the optimal contract.

The principal aims to select a contract $\tilde{w}$, incentive compatible effort levels (a_h, a_l), and payoff allocations (u_h, u_l) to optimize her profit. This can be formally described as:

$$\max_{\tilde{w}(\cdot),\, a_h,\, a_l,\, u_h,\, u_l} \int [x - \tilde{w}(x)] \left[q(u_l, u_h) f(x|a_h) + (1 - q(u_l, u_h)) f(x|a_l)\right] dx \tag{P}$$

$$\text{s.t.} \quad \int (v \circ \tilde{w})(x) f(x|a_t) dx - c_t(a_t) = u_t \quad \text{for each } t \in \{l, h\}$$

$$\int (v \circ \tilde{w})(x) \frac{\partial f(x|a_t)}{\partial a} dx = c_t'(a_t) \quad \text{for each } t \in \{l, h\},$$

$$r_h p G_h(u_h) + r_l (1 - p) G_l(u_l) \geq V,$$

where $q(u_l, u_h)$ is given in (2) and is a function of u_h and u_l.[8] The first set of constraints stipulates that each worker type attains a prescribed utility level, the second set enforces incentive compatibility via first-order conditions, while the third guarantees that all vacancies can be filled with those who pass the screen test.

We maintain Assumptions 1–3 throughout this analysis. Furthermore, we introduce a new assumption regarding the specific functional forms of the workers' effort costs and the distributions of their outside options:

Assumption 4.

i. Workers have isoelastic costs satisfying $c_t'(a) = e^{-\gamma_t/\epsilon} a^{1/\epsilon}$ for some parameters $(\gamma_l, \gamma_h, \epsilon)$; and

ii. the outside option distributions belong to a generalized location-scale distribution family $G_t(z) = \overline{G}((m(z) - \theta_t)/\theta_0)$ with parameters $(\theta_0, \theta_l, \theta_h)$, a strictly increasing $m(\cdot)$, and a log-concave standardized distribution $\bar{G}$ with support unbounded from above.

We also assume the firm observes outcome data from a different contract as detailed below.

Experiment 2. *The firm observes outcome data under some contract $\widehat{w}$ such that $\widehat{w}(x) \neq w(x)$ on a set of outputs with positive Lebesgue measure. In particular, the firm observes the distribution of aggregate output $\overline{f}(x \mid \widehat{w})$, as well as the total mass of applicants $T(\widehat{w})$ and the mass passing the screening test $S(\widehat{w})$.*

[8] We abuse notation to write q as a function of u_h and u_l directly. Note that q is a function of w only through the expected utility for each type. Once (u_l, u_h) are fixed, so it is q.

Experiment 2 generates data that allows us to test—and potentially reject—our model. We say that the model is *rejected* if there exists no model primitives that could generate outcomes consistent with Experiment 2's data. For example, if $u_t(\widehat{w}) > u_t(w)$ for each $t \in \{h, l\}$ yet fewer workers apply under $\widehat{w}$ compared to the baseline contract, then the model should be rejected. Similarly, if the fraction of applicants passing the screen falls outside the interval $[r_l, r_h]$, the model is rejected.[9]

The following result demonstrates that, whenever the model is not rejected, Assumptions 1-4 together with the outcome data from Experiment 2 provide all the information necessary to determine the optimal contract that balances insurance, incentives, and selection considerations.

Theorem 4. *Let Assumptions 1–4 hold. Suppose that under the baseline contract, high and low types exert different and non-zero efforts, i.e.,* $a_h(w) > a_l(w) > 0$*. Furthermore, assume that Experiment 2 does not reject the model, has a non-zero impact on both utilities and incentives, i.e.,* $u_t(w) \neq u_t(\widehat{w})$ *and* $a_t(w) \neq a_t(\widehat{w})$ *for each* $t \in \{l, h\}$*. Then, the information obtained from Experiment 2 suffices to solve problem (P).*

Given the necessary information, (P) can be solved using the two-step methodology outlined by Grossman and Hart (1983): initially finding the profit-maximizing contract while holding the remaining choice variables fixed, and subsequently optimizing over these variables, potentially via grid search. We remark that one consequence of Assumptions 3 and 4 is that choosing a_h implicitly pins down a_l, and (a_h, u_h) jointly pin down u_l. Thus, in the second stage, one only needs to optimize over (a_h, u_h).

6 Conclusion

This paper highlights three main findings regarding the dual role of high-powered incentives: motivating effort and influencing workforce composition. First, workforce sorting depends exclusively on changes in the relative application rates of the different skill types—knowledge of the skill distribution or the firm's screening precision is unnecessary. Second, we show that whenever a contract adjustment benefits all workers—as is common in many reforms—there always exist plausible outside-option distributions un-

[9] Recall that the pass rates are precisely r_h and r_l for high and low types, respectively. Thus, any observed pass rate outside this range contradicts model predictions.

der which low-skilled workers flood in disproportionately. Thus, appealing to "steeper incentives" as a universal recipe for upgrading talent is misplaced; without empirical guidance, the firm is essentially gambling on the shape of workers' outside options.

Our third contribution addresses this uncertainty with empirical evidence. We introduce a sufficient-statistics test based solely on an easily measurable factor: the proportion of applicants who pass the firm's screening mechanism. An increase in this proportion following a contract change indicates improved selection; a decrease signals the opposite. This measure's simplicity, which does not depend on screening precision or labor supply conditions, makes it applicable across various sectors and roles. We complement this assessment with a second experimental approach involving on-the-job bonuses to distinguish the direct effect of incentives on effort. Collectively, these experiments enable firms to assess how profit margins shift in response to any contractual adjustment and select the modification that maximizes profits. In Section 5 we show how the same data, combined with mild structure on effort costs and outside options, extend naturally to the global problem of designing the entire optimal contract.

Several practical implications follow. Firms evaluating compensation reforms should track not only output but also the composition of applicants and hires, and they should embed simple screen-pass metrics into their A/B tests. Policy makers concerned with wage inequality or labor-market sorting can likewise gauge the heterogeneity of responses without full structural estimation. Future work might enrich the framework with multidimensional skill, dynamic learning about match quality, or peer effects in effort. Finally, while we establish what can be learned from two clean experiments, implementing them at scale–especially in environments with complex team production–remains fertile ground for both methodological advances and field applications.

Bibliography

Alfitian, Jakob, Dirk Sliwka, and Timo Vogelsang. 2024. "When Bonuses Backfire: Evidence from the Workplace." *Management Science*, 70(9): 6395–6414.

Ashraf, Nava, and Oriana Bandiera. 2018. "Social Incentives in Organizations." *Annual Review of Economics*, 10(Volume 10, 2018): 439–463.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul. 2005. "Social Preferences and the Response to Incentives: Evidence from Personnel Data." *Quarterly Journal of Economics*, 120(3): 917–962.

Bénabou, Roland, and Jean Tirole. 2003. "Intrinsic and extrinsic motivation." *Review of Economic Studies*, 70(3): 489–520.

Booth, Alison L., and Jeff Frank. 1999. "Earnings, Productivity, and Performance-Related Pay." *Journal of Labor Economics*, 17(3): 447–463.

Bryson, Alex, John Forth, and Lucy Stokes. 2017. "How much performance pay is there in the public sector and what are its effects?" *Human Resource Management Journal*, 27(4): 581–597.

Bénabou, Roland, and Jean Tirole. 2006. "Incentives and Prosocial Behavior." *American Economic Review*, 96(5): 1652–1678.

Casadesus-Masanell, Ramon. 2004. "Trust in Agency." *Journal of Economics & Management Strategy*, 13(3): 375–404.

Castro-Pires, Henrique, and Humberto Moreira. 2021. "Limited liability and non-responsiveness in agency models." *Games and Economic Behavior*, 128: 73–103.

Castro-Pires, Henrique, Hector Chade, and Jeroen Swinkels. 2024. "Disentangling Moral Hazard and Adverse Selection." *American Economic Review*, 114(1): 1–37.

Chetty, Raj. 2009. "Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods." *Annual Review of Economics*, 1(Volume 1, 2009): 451–488.

Conti, Annamaria, Vansh Gupta, Jorge Guzman, and Maria Roche. 2023. "Better Keep the Twenty Dollars: Incentivizing Innovation in Open Source." *Harvard Business School Working Paper*, No. 24-014.

Cornelissen, Thomas, John S. Heywood, and Uwe Jirjahn. 2011. "Performance pay, risk attitudes and job satisfaction." *Labour Economics*, 18(2): 229–239.

Dal Bó, Ernesto, Frederico Finan, and Martín A. Rossi. 2013. "Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service." *Quarterly Journal of Economics*, 128(3): 1169–1218.

Deserranno, Erika. 2019. "Financial Incentives as Signals: Experimental Evidence from the Recruitment of Village Promoters in Uganda." *American Economic Journal: Applied Economics*, 11(1): 277–317.

Finkelstein, Amy, and Matthew J Notowidigdo. 2019. "Take-Up and Targeting: Experimental Evidence from SNAP." *Quarterly Journal of Economics*, 134(3): 1505–1556.

Foarta, Dana, and Takuo Sugaya. 2021. "The management of talent: Optimal contracting for selection and incentives." *The RAND Journal of Economics*, 52(1): 49–77.

Frey, Bruno S., and Felix Oberholzer-Gee. 1997. "The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding- Out." *American Economic Review*, 87(4): 746–755.

Friebel, Guido, Matthias Heinz, Miriam Krueger, and Nikolay Zubanov. 2017. "Team Incentives and Performance: Evidence from a Retail Chain." *American Economic Review*, 107(8): 2168–2203.

Fryer, Roland G. 2013. "Teacher Incentives and Student Achievement: Evidence from New York City Public Schools." *Journal of Labor Economics*, 31(2): 373–407.

Fryer, Roland G., Jr. 2011. " Financial Incentives and Student Achievement: Evidence from Randomized Trials." *Quarterly Journal of Economics*, 126(4): 1755–1798.

Georgiadis, George. 2024. "Contracting with moral hazard." *Elgar Encyclopedia on the Economics of Competition, Regulation and Antitrust*, 24 – 37. Cheltenham, UK:Edward Elgar Publishing.

Georgiadis, George, and Michael Powell. 2022. "A/B Contracts." *American Economic Review*, 112(1): 267–303.

Gielen, Anne C., Marcel J. M. Kerkhofs, and Jan C. van Ours. 2010. "How performance related pay affects productivity and employment." *Journal of Population Economics*, 23(1): 291–301.

Gneezy, Uri, and Aldo Rustichini. 2000. "Pay Enough or Don't Pay at All." *Quarterly Journal of Economics*, 115(3): 791–810.

Gottlieb, Daniel, and Humberto Moreira. 2022. "Simple contracts with adverse selection and moral hazard." *Theoretical Economics*, 17(3): 1357–1401.

Grossman, Sanford J., and Oliver D. Hart. 1983. "An Analysis of the Principal-Agent Problem." *Econometrica*, 51(1): 7–45.

Guiteras, Raymond P., and B. Kelsey Jack. 2018. "Productivity in piece-rate labor markets: Evidence from rural Malawi." *Journal of Development Economics*, 131: 42–61.

Hall, Brian J, Edward Lazear, and Carleen Madigan. 2000. "Performance Pay at Safelite Auto Glass (A) and (B)." *Harvard Business School Case 800-292*.

Hamilton, Barton H, Jack A Nickerson, and Hideo Owan. 2003. "Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation." *Journal of Political Economy*, 111(3): 465–497.

Harberger, Arnold C. 1964. "The Measurement of Waste." *American Economic Review*, 54(3): 58–76.

Holmström, Bengt. 1979. "Moral hazard and observability." *The Bell journal of economics*, 74–91.

Holmstrom, Bengt, and Paul Milgrom. 1987. "Aggregation and linearity in the provision of intertemporal incentives." *Econometrica: Journal of the Econometric Society*, 303–328.

Kreps, David M. 1997. "Intrinsic Motivation and Extrinsic Incentives." *American Economic Review*, 87(2): 359–364.

Lazear, Edward P. 1986. "Salaries and Piece Rates." *The Journal of Business*, 59(3): 405–431.

Lazear, Edward P. 2000. "Performance Pay and Productivity." *American Economic Review*, 90(5): 1346–1361.

Lazear, Edward P. 2009. "Firm-specific human capital: A skill-weights approach." *Journal of Political Economy*, 117(5): 914–940.

Lazear, Edward P. 2018. "Compensation and Incentives in the Workplace." *The Journal of Economic Perspectives*, 32(3): 195–214.

Leuven, Edwin, Hessel Oosterbeek, and Bas van der Klaauw. 2010. "The effect of financial rewards on students' achievement: Evidence from a randomized experiment." *Journal of the European Economic Association*, 8(6): 1243–1265.

Lo, Desmond, Mrinal Ghosh, and Francine Lafontaine. 2011. "The incentive and selection roles of sales force compensation contracts." *Journal of Marketing Research*, 48(4): 781–798.

Milgrom, Paul, and Ilya Segal. 2002. "Envelope Theorems for Arbitrary Choice Sets." *Econometrica*, 70(2): 583–601.

Mirrlees, James A. 1999. "The theory of moral hazard and unobservable behaviour: Part I." *Review of Economic Studies*, 66(1): 3–21.

Moen, Espen R., and Åsa Rosén. 2005. "Performance Pay and Adverse Selection." *The Scandinavian Journal of Economics*, 107(2): 279–298.

Ollier, Sandrine, and Lionel Thomas. 2013. "Ex post participation constraint in a principal–agent model with adverse selection and moral hazard." *Journal of Economic Theory*, 148(6): 2383–2403.

Ortner, Juan, and Sylvain Chassang. 2018. "Making Corruption Harder: Asymmetric Information, Collusion, and Crime." *Journal of Political Economy*, 126(5): 2108–2133.

Paarsch, Harry J., and Bruce Shearer. 2000. "Piece Rates, Fixed Wages, and Incentive Effects: Statistical Evidence from Payroll Records." *International Economic Review*, 41(1): 59–92.

Saez, Emmanuel. 2001. "Using Elasticities to Derive Optimal Income Tax Rates." *Review of Economic Studies*, 68(1): 205–229.

Shearer, Bruce. 2004. "Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment." *Review of Economic Studies*, 71(2): 513–534.

Sliwka, Dirk. 2007. "Trust as a Signal of a Social Norm and the Hidden Costs of Incentive Schemes." *American Economic Review*, 97(3): 999–1012.

Wu, Yanhui. 2017. "Incentive Contracts and the Allocation of Talent." *The Economic Journal*, 127(607): 2744–2783.

A Microfounding the Outside Option Distributions

So far, we have taken the outside option distributions G_l and G_h to be exogenous. In this section we develop a microfoundation for the outside option distributions using the skill-weights approach of Lazear (2009).[10] We also show that, when outside option distributions are endogenous, increasing the steepness of incentives can potentially worsen selection.

Let us consider the following amended model. Each worker possesses two distinct skills, labeled 'A' and 'B'; for example, skill A might represent expertise in economics or accounting, while skill B could correspond to proficiency in computer programming. There are many external firms, each placing different values on these skills. Upon entering the labor market, each worker is endowed with a proportion $\alpha \in [0,1]$ of skill A and $1-\alpha$ of skill B; by convention, we set $\alpha > 1/2$. Each external firm values skill A at a rate $\lambda \in [0,1]$ and skill B at $1-\lambda$, with λ drawn from a distribution H that has density function h. A type-t worker receives K_t independent job offers from firms, with each firm's λ drawn from H; we assume $K_h \geq K_l$, indicating that high-type workers receive at least as many offers as low-type workers. A firm characterized by parameter λ offers a type-t worker utility equal to $\nu_t\big([\lambda\alpha + (1-\lambda)(1-\alpha)]\big)$, where we assume $\nu_t : \mathbb{R} \to \mathbb{R}$ is continuous and strictly increasing—meaning that firms offer higher utility the better the worker's skill profile—and $\nu_h(x) > \nu_l(x)$ for all x to reflect that high-type workers receive a higher utility even within the same firm. Workers choose to apply for a position at the incumbent firm if the expected utility exceeds all alternative offers.

Let u denote the expected utility a worker obtains if employed by the incumbent firm. The likelihood that a randomly selected external offer provides less utility than u is given by

$$\Pr\left\{\nu_t\big([\lambda\alpha + (1-\lambda)(1-\alpha)]\big) \leq u\right\} = \Pr\left\{\lambda \leq \frac{\varphi_t(u) - (1-\alpha)}{2\alpha - 1}\right\} = H\left(\lambda_t(u)\right),$$

where $\varphi_t := \nu_t^{-1}$ and $\lambda_t(u) := [\varphi_t(u) - (1-\alpha)]/(2\alpha - 1)$. Since workers apply only if the utility from the internal position, u, meets or exceeds their best of K_t external offers, the

[10] We thank Canice Prendergast for the suggestion.

probability that a type-t worker applies is

$$G_t(u) = [H(\lambda_t(u))]^{K_t}.$$

Observe that $\lambda_h(u) < \lambda_l(u)$, since

$$\nu_l\big(\varphi_h(u)\big) < \nu_h\big(\varphi_h(u)\big) = u = \nu_l\big(\varphi_l(u)\big) \implies \varphi_h(u) < \varphi_l(u) \implies \lambda_h(u) < \lambda_l(u).$$

Moreover, as H is increasing, it follows that $G_h(u) \leq G_l(u)$ for all u; that is, G_h first-order stochastically dominates G_l.

The most counterintuitive result in this paper is arguably Theorem 1(ii), which shows that for any baseline contract w and any direction of adjustment that has a nontrivial selection effect, there exist outside option distributions for which this adjustment worsens selection. The following theorem shows that our microfoundation can always generate such a pair of distributions.

Theorem 5. *Given any baseline contract w, any local adjustment along direction ℓ such that $\mathcal{D}u_h(w,\ell) \times \mathcal{D}u_l(w,\ell) > 0$, and fixed parameters $K_h \geq K_l$ and $\theta_h > \theta_l$, there exists a distribution H such that this adjustment harms selection.*

B Omitted Proofs

Proof of Lemma 1. Towards establishing this lemma, we first show that $\mathcal{D}u_t(w,\ell)$ exists for every adjustment direction ℓ. To do so, we introduce an intermediate result, Claim 1. Let $\psi(a,\varepsilon) := \int v\big(w(x) + \varepsilon\ell(x)\big) f(x|a)dx - c_t(a)$, and note that $u_t(w+\varepsilon\ell) = \max_a \psi(a,\varepsilon)$, where the maximum is well defined since $\psi(\cdot,\varepsilon)$ is continuous on $a \in [0,\bar{a}]$.

Claim 1. *The family of functions $\{\partial\psi(a,\cdot)/\partial\varepsilon\}_{a\in[0,\bar{a}]}$ is equidifferentiable at $\varepsilon_0 \in [0,1]$ and*

$$\sup_{a\in[0,\bar{a}]} |\partial\psi(a,\varepsilon_0)/\partial\varepsilon| < +\infty.$$

Proof of Claim 1. Note that as v is continuous and w and ℓ bounded, by the Dominated

Convergence Theorem

$$\frac{\partial\psi(a,\varepsilon)}{\partial\varepsilon} = \int v'\big(w(x)+\varepsilon\ell(x)\big)\ell(x)f(x|a)dx < +\infty$$

since v is twice continuously differentiable. Note also that

$$\begin{aligned}\Big|\frac{\partial\psi(a,\tilde{\varepsilon})}{\partial\varepsilon} - \frac{\partial\psi(a,\hat{\varepsilon})}{\partial\varepsilon}\Big| \quad &\leq \int \Big|v'\big(w(x)+\tilde{\varepsilon}\ell(x)\big) - v'\big(w(x)+\hat{\varepsilon}\ell(x)\big)\Big| \cdot |\ell(x)| \cdot f(x|a)dx \\ &\leq |\tilde{\varepsilon}-\hat{\varepsilon}| \cdot \underset{x}{sup}|\ell^2(x)| \cdot \underset{y}{sup}|v''(y)|,\end{aligned}$$

which concludes the Claim's proof. □

Claim 1, together with the boundedness condition $\sup_{a\in[0,\bar{a}]}\left|\frac{\partial\psi(a,\cdot)}{\partial\varepsilon}\right| < +\infty$ and the fact that $\arg\max_{a\in[0,\bar{a}]}\{\psi(a,\varepsilon_0)\} \neq \emptyset$ for all $\varepsilon_0 \in [0,1]$, ensures that the conditions of Theorem 3 in Milgrom and Segal (2002) are satisfied. It follows that the right-hand derivative of $u_t(w+\varepsilon_0\ell)$ with respect to ε_0 exists and is given by $\partial\psi(a_t,\varepsilon_0)/\partial\varepsilon$. In particular, at $\varepsilon_0 = 0$,

$$\mathcal{D}u_t(w,\ell) = \lim_{\varepsilon_0\downarrow 0}\frac{\partial\psi(a_t,\varepsilon_0)}{\partial\varepsilon} = \int (v'\circ w)(x)\ell(x)f(x \mid a_t(w))\,dx. \tag{9}$$

It follows from Remark 1 that $\mathcal{D}q(w,\ell)$ has the same sign as $\mathcal{D}[\ln(G_h\circ u_h)(w) - \ln(G_l\circ u_l)(w)]$. Computing this Gateaux differential delivers the result and concludes the proof. □

Proof of Theorem 1.

Fixing a baseline contract w and a local adjustment ℓ pins $u_t(w)$ and $\mathcal{D}u_t(w,\ell)$ for each $t\in\{l,h\}$. By Lemma 1, this adjustment harms selection if and only if

$$\rho_h(u_h(w)) \times \mathcal{D}u_h(w,\ell) < \rho_l(u_l(w)) \times \mathcal{D}u_l(w,\ell). \tag{10}$$

Part (*i*): Suppose that $\mathcal{D}u_h(w,\ell)$ and $\mathcal{D}u_l(w,\ell)$ have opposite signs. Since $(\rho_t\circ u_t)(w) > 0$, the first part of the theorem follows immediately from (10).

Part (*ii*): Next, suppose that $\mathcal{D}u_h(w,\ell)$ and $\mathcal{D}u_l(w,\ell)$ have the same sign. We will construct distributions G_l and $G_h \succ_{fosd} G_l$ such that (10) is satisfied.

Let $G_t(\cdot)$ be a Gumbel distribution (a.k.a a type-I generalized extreme value distribution) with parameters (μ_t, β). Then $\rho_t(u) = \beta^{-1} \times \exp(-(u - \mu_t)/\beta)$. We will show that there exist parameters μ_h, $\mu_l < \mu_h$, and $\beta > 0$ such that (10) is satisfied.[11]

Note that $G_h \succ_{fosd} G_l$ if and only if $\mu_h > \mu_l$, and

$$\frac{\rho_h(u_h(w))}{\rho_l(u_l(w))} = \exp\left(- \frac{u_h(w) - u_l(w) - (\mu_h - \mu_l)}{\beta} \right).$$

First, suppose that $\mathcal{D}u_t(w, \ell) > 0$ for each $t \in \{l, h\}$. Then (10) can be rewritten as

$$\exp\left(- \frac{u_h(w) - u_l(w) - (\mu_h - \mu_l)}{\beta} \right) < \frac{\mathcal{D}u_l(w, \ell)}{\mathcal{D}u_h(w, \ell)},$$

and note that the right-hand side is positive. Noting that $u_h(w) > u_l(w)$ because high types have smaller absolute and marginal effort costs, for any μ_l and $\mu_h \in \big(\mu_l, \mu_l + u_h(w) - u_l(w)\big)$, the left-hand side converges to zero as $\beta \to 0$, thereby guaranteeing that the inequality is satisfied.

Next, suppose that $\mathcal{D}u_t(w, \ell) < 0$ for each $t \in \{l, h\}$. Then (10) can be rewritten as

$$\exp\left(- \frac{u_h(w) - u_l(w) - (\mu_h - \mu_l)}{\beta} \right) > \frac{\mathcal{D}u_l(w, \ell)}{\mathcal{D}u_h(w, \ell)}.$$

For any β and μ_l, the left-hand side diverges to infinity as $\mu_h \to \infty$, thereby guaranteeing that the inequality is satisfied. □

Proof of Remark 2. Define

$$T(w) := p(G_h \circ u_h)(w) + (1 - p)(G_l \circ u_l)(w), \tag{11}$$

$$S(w) := r_h p(G_h \circ u_h)(w) + r_l(1 - p)(G_l \circ u_l)(w), \text{ and} \tag{12}$$

$$A(w) := S(w)/T(w). \tag{13}$$

$T(w)$ represents the total mass of applicants, $S(w)$ captures the mass of applicants who

[11]Note that $G_h \succ_{fosd} G_l$ if and only if $\mu_h > \mu_l$.

pass the screening test, and $A(w)$ is the passing *rate*. Observe that

$$S(w) = r_l T(w) + (r_h - r_l)\, p\, (G_h \circ u_h)(w) \iff r_h\, p\, (G_h \circ u_h)(w) = \frac{r_h}{r_h - r_l}\big[S(w) - r_l\, T(w)\big].$$

Using

$$q(w) = \frac{r_h\, p\, (G_h \circ u_h)(w)}{S(w)} \quad \text{and} \quad T(w) = \frac{S(w)}{A(w)}$$

we have

$$q(w) = \frac{r_h}{r_h - r_l} \times \frac{S(w) - r_l\, T(w)}{S(w)} = \frac{r_h}{r_h - r_l} \times \left[1 - \frac{r_l\, T(w)}{S(w)}\right] = \frac{r_h}{r_h - r_l} \times \frac{A(w) - r_l}{A(w)},$$

and notice that $q(w)$ increases in $A(w)$. □

Proof of Theorem 2.

For any direction ℓ, we have that

$$\mathcal{D}q(w, \ell) = \frac{G_h(u_h) G_l(u_l) p(1-p) r_h r_l}{[S(w)]^2} \times \left[\rho_h(u_h)\mathcal{D}u_h(w, \ell) - \rho_l(u_l)\mathcal{D}u_l(w, \ell)\right], \qquad (14)$$

where we abuse notation and omit the dependence of u_h and u_l on w for notational simplicity.

By using (9) and Assumptions 1 and 2 (i.e., the knowledge from the baseline environment, $v(\cdot)$, r_h, r_l, $f(\cdot|a_h(w))$ and $f(\cdot|a_l(w))$), one can compute $\mathcal{D}u_t(w, \ell)$ for every adjustment ℓ.

Using (11) and (12) and taking the Gateaux differential along direction ℓ yields

$$\mathcal{D}T(w, \ell) = p\rho_h(u_h) G_h(u_h) \mathcal{D}u_h(w, \ell) + (1-p)\rho_l(u_l) G_l(u_l) \mathcal{D}u_l(w, \ell), \text{and} \qquad (15)$$

$$\mathcal{D}S(w, \ell) = p r_h \rho_h(u_h) G_h(u_h) \mathcal{D}u_h(w, \ell) + (1-p) r_l \rho_l(u_l) G_l(u_l) \mathcal{D}u_l(w, \ell). \qquad (16)$$

Letting $\mathcal{E}_l := (\rho_l \circ u_l)(w)$ and $\mathcal{E}_h := (\rho_h \circ u_h)(w)$, we can solve this linear system with respect to $\mathcal{E}_l$ and $\mathcal{E}_h$ to obtain

$$\mathcal{E}_l = \frac{r_h \mathcal{D}T(w, \ell_1) - \mathcal{D}S(w, \ell_1)}{[r_h T(w) - S(w)]\mathcal{D}u_l(w, \ell_1)}, \quad \text{and} \quad \mathcal{E}_h = \frac{\mathcal{D}S(w, \ell_1) - r_l \mathcal{D}T(w, \ell_1)}{[S(w) - r_l T(w)]\mathcal{D}u_h(w, \ell_1)}.$$

Hence, the data from Experiment 1—specifically, knowledge of $\mathcal{D}T(w,\ell_1)$ and $\mathcal{D}S(w,\ell_1)$, coupled with Assumptions 1 and 2 enables one to recover $\mathcal{E}_l$ and $\mathcal{E}_h$.

Finally, we must recover the term multiplying the square brackets in (14). From the definitions of $T(w)$ and $S(w)$, we have

$$r_h T(w) - S(w) = (1-p)(r_h - r_l)G_l(u_l) \qquad \text{and} \qquad S(w) - r_l T(w) = p(r_h - r_l)G_h(u_h)$$

Hence,

$$K := \frac{G_h(u_h)G_l(u_l)p(1-p)r_h r_l}{[S(w)]^2} = \frac{[r_h T(w) - S(w)]\,[S(w) - r_l T(w)]}{[S(w)]^2} \times \frac{r_h r_l}{(r_h - r_l)^2},$$

which can be recovered using Assumptions 1 and 2. We have therefore shown that

$$\mathcal{D}q(w,\ell) = K \times \int (v' \circ w)(x)\ell(x)\left[\mathcal{E}_h f(x|a_h(w)) - \mathcal{E}_l f(x|a_l(w))\right] dx,$$

and Assumptions 1 and 2 coupled with data from Experiment 1 suffice to compute this Gateaux differential for every adjustment ℓ. □

Proof of Theorem 3. We can write each of the three effects of changing the contract in direction ℓ as a function of observables:

$$\textbf{Incentive effect } = \mu^* \int \ell(x)(v' \circ w)(x) f_a(x) dx,$$

where

$$\mu^* := \frac{\int [s - w(s)] f_a(s) ds}{\int \ell_1(s)(v' \circ w)(s) f_a(s) ds} \cdot \Big[\mathcal{D}\bar{a}(w,\ell_1) - [a_h(w) - a_l(w)]\mathcal{D}q(w,\ell_1)\Big].$$

$$\textbf{Selection effect } = \lambda^* \int \ell(x)(v' \circ w)(x)\left[\mathcal{E}_h f(x|a_h(w)) - \mathcal{E}_l f(x|a_l(w))\right] dx,$$

where $\lambda^* := K \times [a_h(w) - a_l(w)] \cdot \int [s - w(s)] f_a(s) ds$, and

$$\textbf{Direct effect } = -\int \ell(x)[q(w) f(x|a_h(w)) + (1 - q(w)) f(x|a_l(w))] dx.$$

We can then write the problem of finding the best direction of improvement as

$$\max_{\ell(x)} \int \ell(x) \left[\lambda^*(v' \circ w)(x) \left[\mathcal{E}_h f(x|a_h(w)) - \mathcal{E}_l f(x|a_l(w))\right] + \mu^*(v' \circ w)(x) f_a(x) - \bar{f}(x|w)\right] dx$$

subject to

$$\int \ell^2(x) dx \leq 1.$$

By Assumption 1, the firm observes outcome data under w; that is, $f(x|a_l(w))$ and $f(x|a_h(w))$ (and, therefore, $a_l(w)$ and $a_h(w)$. Using the data from Experiment 1, coupled with Assumptions 1 and 2, the firm can recover $K,\ \mathcal{E}_l$, $\mathcal{E}_h$, and $\mathcal{D}q(w, \ell_1)$. In addition, Assumption 3 allows the firm to also recover $f_a(\cdot)$ and $\mathcal{D}\bar{a}(w, \ell_1)$. Therefore, it has all the information needed to determine λ^*, μ^*, and to solve the above maximization program. We now characterize its solution.

Letting $\nu \geq 0$ denote the dual multiplier associated with the constraint, we have the Lagrangian

$$\mathcal{L}(\nu) = \max_{\ell(x)} \nu + \int \ell(x) \Big[(v' \circ w)(x) \big(\mu^* f_a(x) + \lambda^* \left[\mathcal{E}_h f(x|a_h(w)) - \mathcal{E}_l f(x|a_l(w))\right] \big) - \bar{f}(x|w)\Big] - \nu \ell^2(x) dx.$$

For any $\nu \geq 0$, the integrand is differentiable and strictly concave in ℓ. Maximizing it pointwise yields, for any $\nu > 0$, the first-order condition

$$\ell_\nu(x) = \frac{\Big(\lambda^* \left[\mathcal{E}_h f(x|a_h(w)) - \mathcal{E}_l f(x|a_l(w))\right] + \mu^* f_a(x)\Big)(v' \circ w)(x) - \bar{f}(x|w)}{2\nu}.$$

Note that for $\nu = 0$,

$$\mathcal{L}(0) = \begin{cases} 0 & \text{if } \Big(\lambda^* \left[\mathcal{E}_h f(x|a_h(w)) - \mathcal{E}_l f(x|a_l(w))\right] + \mu^* f_a(x)\Big)(v' \circ w)(x) = \bar{f}(x|w) \text{ for almost all } x, \\ \infty & \text{otherwise.} \end{cases}$$

Next, we solve the dual problem:

$$\min_{\nu \geq 0} \mathcal{L}(\nu).$$

This problem is convex, and using $\ell_\nu(x)$, the Lagrangian is minimized at

$$\nu^* = \frac{1}{2} \sqrt{\int \left\{ \Big(\lambda^* \left[\mathcal{E}_h f(x|a_h(w)) - \mathcal{E}_l f(x|a_l(w))\right] + \mu^* f_a(x)\Big)(v' \circ w)(x) - \bar{f}(x|w) \right\}^2 dx}.$$

Up to now, we have shown that ℓ_{ν^*} solves the dual problem. To show that it solves the primal problem, we must establish that strong duality holds. Denote the optimal value of the primal by Π^*. First, by weak duality, we have that $\mathcal{L}(\nu^*) \geq \Pi^*$. Second, it is straightforward to verify that $\int \ell^2_{\nu^*}(x)dx = 1$, and so ℓ_{ν^*} is feasible for the primal. the original problem, and that ν^* is strictly positive if and only if the respective (primal) constraint binds; meaning that the complementary slackness conditions are satisfied. This implies that $\mathcal{L}(\nu^*) \leq \Pi^*$. Therefore, $\mathcal{L}(\nu^*) = \Pi^*$, which proves that strong duality holds, and ℓ_{v^*} solves the primal problem. □

Proof of Theorem 4.

Step 1 (Recovering the basis f_0, f_a). Per Assumption 3, $f(x|a) = f_0(x) + a\, f_a(x)$. By Assumption 1, the firm observes, under the baseline contract, w, the densities $f(x|a_h(w))$ and $f(x|a_l(w))$, from which it can recover the corresponding efforts $a_t(w) = \int x f(x|a_t(w))dx$ for each $t \in \{l, h\}$. As $a_h(w) \neq a_l(w)$, solving the two linear equations delivers the unique pair (f_0, f_a).

Step 2 (Recovering $q(\widehat{w})$). By Remark 2

$$q(\widehat{w}) = \frac{r_h}{r_h - r_l} \times \frac{A(\widehat{w}) - r_l}{A(\widehat{w})},$$

where (r_h, r_l) are known by Assumption 2, and $A(\widehat{w}) = S(\widehat{w})/T(\widehat{w})$ is observable from Experiment 2.

Step 3 (Effort levels under the new contract). For any contract w', let

$$I(w') \;:=\; \int (v \circ w')(x)\, f_a(x)\, dx.$$

denote the *marginal incentives* associated with this contract, and note that by Assumptions 2 and 3 and the data from Experiment 2, the firm has all the information needed to compute this function.

By Assumption 4(i) the first-order condition of a type-$t \in \{l, h\}$ worker facing contract w' is

$$I(w') = c_t'\big(a_t(w')\big) \iff \log a_t(w') = \gamma_t + \epsilon \log I(w'). \tag{17}$$

Equation (17) holds for the four combinations $(t, w') \in \{l, h\} \times \{w, \widehat{w}\}$ and involves the five unknowns $\gamma_l, \gamma_h, \epsilon, a_h(\widehat{w})$, and $a_l(\widehat{w})$. Moreover, as the Experiment has an non-zero impact on incentives, $I(w) \neq I(\widehat{w})$. Taking the difference of (17) between w and $\widehat{w}$ eliminates γ_t and yields

$$\epsilon = \frac{\log a_t(\widehat{w}) - \log a_t(w)}{\log I(\widehat{w}) - \log I(w)}, \qquad t \in \{l, h\}. \tag{18}$$

Since ϵ is common across types, it follows that $a_h(\widehat{w})/a_h(w) = a_l(\widehat{w})/a_l(w)$. Therefore, we have

$$a_t(\widehat{w}) = \bar{a}(\widehat{w}) \times \frac{a_t(w)}{q(\widehat{w})a_h(w) + (1 - q(\widehat{w}))a_l(w)}, \qquad t \in \{l, h\},$$

where $\bar{a}(w')$ has been defined as the expectation of aggregate output under w'.

Note that $a_h(w)$ and $a_l(w)$ are known by Assumption 1, $\bar{a}(\widehat{w})$ is known from Experiment 2, and $q(\widehat{w})$ is known from Step 2 of this proof. Therefore, the firm has all of the information needed to compute $a_l(\widehat{w})$ and $a_h(\widehat{w})$.

Step 4 (Cost parameters). Per the previous step, the firm can compute $a_l(\widehat{w})$ and $a_h(\widehat{w})$, as well as $I(w)$ and $I(\widehat{w})$. Therefore, it can determine the cost parameter ϵ using (18). Then, $\gamma_t = \log a_t(w) - \epsilon \log I(w)$ for $t \in \{l, h\}$.

Step 5 (Outside-option distribution parameters). From Step 4, we can compute $u_t(w')$ for $(t, w') \in \{l, h\} \times \{w, \widehat{w}\}$. Define for each $w' \in \{w, \widehat{w}\}$

$$\eta_h(w') := \frac{S(w') - r_l T(w')}{r_h - r_l} = p\bar{G}\left(\frac{m(u_h(w')) - \theta_h}{\theta_0}\right) \tag{19}$$

$$\eta_l(w') := \frac{r_h T(w') - S(w')}{r_h - r_l} = (1 - p)\bar{G}\left(\frac{m(u_l(w')) - \theta_l}{\theta_0}\right). \tag{20}$$

These 4 constants can be computed by Assumptions 1-2 and Experiment 2. The facts that $p \in (0, 1)$ and $\bar{G}$ is a CDF imply that

$$p \in (\max\{\eta_h(w), \eta_h(\widehat{w})\},\ 1 - \max\{\eta_l(w), \eta_l(\widehat{w})\}).$$

Hence, the model not being rejected implies that

$$0 < \bar{\eta}_h < \ 1 - \bar{\eta}_l < 1,$$

where $\bar{\eta}_t := max\{\eta_t(w), \eta_t(\widehat{w})\}$. Moreover, as $u_t(w) \neq u_t(\widehat{w})$ for all t, we have from (19) and (20) that:

$$\frac{1}{\theta_0} = \frac{\bar{G}^{-1}(\eta_h(w)/p) - \bar{G}^{-1}(\eta_h(\widehat{w})/p)}{m(u_h(w)) - m(u_h(\widehat{w}))} = \frac{\bar{G}^{-1}(\eta_l(w)/(1-p)) - \bar{G}^{-1}(\eta_l(\widehat{w})/(1-p))}{m(u_l(w)) - m(u_l(\widehat{w}))}.$$

Define

$$\Psi(p) := \frac{\bar{G}^{-1}(\eta_h(w)/p) - \bar{G}^{-1}(\eta_h(\widehat{w})/p)}{m(u_h(w)) - m(u_h(\widehat{w}))} - \frac{\bar{G}^{-1}(\eta_l(w)/(1-p)) - \bar{G}^{-1}(\eta_l(\widehat{w})/(1-p))}{m(u_l(w)) - m(u_l(\widehat{w}))}.$$

The solution to the system must be such that $\Psi(p) = 0$. Note that Ψ is continuous, and because $\bar{G}$ is log-concave, it is strictly decreasing. Moreover, $\lim_{p \to \bar{\eta}_h} \Psi(p) > 0$ and $\lim_{p \to 1-\bar{\eta}_l} \Psi(p) < 0$. Hence, by the Intermediate Value Theorem there exists a unique p such that $\Psi(p) = 0$. Given this p, the firm can recover

$$\begin{aligned}
\theta_0 =& \frac{m(u_h(w)) - m(u_h(\widehat{w}))}{\bar{G}^{-1}(\eta_h(w)/p) - \bar{G}^{-1}(\eta_h(\widehat{w})/p)} \\
\theta_l =& m(u_l(w)) - \theta_0 \bar{G}^{-1}(\eta_l(w)/(1-p)) \\
\theta_h =& m(u_h(w)) - \theta_0 \bar{G}^{-1}(\eta_h(w)/p).
\end{aligned}$$

Step 6 (Solving the principal's problem). Steps 1–5 identify the primitives $c_t(\cdot)$, $f(\cdot|\cdot)$, $G_t(\cdot)$, and p, so the firm's problem (P) is fully specified. We proceed with the Grossman and Hart (1983) two-step procedure—first maximizing the objective for fixed (u_h, u_l, a_h, a_l); then searching over all feasible such vectors to find the global optimum.

It follows from Assumption 4(i) that for any contract $\tilde{w}$ and incentive compatible effort recommendations, $a_h(\tilde{w})/a_l(\tilde{w}) = e^{\gamma_h - \gamma_l}$. Hence, for each a_h, there is a unique feasible a_l. Additionally, using Assumption 2, we can write

$$u_t(\tilde{w}) = \int (v \circ \tilde{w})(x) f_0(x) dx + a_t \underbrace{\int (v \circ \tilde{w})(x) f_a(x) dx}_{=I(\tilde{w}) = c'_h(a_h)} - c_t(a_t),$$

which implies that

$$u_l = u_h - (a_h - a_l)c_h'(a_h) + c_h(a_h) - c_l(a_l).$$

Hence, for every (a_h, u_h), there is a unique feasible pair (a_l, u_l).

For each pair (a_h, u_h), we compute the corresponding a_l, u_l and q, and solve

$$\begin{aligned} \Pi(a_h, u_h) := \max_{\tilde{w}(x)} \quad & \int [x - \tilde{w}(x)] \left[q f(x|a_h) + (1-q) f(x|a_l) \right] dx \\ \text{s.t.} \quad & \int (v \circ \tilde{w})(x) f(x|a_h) dx - c_h(a_h) = u_h \\ & \int (v \circ \tilde{w})(x) f_a(x) dx = c_h'(a_h) \end{aligned} \tag{21}$$

The result delivers a modified Holmström-Mirrlees type of contract

$$\frac{1}{(v' \circ \tilde{w})(x)} = \lambda \times \frac{f(x|a_h)}{q f(x|a_h) + (1-q) f(x|a_l)} + \mu \times \frac{f_a(x)}{q f(x|a_h) + (1-q) f(x|a_l)},$$

where λ and μ are the multipliers associated with the two constraints in problem (21). The final step is a grid search over (a_h, u_h) to maximize $\Pi(a_h, u_h)$. □

Proof of Theorem 5.

The baseline contract specifies u_h and $u_l < u_h$, and the modified contract provides increments $\mathcal{D}u_h$ and $\mathcal{D}u_l$. Define $\kappa := \mathcal{D}u_l / \mathcal{D}u_h$. The ratio of the application probability density to its level for type t is

$$\frac{g_t(u)}{G_t(u)} = \frac{K_t \varphi_t'(u)}{2\alpha - 1} \rho_H(\lambda_t(u)),$$

where $\rho_H(\lambda) := h(\lambda)/H(\lambda)$. Define the constant

$$C := \frac{\mathcal{D}u_l}{\mathcal{D}u_h} \frac{\varphi_l'(u_l)}{\varphi_h'(u_h)} \frac{K_l}{K_h} > 0.$$

According to Lemma 1, the adjustment harms selection if and only if

$$K_h\varphi_h'(u_h)\rho_H(\lambda_h(u_h))\mathcal{D}u_h < K_l\varphi_l'(u_l)\rho_H(\lambda_l(u_l))\mathcal{D}u_l$$

$$\Longleftrightarrow \quad \begin{cases} \frac{\rho_H(\lambda_h(u_h))}{\rho_H(\lambda_l(u_l))} < C & \text{if } \mathcal{D}u_h > 0 \\ \frac{\rho_H(\lambda_h(u_h))}{\rho_H(\lambda_l(u_l))} > C & \text{if } \mathcal{D}u_h < 0 \end{cases} \tag{22}$$

Since this inequality depends solely on the reverse hazard rate of H, we can construct a function ρ_H on $[0,1)$ to satisfy it. The corresponding cumulative distribution function is $H(\lambda) = \exp(-\int_\lambda^1 \rho_H(s)ds)$ for $\lambda \in (0,1]$, with H assigning a mass point at 0.

If $[\lambda_l(u_l) - \lambda_h(u_h)] \times \mathcal{D}u_h \geq 0$, set

$$\rho_H(\lambda) = \begin{cases} r & \text{for } \lambda \in \big[0, \underline{\lambda}\big], \\ R & \text{for } \lambda \in (\underline{\lambda}, 1\big], \end{cases}$$

where $\underline{\lambda} := \min\{\lambda_l(u_l), \lambda_h(u_h)\}$, $r, R > 0$, and $r/R < \min\{1/C, C\}$. Observe that

$$\frac{\rho_H(\lambda_h(u_h))}{\rho_H(\lambda_l(u_l))} = \begin{cases} r/R & \text{if } \lambda_h(u_h) \leq \lambda_l(u_l), \\ R/r & \text{if } \lambda_h(u_h) > \lambda_l(u_l). \end{cases}$$

In either case, the inequality in (22) is satisfied.

If instead $[\lambda_l(u_l) - \lambda_h(u_h)] \times \mathcal{D}u_h < 0$. Define

$$\rho_H(\lambda) = \begin{cases} R & \text{for } \lambda \in [0, \underline{\lambda}], \\ r & \text{for } \lambda \in (\underline{\lambda}, 1], \end{cases}$$

ensuring

$$\frac{\rho_H(\lambda_h(u_h))}{\rho_H(\lambda_l(u_l))} = \begin{cases} R/r & \text{if } \lambda_h(u_h) \leq \lambda_l(u_l), \\ r/R & \text{if } \lambda_h(u_h) > \lambda_l(u_l). \end{cases}$$

Again, in either case the condition is satisfied. □