Incentives and Selection*

Henrique Castro-Pires† and George Georgiadis‡

March 31, 2025

Click here for the latest version

Abstract

Performance pay schemes influence not only workers' incentives to exert effort but also the skill composition of the applicant pool. Conventional wisdom suggests that stronger incentives should attract higher-skilled workers. We demonstrate, however, this is not always the case: under certain distributions of outside options, stronger incentives may disproportionately cause low-skilled workers to flood into the applicant pool, reducing overall workforce quality. We identify necessary and sufficient conditions under which selection necessarily improves, as well as a simple, sufficient-statistics test—based on approval rates in a screening process—that reveals whether a given contract modification improves or harms selection. We then develop a framework for characterizing the optimal incentive adjustments that maximize performance without harming selection.

*We are grateful to Eduardo Azevedo, Arjada Bardhi, Rahul Deb, Michael Grubb, Johannes Horner, Jan Knoepfle, Erik Madsen, Lucas Maestri, Humberto Moreira, Doron Ravid, Julia Salmi, João Thereze, and several participants of the QMUL-City Theory Workshop, the Game Theory & Information Economics Conference at EPGE/FGV-Rio de Janeiro, and seminar participants at Boston College.

†University of Miami, Herbert Business School; h.castro-pires@miami.edu

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

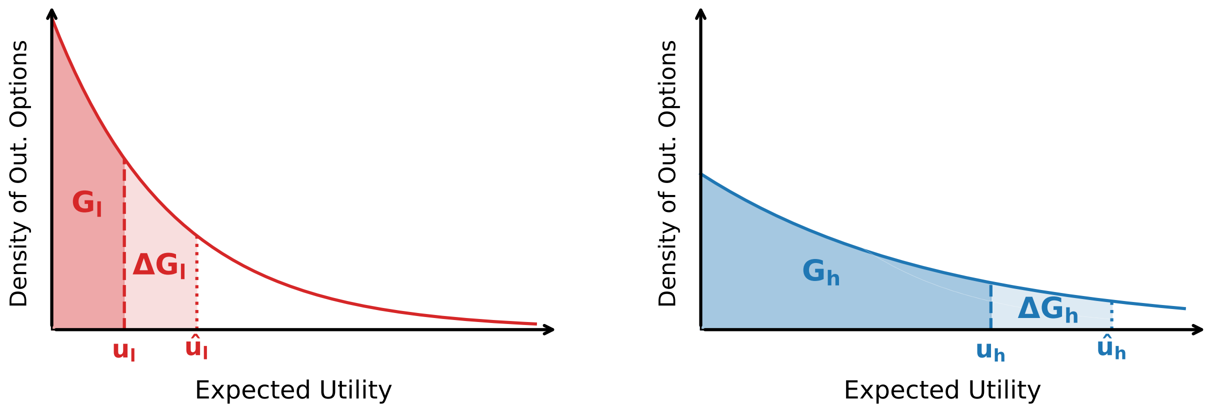
1 Introduction

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"). Indeed, the offer of a steeper incentive scheme can attract a different mix of job-seekers. 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 workers. 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 a performance-based contract: each worker exerts costly effort to generate output, and workers differ in skill. High-skilled workers face lower 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?



(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 shaded areas indicate the mass of new applicants drawn in by the modified contract.

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] 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. Critically, whether selection improves or worsens depends solely on how much each type benefits and how their outside options are distributed. Surprisingly,

[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.

neither the fraction of high-skilled workers in the overall population nor the accuracy of the firm's screening test plays any direct role in determining this outcome.

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). 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, such as 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. Thus, we focus on the subtler case of contract modifications that move payoffs for both types in the same direction.

We extend our analysis beyond the simple linear-contract setting by considering local adjustments to any incentive contract—potentially nonlinear—and examining their effect on effort, the share of high-skilled workers in the firm, and overall profit. This general approach lets us define a broad notion of "steepening of incentives," which means that, however the contract is tweaked, both high- and low-skilled workers face stronger incentives to exert effort.

Our first main result (Theorem 2) reveals a sharp negative possibility: *for any given contract* and *any* way of steepening its incentives, 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. Then, in a partial converse (Theorem 3), we characterize the shape of the outside-option distributions under which any such steepening improves selection. This condition requires that the reverse hazard rate (i.e., the ratio between the density and cumulative distribution) of outside options evaluated at the utility under the status quo contract is weakly larger for high types than for low types. The intuition is that the mass of high-type applicants would respond more in percentage terms than the mass of low-type applicants and, hence, the share of high-skilled workers hired would increase.

A practical hurdle to applying these theorems is that firms rarely observe detailed data on outside-option distributions. Hence, we propose a sufficient-statistics style test: namely, Theorem 4 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 a greater inflow of high-skilled workers relative to 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 5) that by running 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 all local adjustments that would improve selection.

After clarifying when steepened incentives improve selection, we then ask what is the *optimal* contract adjustment. An exact solution is typically difficult because a contract change affects profits through three channels: how workers adjust their effort (the "incentive effect"), how workforce composition changes (the "selection effect"), and any direct change in wage costs. Fully optimizing over these requires substantial knowledge of worker preferences and market structure.

To tackle this complexity, we propose a simpler strategy: focus on maximizing the incentive and direct effects only, subject to not harming selection. By ignoring the possibility that harming selection might raise short-run profits, we ensure that the firm reaps higher effort without risking long-term damage from drawing in less-skilled recruits.Under an assumption about how effort translates into output, Theorem 6 demonstrates that firms can identify the optimal adjustment—among those that do not harm selection—by running just two simple experiments. The first is an adjustment made before hiring, which gauges how many applicants enter and pass the test. The second is a post-hiring modification—a bonus or other incentive that applies only after workers are on the job—so that selection is unaffected, and the principal can observe how effort (and thus output) responds in isolation. Combining these two data points, the firm can pinpoint exactly which modifications to the existing contract improve profits the most without reducing the share of high-skilled workers.

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 et al. (2024)) or choose not to screen (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.

Several papers document productivity gains following an increase in incentives, including influential studies by Lazear (2000, 2018), Shearer (2004), Bandiera et al. (2005), and Friebel et al. (2017). Yet, others find monetary incentives to be ineffective or even counterproductive; e.g., Leuven et al. (2010), Fryer (2011, 2013), and Alfitian et al. (2024). Existing theoretical 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)), interactions with social norms (e.g., Gneezy and Rustichini (2000), Sliwka (2007)), and social preferences or peer pressure (e.g., Hamilton et al. (2003), Ashraf and Bandiera (2018)). We contribute to this literature by demonstrating 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 monetary rewards.

This paper also relates to a recent literature examining the effect of monetary incentives on the selection of employees. Similarly to the literature on productivity, the empirical evidence here is mixed. Following increases in financial rewards, Dal Bó et al. (2013) find evidence of improved selection among civil servants in Mexico, whereas Guiteras and Jack (2018) document no improvement in the context of informal labor in rural Malawi, while Deserranno (2019) reports negative selection effects among health-promoters in Uganda. We extend this literature in two main ways: First, we propose another mechanism for negative selection beyond attracting less intrinsically (or pro-socially) motivated workers. We show that even if intrinsic motivation plays no role, selection might be harmed depending on the shape of the workers' outside option distributions. Second and most importantly, we develop a simple test to evaluate whether selection improves following changes in incentives. Given that empirical studies on selection often struggle with data limitations, our parsimonious test (Theorem 4) can help overcome these challenges.

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 extensively 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 in two significant ways: we incorporate worker selection into the analysis, and we derive 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 aims to hire a fixed mass of agents and motivate them to exert hidden effort.

Events unfold in the following order:

i. The principal posts a fixed number of (identical) job openings and a wage contract $w(\cdot)$, which is a bounded and upper-semicontinuous mapping from output x to payments.

ii. Each agent has a privately known type $t \in \{l, h\}$, where the share of high types ($t = h$) is p. Each agent, conditional on his type, draws his outside option $\overline{u}$ from a type-dependent distribution $G_t(\cdot)$ and decides whether to apply for an opening.

iii. The principal has an imperfect screening technology, whereby each type-t applicant passes the screen with probability $1 - r_t$. We assume that $r_l > r_h$ so that high types pass with strictly higher probability than low types. Then, the principal hires at random among the agents who pass the screen to fill the vacancies.

iv. Each hiree then chooses how much effort, $a \in [0, \overline{a}] \subset \mathbb{R}_+$ to exert, individual output x is drawn according to the probability density function $f(\cdot|a)$, payoffs are realized, and the game ends.

If a type-t agent is hired, is paid y, and exerts effort a, then his payoff is $v(y) - c_t(a)$, where

v is strictly increasing, twice continuously differentiable, and weakly concave, while c_t is strictly increasing, strictly convex, and twice continuously differentiable. Thus, a type-t agent applies whenever his expected utility from taking the job is larger than his outside option:

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

From Bayes' rule, the probability of hiring a high-type agent given contract w is

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

The numerator represents the probability that a high-type chooses to apply and passes the screening test, while the denominator represents the probability that a randomly chosen agent applies and passes the test.

Then the principal's payoff (per opening) when she posts contract w is

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

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

We will study how the share of high types among the hired agents q and the principal's payoff π respond to changes in some arbitrary status quo contract w. We will say that a change in the contract causes *positive selection* if it causes the share of high types among the hirees, $q(\cdot)$, to rise, and it causes *negative selection* otherwise.

Finally, we impose the following assumptions:

A.1. High types have weakly better outside options; i.e., G_h weakly first-order stochastically dominates G_l. Moreover, both distributions are continuously differentiable.

A.2. High types have strictly smaller absolute and marginal effort costs than low types;

i.e., $c_l(a) > c_h(a)$ and $c_l'(a) > c_h'(a)$ for all $a > 0$.

A.3. The density function $f(\cdot|a)$ has full support over the output space, a finite first moment, and its derivative with respect to a, denoted f_a, exists. Without loss of generality, we normalize $a \equiv \mathbb{E}[x|a]$, so that an agent's effort is interpreted as their expected output.

A.4. At the status quo contract, strictly positive masses of both types apply for the job; i.e., $(G_h \circ u_h)(w), (G_l \circ u_l)(w) > 0$.

3 Steeper Incentives, Worse Selection?

Conventional wisdom suggests that higher-powered incentives attract higher-ability workers because they will have the opportunity to earn more under the new scheme. For example, when Safelite, a windshield repair firm, switched from hourly wages to piece rates, productivity increased by 44%, half of which was attributable to increased motivation and the other half to 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

Consider a simple example where agents are risk-neutral, they have isoelastic effort costs, and the principal restricts attention to linear contracts. That is,

$$v(y) \equiv y,\ c_t(a) = \frac{a^{1+\gamma}}{(1+\gamma)\theta_t},\ G_l \equiv G_h := G, \text{ and } w(x) = \alpha x,$$

for some $\theta_l < \theta_h$ and $\gamma > 0$. Using the normalization $a \equiv \mathbb{E}[x|a]$ and solving for the agent's optimal effort, we have that the expected utility of a type$-t$ agent

$$u_t(w) = \frac{\alpha\gamma}{1+\gamma} \left(\alpha\theta_t\right)^{1/\gamma}.$$

Consider a marginal increase of the slope α. The expected utility of both types will increase, so more both low and high types will apply. Notice that $du_t/d\alpha = (\alpha\theta_t)^{1/\gamma}$ increases in θ_t; that is, high types benefit more from the larger slope than the low types. Of course, this is not sufficient to guarantee an improvement in selection. In fact, by Remark 1, selection *worsens* if (and only if) $(G \circ u_h)(w)/(G \circ u_l)(w)$ decreases in α. We have

$$\frac{d}{d\alpha}\frac{(G \circ u_h)(w)}{(G \circ u_l)(w)} =_s \frac{g(u_h)}{G(u_h)}\frac{du_h}{d\alpha} - \frac{g(u_l)}{G(u_l)}\frac{du_l}{d\alpha} =_s u_h\rho(u_h) - u_l\rho(u_l),$$

where $\rho := g/G$ is the reverse hazard rate function of G.[2] Thus selection worsens if $u\rho(u)$ *decreases* in u, or equivalently, if the reverse hazard rate function, ρ, has elasticity smaller than -1.[3]

Several distributions have this property. For example, the reverse hazard rate function of the uniform distribution has elasticity smaller than -1. Similarly, the exponential distri-

[2]The symbol "$=_s$" indicates that the objects on either side have strictly the same sign.
[3]Conversely, selection is improved if ρ has elasticity greater than -1.

bution has this property, as does the Pareto distribution, as long as the shape parameter is larger than 1, and the right tail of the normal distribution also has this property; i.e., $u\rho(u)$ is decreasing for u sufficiently large.

3.2 General Framework

In this section, we extend the above insight to a more general contracting environment and identify conditions under which a marginal change to an arbitrary contract in a given direction might improve or harm selection.

To carry out this exercise, we must describe how the share of high types among hired workers changes as we locally adjust the status quo 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 direction of adjustment $\ell : X \to \mathbb{R}$ is the difference between an adjustment contract $\hat{w}$ and the status quo contract w. Our main interest is how $q(w)$ varies as we change the status quo contract w in direction ℓ, i.e., whether $\mathcal{D}q(w,\ell)$ is positive.

Definition. Adjusting w in direction ℓ *improves* selection if and only if $\mathcal{D}q(w,\ell) > 0$.

The following lemma provides a necessary and sufficient condition such that an adjustment improves or harms selection.

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

$$\frac{(g_h \circ u_h)(w)}{(G_h \circ u_h)(w)} \times \mathcal{D}u_h(w,\ell) > \frac{(g_l \circ u_l)(w)}{(G_l \circ u_l)(w)} \times \mathcal{D}u_l(w,\ell) \tag{4}$$

The interpretation is similar to (1): On the margin, adjusting the status quo contract in the direction of ℓ increases the utility of type$-t$ agents by $\mathcal{D}u_t(w, \ell)$ utils, and their probability of applying by $(g_h \circ u_h)(w)/(G_h \circ u_h)(w)$ percentage points per util. Thus, selection is improve if (and only if) the percentage change of high-type applicants exceeds that of low-type applicants.

Our first result is a direct implication of Lemma 1, and it shows that any adjustment that impacts the expected utility of high types and low types in opposite directions has an unambiguous effect on selection.

Theorem 1. *Consider a marginal adjustment in direction ℓ of the status quo contract w.*

(i) If $\mathcal{D}u_h(w, \ell) \geq 0 \geq \mathcal{D}u_l(w, \ell)$, then this adjustment improves selection.

(ii) If $\mathcal{D}u_h(w, \ell) \leq 0 \leq \mathcal{D}u_l(w, \ell)$, then this adjustment harms selection.

If the adjustment raises the expected utility of high types while decreasing that of low types, then the firm attracts more high-type applicants and fewer low-type ones, thus improving selection. One example of such an adjustment is a rotation of the contract, where the firm increases its slope and reduces base pay by the appropriate amount. The converse is true in case (ii). Notice, however, that to determine how to implement such a rotation (e.g., how much to reduce base pay), the firm must know the agents' preferences, as well as the mapping from effort into output. Otherwise, it might inadvertently increase or decrease the expected utility of both types which, as the next theorem shows, can lead to negative selection for *any* adjustment.

In the remainder of this section, we focus on adjustments that impact both types' expected utilities in the same direction. Specifically, we are interested in conditions such that *steepening* incentives improves or harms selection. When contracts are linear and agents are risk-neutral, the definition of "steepening incentives" corresponds to increasing the contract's slope. Towards generalizing the notion of "steepening incentives" to arbitrary contracts observe that each agent's optimal effort, assuming its interior, satisfies

$$\int (v \circ w)(x) f_a(x|a) dx = c_t'(a),$$

that is, it equates marginal incentives to marginal cost. We will say that an adjustment ℓ to the status quo contract *steepens* incentives if it increases marginal incentives for all levels of effort.

Definition. An adjustment to w in the direction of ℓ *steepens incentives* if

$$\int (v' \circ w)(x) \ell(x) f_a(x|a) dx > 0 \quad \forall a.$$

Theorem 2. *Consider an arbitrary status quo contract w and a steepening of incentives in direction ℓ, where $\mathcal{D}u_h(w,\ell) \times \mathcal{D}u_l(w,\ell) > 0$. Then, there exist $G_h(\cdot) \succ_{FOSD} G_l(\cdot)$ for which a local adjustment in direction ℓ harms selection.*

Theorem 2 shows that *any* steepening of incentives that affects both types' utilities in the same direction may lead to negative selection. For instance, consider a change in contracts that steepens incentives but weakly raises payments for all output realizations (à la the change studied by Lazear (2000)). As high types have lower effort cost, $u_h \geq u_l$. Moreover, as incentives are steepened, high types' utility increases more than low types' (i.e., $\mathcal{D}u_h(w,\ell) \geq \mathcal{D}u_l(w,\ell)$), which favors selection. This theorem shows that one can always construct outside option distributions with a reverse hazard rate that decreases fast enough that reverses the effect of bigger utility gains for high types by attracting a sufficiently larger mass of low types.

In Theorem 2, we take as given the workers' preferences and show that for any given status quo contract and any steepening of incentives that changes the utility of both types in the same direction—as was the case at Safelite (Lazear, 2000), there exist outside option distributions under which such steepening harms selection. We now look at the converse: we characterize the outside option distributions under which any strict Pareto-improving steepening of incentives improves selection for all possible effort cost functions.

Whether a local adjustment generates a strict Pareto improvement or not may depend on the effort cost function as it affects the agents' effort choices. We first state a condition on the direction of the adjustment $\ell(\cdot)$ that guarantees a strict Pareto improvement for every effort cost function. This condition requires that this adjustment increases the worker's expected monetary utility for any fixed effort level, i.e.,

$$\int v'\big(w(x)\big)\ell(x)f(x|a)dx > 0 \text{ for all } a. \tag{5}$$

Condition (5) guarantees that any agent strictly benefits from the contract change even if they do not change their efforts, regardless of what level it was originally at. As agents can always keep the same initial effort level, both types would strictly benefit from that change once they can also adjust their effort.

Theorem 3. *Consider a marginal steepening of the status quo contract in the direction of ℓ, and assume it satisfies* (5)*.*

(i) If $\rho_h(\tilde{u}) \geq \rho_l(\hat{u})$ for all $\tilde{u} \geq \hat{u}$, this adjustment improves selection for all effort cost functions.

(ii) Otherwise, there exist effort cost functions such that this adjustment harms selection.

Theorem 3 generalizes the insights provided by our linear contracts example by showing that a decreasing reverse hazard rate is the key feature of outside option distributions potentially generating negative selection. As in our example, if the reverse hazard rate of the high types' outside option distribution is significantly smaller than the low types' one, it can outweigh the fact that high types benefit more from steeper incentives than low types. This theorem shows that if there is a pair of utility levels $\tilde{u} \geq \hat{u}$ for which

$\rho_h(\tilde{u}) < \rho_l(\hat{u})$, then we can construct effort cost functions for which

$$u_h(w) = \tilde{u},\ u_l(w) = \hat{u},\ \text{and}\ \rho_h(\tilde{u})\mathcal{D}u_h(w,\ell) < \rho_l(\hat{u})\mathcal{D}u_l(w,\ell),$$

which by Lemma 1 implies negative selection.

3.3 A Sufficient Statistic for improved Selection

Theorem 3 allows us to construct examples under which any steepening of incentives improves selection. However, it relies on knowledge about the outside option distributions, which are not typically observable. The following result finds conditions over observables to assess whether a particular adjustment to the status quo contract improves or harms selection.

Before we formulate the principal's payoff, we introduce some notations. For a given contract w, we denote by $T(w)$ the *total* mass of applicants, by $R(w)$ the mass of *rejected* applicants, and by $A(w)$ the share of agents who are approved among the ones who take the hiring test. That is,

$$\begin{aligned} T(w) &= pG_h(u_h(w)) + (1-p)G_l(u_l(w)) \\ R(w) &= r_h pG_h(u_h(w)) + r_l(1-p)G_l(u_l(w)) \\ A(w) &= \frac{T(w) - R(w)}{T(w)}. \end{aligned}$$

Theorem 4. *A local adjustment of w in direction ℓ improves selection if and only if $\mathcal{D}A(w,\ell) \geq 0$.*

Theorem 4 shows that selection improves if and only if the *share* of agents who are approved by the hiring test increases. Intuitively, if a change in the contract generates an increase in the approval rate, then it must have attracted more (fewer) high types than low types since high types are less likely to be rejected.

Theorem 4 implies that knowing how the approval rate changed suffices to answer whether a given local contract adjustment has improved or harmed selection. In particular, answering whether selection has improved does not require knowledge of many of the model's primitives, including the quality of the principal's screening technology, the distribution of outside options, the prevalence of each type in the market, the workers' utility function, or each type's effort cost function.

Theorem 4 tells us that the principal can tell if selection improved even before employees get to work –> can use this information in various ways –> discuss

3.4 Adjustments that improve Selection

Theorem 4 provides a test of whether an adjustment *in a given direction* ℓ has improved or harmed selection. We extend the result to identify the information needed to characterize *all* local adjustments that do not harm selection.

Suppose the firm knows the workers' monetary utility function, $v : \mathbb{R} \to \mathbb{R}$, and its own screening technology, r_l and r_h. Moreover, suppose the firm observes the outcome data generated under a status quo contract w, where the outcome data consists of the distribution of output generated by each worker-type, $f(\cdot|a_l(w))$ and $f(\cdot|a_h(w))$, and the masses of total and rejected applicants $T(w)$ and $R(w)$.

We consider data from one experiment (Experiment 1) where the firm marginally changes the status quo contract in direction ℓ. Upon conducting such an experiment, the firm observes how the total and rejected masses of applicants change; that is, it observes $\mathcal{D}T(w,\ell)$ and $\mathcal{D}R(w,\ell)$. We shall argue that the data from this experiment suffices for the firm to infer all adjustments to w that do not hurt selection.

By the Envelope Theorem, marginally changing the contract in direction $\hat{\ell}$ changes the utility of a type-t worker by

$$\mathcal{D}u_t(w,\hat{\ell}) = \int v'\big(w(x)\big)\hat{\ell}(x)f(x|a_t)dx \quad \forall t \in \{l,h\}. \tag{6}$$

Hence, even without information about the effort costs, we can compute how the utility of each agent's type varies by marginally changing the contract in any direction $\hat{\ell}$. Next, we combine this observation with the data observed in Experiment 1 to establish the following result:

Theorem 5. *Consider the data generated by Experiment 1, and suppose that $\mathcal{D}u_t(w,\ell) \neq 0$ for all $t \in \{l,h\}$. Then, a local change in direction $\hat{\ell}$ weakly improves selection if and only if*

$$\mathcal{D}u_h(w,\hat{\ell})\underbrace{\frac{r_l\mathcal{D}T(w,\ell)-\mathcal{D}R(w,\ell)}{[r_lT(w)-R(w)]\mathcal{D}u_h(w,\ell)}}_{:=K_h} \geq \mathcal{D}u_l(w,\hat{\ell})\underbrace{\frac{\mathcal{D}R(w,\ell)-r_h\mathcal{D}T(w,\ell)}{[R(w)-r_hT(w)]\mathcal{D}u_l(w,\ell)}}_{:=K_l}. \tag{7}$$

Note that K_h and K_l depend only on ℓ and not on $\hat{\ell}$. This implies that by observing the data from a single experiment, the principal can find *all* adjustments that do not harm selection. Moreover, (7) is linear in $\hat{\ell}$, implying not only that it is a simple condition to check but also that the set of directions that do not harm selection is convex.

4 Optimal Local Adjustments

So far, we have explored when a local adjustment improves or harms selection and what information the principal needs to make this assessment. In this section, we will be interested in the profit-maximizing adjustment to the status quo contract. When the principal adjusts w in the direction $\hat{\ell}$, assuming effort responses are differentiable, the effect in her payoff can be decomposed into three terms:[4]

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

The incentive effect is the change in profit that stems from workers adjusting their effort, holding the workforce composition fixed. The selection effect is the variation in profits due to changes the workforce composition, holding effort fixed. Finally, the direct effect computes the direct cost of changing the payments in direction $\hat{\ell}$.

Ideally, the principal would like to adjust the status quo contract in the direction that leads to the largest profit gain. However, computing $\mathcal{D}\pi(w,\hat{l})$ for all possible directions $\hat{l}$ demands substantial information about the environment. As a simplification, we find the adjustment that maximizes the sum of incentive and direct effects, subject to not hurting selection. This can be interpreted as adjusting the contract in a way that maximizes short-term gains without harming selection, which may have significant long-run effects. We call this the *prescriptive problem (PP)*.

Denote by $I^{\hat{\ell}}(w)$ the incentive effect on direction $\hat{\ell}$ under the status quo contract w. We can then write the prescriptive problem as:

$$\max_{\hat{\ell}:\|\hat{\ell}\|\leq 1} \left\{ I^{\hat{\ell}}(w) - \int \hat{\ell}(x)[q(w)f(x|a_h) + (1-q(w))f(x|a_l)]dx \right\} \qquad (PP)$$

[4]We shall soon impose a sufficient condition for effort responses to be differentiable (Condition 1). We present the argument in this order for expositional convenience.

subject to

$$\mathcal{D}q(w,\hat{l}) \geq 0, \tag{NHS}$$

where (NHS) is the *not-harming selection* constraint.

Recall that Theorem 5 allows us to represent (NHS) as a linear inequality that can be constructed using the data generated by Experiment 1. However, to solve (PP), we still need to characterize how the agent's incentives change for every direction $\hat{\ell}$.

To achieve that, we must introduce an additional condition and a second experiment. The following condition assumes that the output distribution is affine in effort, which implies that marginal incentives are independent of the effort level.

Condition 1. *The output distribution $f(x|a)$ is affine in a, that is, $f(x|a) = h_1(x) + ah_0(x)$ for some $h_0(x)$ and $h_1(x)$ satisfying $\int h_0(x)dx = 0$ and $\int h_1(x)dx = 1$.*

An important implication of this condition is that the agent's effort choice is fully characterized by its first-order condition. Hence, optimal effort is implicitly characterized by $c_t'(a_t) = \int v\big(w(x)\big)h_0(x)dx$. Therefore, locally adjusting a contract w in the direction ℓ changes the agent's effort by

$$\mathcal{D}a_t(w,\ell) = \frac{\int \ell(x)v'\big(w(x)\big)h_0(x)dx}{c_t''(a_t)}.$$

We can then write the effort response in any direction $\hat{\ell}$ as a function of the effort response in direction ℓ. That is,

$$\mathcal{D}a_t(w,\hat{\ell}) = \underbrace{\frac{\mathcal{D}a_t(w,\ell)}{\int \ell(x)v'(w(x))h_0(x)dx}}_{\text{Does not depend on } \hat{\ell}} \cdot \int \hat{\ell}(x)v'(w(x))h_0(x)dx. \tag{8}$$

Equation (8) provides a path to recover effort responses in all directions upon observing the responses from a single direction. However, we must still recover the effort responses in a given direction. The main challenge is that changing contracts affects not only incentives but simultaneously the skill composition of the applicant group. We shall show that this challenge can be addressed with a second experiment (Experiment 2) that keeps the workforce composition fixed while affecting incentives.

Consider a second experiment where the principal advertises the position at the original contract w but, after hiring, changes the worker's contracts in direction ℓ^+, where ℓ^+ ensures that both types are strictly better off and $\int \ell^+(x)v'\big(w(x)\big)h_0(x)dx > 0$. As locally changing contracts in direction ℓ^+ strictly benefits both types, all hired types still accept the job after the change. Moreover, as the change occurs after workers are hired, it does not affect the composition of the applicant pool and, hence, the share of high types among the hired workers. Finally, the relevant data generated by this experiment is how the average output (effort) and the distribution of outputs change when the contract is adjusted in direction ℓ^+ while keeping q fixed at the status quo. That is, upon running Experiment 2, the principal observes

$$\mathcal{D}\bar{a}(w, \ell^+) = q\mathcal{D}a_h(w, \ell^+) + (1 - q)\mathcal{D}a_l(w, \ell^+),$$

and

$$\mathcal{D}\bar{f}(w, \ell^+)(x) = qh_0(x)\mathcal{D}a_h(w, \ell^+) + (1 - q)h_0(x)\mathcal{D}a_l(w, \ell^+) = h_0(x)\mathcal{D}\bar{a}(w, \ell^+),$$

where $\bar{a}(w)$ and $\bar{f}(w)$ respectively denote the average effort and distribution of outputs under status quo contract w.

From Experiment 2, the principal can recover not only the average effort response $\mathcal{D}\bar{a}(w, \ell^+)$ but also the function $h_0(x) = \mathcal{D}\bar{f}(w, \ell^+)(x)/\mathcal{D}\bar{a}(w, \ell^+)$. However, even with the data generated by Experiment 2, we still cannot fully reconstruct how each type responds to a local adjustment of the status quo contract in any direction, since the experiment only reveals the average effort response and not the type-specific response. Nevertheless, the following result shows that the data generated from Experiments 1 and 2 suffices to solve (PP).

Theorem 6. *Let w be the status quo contract and assume Condition 1 holds. The data generated by Experiments 1 and 2 is a sufficient statistic for problem (PP). Moreover, there exists $\lambda^*, \mu^* \geq 0$ such that the solution*

$$\hat{\ell}^*(x) = C \times \left[\mu^* \frac{h_0(x)}{f(x)} + \lambda^* \frac{[f(x|a_h)K_h - f(x|a_l)K_l]}{f(x)} - \frac{1}{v'\big(w(x)\big)}\right] f(x)v'\big(w(x)\big),$$

where C is a constant characterized in the Appendix, and K_h and K_l are defined in (7).

Theorem 6 constructs the local adjustment direction that increases the most the sum of the incentive and direct effects subject to not hurting selection. The optimal local adjustment is in the direction of a modified Holmström-Mirlees-type contract. While the traditional Holmström-Mirrlees optimal contract (Mirrlees, 1999 and Holmström, 1979), balances incentives and insurance, here, the optimal adjustment shifts payments to outputs that increase incentives but provide sufficiently more utility to high types than low types to ensure positive selection. The optimal way to balance these two considerations is determined by the coefficients λ^* and μ^*, which are characterized in the Appendix.

The prescriptive problem aims to construct improvements to the status quo contract using only observable information stemming from Experiments 1 and 2. An appealing property of this approach is that it guarantees that, regardless of the unobservable primitives, i.e., G_t and $c_t(\cdot)$ for $t \in \{l, h\}$, the resulting direction of adjustment will not decrease profits. In particular, if the status quo contract is optimal, the prescriptive approach will recommend that no changes are made to the contract. The following result formalizes these assertions.

Corollary 1. *Suppose that the status quo contract satisfies*

$$\int [x - w(x)][f(x|a_h) - f(x|a_l)]dx \geq 0. \tag{9}$$

Then $\mathcal{D}\pi(w, \hat{\ell}^*) \geq 0$*. Moreover, if* $w \in \underset{\tilde{w}}{argmax}\ \{\pi(\tilde{w})\}$*, then* $\hat{\ell}^*(x) = 0$ *for all* $x \in X$*.*

Our approach looks for adjustments such that the gain in incentives outweighs the direct cost. Moreover, as we look only for directions that do not harm selection, profits will not decrease due to worse selection as long as the principal benefits from a more able workforce (condition (9)). Intuitively, we can think about the principal's profit depending on two dimensions: how strongly it motivates workers and the workforce skill composition. Our procedure finds local profit increases whenever you can improve the former without worsening the latter.

The prescriptive problem has two important limitations. First, as it only looks for local improvements, it would not prescribe any change when the status quo contract is a local but not a global maximum. The second issue is that we would miss adjustments that increase profits while harming selection. However, finding global maxima and fully considering the trade-off between incentives and selection is a complex task that requires detailed knowledge of the environment, including knowing the workers' effort costs and outside option distributions. The advantage of our approach is that imposes more modest informational demands.

5 Conclusion

This paper speaks to the dual role of incentive schemes in motivating workers and shaping a firm's workforce. Our findings challenge the straightforward intuition that steeper incentives automatically improve workforce quality by attracting more skilled workers. Instead, we show that depending on the distribution of outside options, increasing incentives may actually lower the average skill level of the workforce.

We then establish conditions under which steeper incentives improve the workforce's skill composition. We show that the reverse hazard rate of outside option distributions plays a critical role, as it determines whether high-skilled workers will apply in greater numbers relative to low-skilled ones in response to a steepening of incentives. This nuance highlights the importance of considering external labor market conditions and the shape of outside options when designing incentive schemes.

Our contribution goes beyond theoretical insights and offers a practical test for firms to assess whether changes to their incentive structure have improved or harmed workforce composition. We show that a simple comparison between the elasticities of total and rejected masses of applicants with respect to the payment scheme is sufficient to determine the direction of the change in the skill distribution in the firm. This test enables firms to make informed decisions about the effects of incentive changes on selection, even in the absence of detailed information about workers' outside options or effort costs.

Our approach also offers broader implications for firms seeking to improve their incentive schemes. We show that when output is affine in effort, firms can, from two simple experiments, construct improvements to their compensation scheme that increase the incentives without harming the skill composition of their workforce. The first experiment adjusts incentives prior to hiring, allowing the firm to characterize all directions of changes that do not harm selection. The second experiment, executed post-hiring, offers additional pay conditional on good performance and permits firms to calculate effort responses. Combining the data generated by both, firms can construct the best increase of incentives among all non-selection-harming ones.

We hope that our insights are taken to the data by future empirical research, particularly by studying monetary incentive effects across different industries and labor markets where the distribution of outside options may vary significantly. Such studies could further refine our understanding of how incentive schemes interact with labor market dynamics.

Bibliography

Alfitian, J., D. Sliwka, and T. Vogelsang (2024). When bonuses backfire: Evidence from the workplace. *Management Science 70*(9), 6395–6414.

Ashraf, N. and O. Bandiera (2018). Social incentives in organizations. *Annual Review of Economics 10*(Volume 10, 2018), 439–463.

Bandiera, O., I. Barankay, and I. Rasul (2005, 08). Social Preferences and the Response to Incentives: Evidence from Personnel Data*. *The Quarterly Journal of Economics 120*(3), 917–962.

Bénabou, R. and J. Tirole (2003). Intrinsic and extrinsic motivation. *The review of economic studies 70*(3), 489–520.

Bénabou, R. and J. Tirole (2006, December). Incentives and prosocial behavior. *American Economic Review 96*(5), 1652–1678.

Casadesus-Masanell, R. (2004). Trust in agency. *Journal of Economics & Management Strategy 13*(3), 375–404.

Castro-Pires, H., H. Chade, and J. Swinkels (2024). Disentangling moral hazard and adverse selection. *American Economic Review 114*(1), 1–37.

Castro-Pires, H. and H. Moreira (2021). Limited liability and non-responsiveness in agency models. *Games and Economic Behavior 128*, 73–103.

Chetty, R. (2009). Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods. *Annual Review of Economics 1*(Volume 1, 2009), 451–488.

Dal Bó, E., F. Finan, and M. A. Rossi (2013, 04). Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service*. *The Quarterly Journal of Economics 128*(3), 1169–1218.

Deserranno, E. (2019, January). Financial incentives as signals: Experimental evidence from the recruitment of village promoters in uganda. *American Economic Journal: Applied Economics 11*(1), 277–317.

Finkelstein, A. and M. J. Notowidigdo (2019, 05). Take-Up and Targeting: Experimental Evidence from SNAP*. *The Quarterly Journal of Economics 134*(3), 1505–1556.

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

Frey, B. S. and F. Oberholzer-Gee (1997). The cost of price incentives: An empirical analysis of motivation crowding- out. *The American Economic Review 87*(4), 746–755.

Friebel, G., M. Heinz, M. Krueger, and N. Zubanov (2017, August). Team incentives and performance: Evidence from a retail chain. *American Economic Review 107*(8), 2168–2203.

Fryer, Roland G., J. (2011, 11). Financial Incentives and Student Achievement: Evidence from Randomized Trials *. *The Quarterly Journal of Economics 126*(4), 1755–1798.

Fryer, R. G. (2013). Teacher incentives and student achievement: Evidence from new york city public schools. *Journal of Labor Economics 31*(2), 373–407.

Georgiadis, G. (2024). *4: Contracting with moral hazard*, pp. 24 – 37. Cheltenham, UK: Edward Elgar Publishing.

Georgiadis, G. and M. Powell (2022, January). A/b contracts. *American Economic Review 112*(1), 267–303.

Gneezy, U. and A. Rustichini (2000). Pay enough or don't pay at all. *The Quarterly Journal of Economics 115*(3), 791–810.

Gottlieb, D. and H. Moreira (2022). Simple contracts with adverse selection and moral hazard. *Theoretical Economics 17*(3), 1357–1401.

Guiteras, R. P. and B. K. Jack (2018). Productivity in piece-rate labor markets: Evidence from rural malawi. *Journal of Development Economics 131*, 42–61.

Hamilton, B. H., J. A. Nickerson, and H. 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, A. C. (1964). The measurement of waste. *The American Economic Review 54*(3), 58–76.

Holmström, B. (1979). Moral hazard and observability. *The Bell journal of economics*, 74–91.

Kreps, D. M. (1997). Intrinsic motivation and extrinsic incentives. *The American Economic Review 87*(2), 359–364.

Lazear, E. P. (2000, December). Performance pay and productivity. *American Economic Review 90*(5), 1346–1361.

Lazear, E. P. (2018). Compensation and incentives in the workplace. *The Journal of Economic Perspectives 32*(3), 195–214.

Leuven, E., H. Oosterbeek, and B. 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.

Milgrom, P. and I. Segal (2002). Envelope theorems for arbitrary choice sets. *Econometrica 70*(2), 583–601.

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

Ortner, J. and S. Chassang (2018). Making corruption harder: Asymmetric information, collusion, and crime. *Journal of Political Economy 126*(5), 2108–2133.

Saez, E. (2001, 01). Using Elasticities to Derive Optimal Income Tax Rates. *The Review of Economic Studies 68*(1), 205–229.

Shearer, B. (2004). Piece rates, fixed wages and incentives: Evidence from a field experiment. *The Review of Economic Studies 71*(2), 513–534.

Sliwka, D. (2007, June). Trust as a signal of a social norm and the hidden costs of incentive schemes. *American Economic Review 97*(3), 999–1012.

A Omitted Proofs

Proof of Lemma 1. The first step is to show that $\mathcal{D}u_t(w,\ell)$ exists, which is done through two intermediate results we name as Claims.

Let

$$\psi(a,\varepsilon) = \int v\big(w(x)+\varepsilon\ell(x)\big)f(x|a)dx - c_t(a).$$

Note that $u_t(w+\varepsilon\ell) = \underset{a}{max}\ \psi(a,\varepsilon)$.

Claim 1. *The family of functions* $\{\partial\psi(a,\cdot)/\partial\varepsilon\}_{a\in[0,\overline{a}]}$ *is equidifferentiable at* $\varepsilon_0 \in [0,1]$ *and* $\underset{a\in[0,\overline{a}]}{sup}\ |\partial\psi(a,\varepsilon_0)/\partial\varepsilon| < +\infty$.

Proof. 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

$$\Big|\frac{\partial \psi(a,\tilde{\varepsilon})}{\partial \varepsilon} - \frac{\partial \psi(a,\hat{\varepsilon})}{\partial \varepsilon}\Big| \quad \leq \int \Big| v'\big(w(x)+\tilde{\varepsilon}\ell(x)\big) - v'\big(w(x)+\hat{\varepsilon}\ell(x)\big)\Big| \cdot |\ell(x)| \cdot f(x|a)dx$$

$$\leq |\tilde{\varepsilon}-\hat{\varepsilon}| \cdot \underset{x}{sup}|\ell^2(x)| \cdot \underset{y}{sup}|v''(y)|,$$

which concludes the Claim's proof. □

Claim 2. *$\mathcal{D}u_t(w,\ell)$ exists and*

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

Proof. Claim 1 shows that the conditions of Theorem 3 in Milgrom and Segal (2002) are satisfied. Hence, the right derivative of $u_t\big(w+\varepsilon_0\ell\big)$ exists and is equal to $\partial\psi(a_t,\varepsilon_0)/\partial\varepsilon$. In particular, at $\varepsilon_0 = 0$

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

□

As noted before, q increases if and only if G_h/G_l increases or, equivalently, $ln(G_h) - ln(G_l)$ is decreasing. Taking the Gateaux differential in the direction ℓ delivers the result. □

Proof of Theorem 2. Fix an arbitrary status quo contract w and a steepening of incentives ℓ for which $\mathcal{D}u_l(w,\hat{\ell}) \cdot \mathcal{D}u_h(w,\ell) > 0$. We then construct two distributions $G_h(\cdot) \succ_{FOSD} G_l(\cdot)$ such that adjusting w in direction ℓ harms selection.

Under contract w, each type-t gets a utility $u_t = u_t(w)$ if hired. Moreover, as $c_h(a) < c_l(a)$ for all a, we know that $u_l < u_h$. Also, as $c_h'(a) < c_l'(a)$, we have that $a_h \geq a_l$. By the definition of steepening of incentives we have that

$$\mathcal{D}u_h(w,\ell) = \int v'\big(w(x)\big) f(x|a_h)dx \geq \int v'\big(w(x)\big) f(x|a_l)dx = \mathcal{D}u_l(w,\hat{\ell}).$$

By Lemma 1, a local adjustment in direction ℓ harms selection if and only if

$$\frac{g_h}{G_h}\mathcal{D}u_h(w,\ell) < \frac{g_l}{G_l}\mathcal{D}u_l(w,\hat{\ell}).$$

The main argument in the proof is to construct distributions with reversed hazard rates g/G such that the inequality above holds. We then split the construction of G's into two possible cases: $\mathcal{D}u_h(w,\ell) > 0$ or $\mathcal{D}u_h(w,\ell) < 0$.

Case I: $\mathcal{D}u_h(w,\ell) > 0$.

Let $G_l(y) = exp\big(-[y+b]^{-\gamma}\big)$ and $G_h(y) = \big[G_l(y)\big]^{\delta}$, where $b > -u_l$, $\gamma > 0$ and $\delta > 1$. Note that as $\delta > 1$ and $G_l(y) < 1$ for all $y \geq -b$, then $G_h(\cdot) \leq G_l(\cdot)$, which implies that $G_h(\cdot)$ first-order stochastically dominates $G_l(\cdot)$. We then show that selection is harmed when γ is sufficiently large.

Note that

$$\frac{g_h}{G_h}\frac{G_l}{g_l} = \delta\left[\frac{u_h+b}{u_l+b}\right]^{-(\gamma+1)}.$$

Hence, for γ large enough

$$\frac{g_h}{G_h}\frac{G_l}{g_l} < \frac{\mathcal{D}u_l(w,\hat{\ell})}{\mathcal{D}u_h(w,\ell)} \implies \mathcal{D}q(w,\ell) < 0.$$

Case II: $\mathcal{D}u_h(w,\ell) < 0$.

Let $G(y|t) = exp\big(-\lambda_t[b-y]\big)$, where $b > u_h$ and $\lambda_h > \lambda_l > 0$. For any point in the support $G_h(y) < G_l(y)$. Hence, $G_h(\cdot)$ first-order stochastically dominates $G_l(\cdot)$. Moreover,

$g_t/G_t = \lambda_t$. Therefore, for a sufficiently large λ_h/λ_l

$$\frac{\lambda_h}{\lambda_l} > \frac{\mathcal{D}u_l(w,\hat{\ell})}{\mathcal{D}u_h(w,\ell)} \implies \mathcal{D}q(w,\ell) < 0.$$

□

Proof of Theorem 3.

The "if" part: Suppose that $\rho_h(\tilde{u}_h) \geq \rho_l(\tilde{u}_l)$ for all $\tilde{u}_h \geq \tilde{u}_l$.

Recall that

$$u_t(w) = \max_a \Big\{ \int v\big(w(x)\big) f(x|a)dx - c_t(a) \Big\}.$$

As $c_h(a) \leq c_l(a)$, we have that $u_h(w) \geq u_l(w)$ and $\rho_h(u_h(w)) \geq \rho(u_l(w))$.

Moreover, by the Envelope Theorem

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

Also, as $c_h'(a) < c_l'(a)$, then $a_h \geq a_l$.

The fact that ℓ is a steepening of incentives implies that $\mathcal{D}u_h(w,\ell) \geq \mathcal{D}u_l(w,\hat{\ell})$, while by the Theorem's statement $\mathcal{D}u_l(w,\hat{\ell}) > 0$. Hence,

$$\rho_h\big(u_h(w)\big)\mathcal{D}u_h(w,\ell) - \rho_l\big(u_l(w)\big)\mathcal{D}u_l(w,\hat{\ell}) =_s \rho_h\big(u_h(w)\big)\frac{\mathcal{D}u_h(w,\ell)}{\mathcal{D}u_l(w,\hat{\ell})} - \rho_l\big(u_l(w)\big) \geq \rho_h\big(u_h(w)\big) - \rho_l\big(u_l(w)\big) \geq 0,$$

which by Lemma 1 implies that selection must be improved.

The "only if" part: Suppose that there exists $\tilde{u}_h \geq \tilde{u}_l$ such that $\rho_h(\tilde{u}_h) < \rho_l(\tilde{u}_l)$. We will construct cost functions $c_l(\cdot)$ and $c_h(\cdot)$ such that selection is harmed by an adjustment in direction ℓ.

Let $c_t(a) = \beta_t(a+1)^2 + \gamma_t$. We now construct $(\beta_l, \beta_h, \gamma_l, \gamma_h)$ such that all properties assumed for the effort cost function are satisfied and $\rho_h(u_h(w))\mathcal{D}u_h(w, \ell) < \rho_l(u_l(w))\mathcal{D}u_l(w, \hat{\ell})$.

Let

$$\psi^*(\beta) = \underset{a \in [0, \overline{a}]}{max} \Big\{ \int v\big(w(x)\big) f(x|a) dx - \beta(a+1)^2 \Big\}, \text{ and}$$

$$a^*(\beta) = \underset{a \in [0, \overline{a}]}{argmax} \Big\{ \int v\big(w(x)\big) f(x|a) dx - \beta(a+1)^2 \Big\},$$

where $\beta > 0$. Let $\beta_H > inf\{\beta \in \mathbb{R}_{++} : a^*(\beta) = 0\}$ and $\beta_L = \beta_H + \varepsilon$, where $\varepsilon > 0$. Consequently, $a^*(\beta_H) = a^*(\beta_L) = 0$. Let $\gamma_t = \psi^*(\beta_t) - \tilde{u}_t$, which implies $u_t(w) = \tilde{u}_t$.

Note that for ε sufficiently small, $c_l(0) > c_h(0)$. Also, $c_t(a)$ is a valid cost function since

- c_t' and $c_t'' > 0$;
- $c_l(a) > c_h(a)$ and $c_l'(a) > c_h'(a)$.

Observe then that

$$u_t(w) = \underset{a}{max} \Big\{ \int v\big(w(x)\big) f(x|a) dx - c_t(a) \Big\} = \tilde{u}_t.$$

Therefore,

$$\rho_h(u_h(w))\mathcal{D}u_h(w, \ell) - \rho_l(u_l(w)\mathcal{D}u_l(w, \hat{\ell}) = \int v'(w(x))\ell(x) f(x|0) dx \Big[\rho_h(\tilde{u}_h) - \rho_l(\tilde{u}_l)\Big] < 0.$$

Hence, by Lemma 1, an adjustment in direction ℓ harms selection. □

Proof of Theorem 4. Recall that

$$T(w) = pG_h \circ u_h(w) + (1-p)G_l \circ u_l(w)$$

$$R(w) = r_h p G_h \circ u_h(w) + r_l(1-p)G_l \circ u_l(w).$$

Hence,

$$G_l \circ u_l(w) = \frac{R(w) - r_h T(w)}{(1-p)(r_l - r_h)} \quad \text{and} \quad G_h \circ u_h(w) = \frac{r_l T(w) - R(w)}{p(r_l - r_h)}. \tag{10}$$

Also,

$$\mathcal{D}T(w,\ell) = p[g_h \circ u_h(w)]\mathcal{D}u_h(w,\ell) + (1-p)[g_l \circ u_l(w)]\mathcal{D}u_l(w,\hat{\ell})$$

$$\mathcal{D}R(w,\ell) = r_h p[g_h \circ u_h(w)]\mathcal{D}u_h(w,\ell) + r_l(1-p)[g_l \circ u_l(w)]\mathcal{D}u_l(w,\hat{\ell}).$$

Hence,

$$[g_l \circ u_l]\mathcal{D}u_l(w,\hat{\ell}) = \frac{\mathcal{D}R(w,\ell) - r_h\mathcal{D}T(w,\ell)}{(1-p)(r_l - r_h)} \quad \text{and} \quad [g_h \circ u_h]\mathcal{D}u_h(w,\ell) = \frac{r_l\mathcal{D}T(w,\ell) - \mathcal{D}R(w,\ell)}{p(r_l - r_h)}. \tag{11}$$

Replacing (10) and (11) into (4), we get

$$\begin{aligned}
\mathcal{D}q(w,\ell) &=_s \frac{g_h \circ u_h}{G_h \circ u_h}\mathcal{D}u_h(w,\ell) - \frac{g_l \circ u_l}{G_l \circ u_l}\mathcal{D}u_l(w,\hat{\ell}) \\
&= \frac{r_l\mathcal{D}T(w,\ell) - \mathcal{D}R(w,\ell)}{r_l T - R} - \frac{\mathcal{D}R(w,\ell) - r_h\mathcal{D}T(w,\ell)}{R - r_h T} \\
&= \frac{(r_l - r_h)TR}{(r_l T - R)(R - r_h T)}\left[\frac{\mathcal{D}T(w,\ell)}{T} - \frac{\mathcal{D}R(w,\ell)}{R}\right] \\
&=_s \frac{\mathcal{D}T(w,\ell)}{T(w)} - \frac{\mathcal{D}R(w,\ell)}{R(w)}.
\end{aligned}$$

Where the last $=_s$ stems from $(r_l T - R)(R - r_h T) > 0$, which is a consequence of $G_l \circ u_l(w),\ G_h \circ u_h(w) > 0$.

Finally, note that

$$\mathcal{D}A(w,\ell) = \frac{-T(w) \times \mathcal{D}R(w,\ell) + R(w) \times \mathcal{D}T(w,\ell)}{\big(T(w)\big)^2} =_s \frac{\mathcal{D}T(w,\ell)}{T(w)} - \frac{\mathcal{D}R(w,\ell)}{R(w)}.$$

Hence, $\mathcal{D}q(w,\ell) =_s \mathcal{D}A(w,\ell)$. $\square$

Proof of Theorem 5. Recall that a marginal change in the contract in direction $\hat{\ell}$ improves selection if and only if

$$\mathcal{D}u_h(w,\hat{\ell})\frac{g_h}{G_h} \geq \mathcal{D}u_l(w,\hat{\ell})\frac{g_l}{G_l}.$$

By equation (6), we can construct $\mathcal{D}^{\hat{\ell}}_{u_t(w)}$ for any direction $\hat{\ell}$. It remains to find g_t/G_t as a function of observables. By equations (10) and (11), we have that

$$\frac{g_l}{G_l} = \frac{\mathcal{D}R(w,\ell) - r_h\mathcal{D}T(w,\ell)}{(R - r_hT)\mathcal{D}u_l(w,\hat{\ell})}, \quad \text{and} \quad \frac{g_h}{G_h} = \frac{r_l\mathcal{D}T(w,\ell) - \mathcal{D}R(w,\ell)}{(r_lT - R)\mathcal{D}u_h(w,\ell)},$$

which concludes the proof. □

Proof of Theorem ??. The proof is divided into two parts: first, we rewrite problem (PP) and argue that all the information needed to state the problem can be recovered from Experiments 1 and 2. Second, we characterize its solution.

Problem (PP) can be written as

$$\max_{\hat{\ell}} \left\{ \mu^* \cdot \int \hat{\ell}(x) v'\big(w(x)\big) h_0(x) dx - \int \hat{\ell}(x) f(x) dx \right\}$$

subject to

$$\int v'\big(w(x)\big)\hat{\ell}(x)\big[f(x|a_h)K_h - f(x|a_l)K_l\big]dx \geq 0,$$

$$\int \hat{\ell}^2(x)dx \leq 1,$$

where

$$\mu^* := \frac{\int [s - w(s)]h_0(s)ds}{\int \ell^+(s)v'\big(w(s)\big)h_0(s)ds} \cdot \underbrace{\Big[q\mathcal{D}a_h(w,\ell^+) + (1-q)\mathcal{D}a_l(w,\ell^+)\Big]}_{\equiv\, \mathcal{D}\bar{a}(w,\ell^+)}.$$

From Experiment 1, the firm can reconstruct K_h and K_l. From Experiment 2, the principal can recover $\mathcal{D}\bar{a}(w,\ell^+)$ and $h_0(\cdot)$. Therefore, the two experiments provide all the necessary information to solve problem (PP). We now find its solution.

Letting $\lambda \geq 0$ and $\nu \geq 0$ denote the dual multipliers associated with the first and second constraint, we have the Lagrangian

$$\mathcal{L}(\lambda, \nu) = \underset{\hat{\ell}}{max} \left\{ \nu + \int \hat{\ell}(x) \Big[v'\big(w(x)\big)\big(\mu^* h_0(x) + \lambda \big[f(x|a_h)K_h - f(x|a_l)K_l\big]\big) - f(x) - \nu \hat{\ell}(x) \Big] dx \right\}.$$

For any $\lambda, \nu \geq 0$, note that the integrand is differentiable and strictly concave. We can then maximize it pointwise with respect to $\hat{\ell}$, with each respective first-order condition delivering

$$\hat{\ell}_{\lambda,\nu}(x) = \frac{\Big[\mu^* h_0(x) + \lambda\big(f(x|a_h)K_h - f(x|a_l)K_l\big)\Big] v'\big(w(x)\big) - f(x)}{2\nu}.$$

Next, we find the optimal λ and ν by solving the dual problem:

$$\underset{\lambda \geq 0, \nu \geq 0}{min} \mathcal{L}(\lambda, \nu).$$

This problem is convex, and using $\hat{\ell}_{\lambda,\nu}$, the solution to the dual problem is

$$\lambda^* = max\left\{0, \frac{\int \big[f - \mu^* h_0\big(f_h K_h - f_l K_l\big) v'\big(w\big)\big] \big[f_h K_h - f_l K_l\big] v'\big(w\big) dx}{\int \big[v'\big(w\big)\big]^2 \big[f_h K_h - f_l K_l\big]^2 dx}\right\}$$

and

$$\nu^* = \frac{1}{2}\sqrt{\int \Big\{ v'\big(w(x)\big)\big[\mu^* h_0(x) + \lambda^*\big(f(x|a_h)K_h - f(x|a_l)K_l\big)\big] - f(x) \Big\} dx}.$$

Thus, the optimal adjustment direction is

$$\hat{\ell}^*(x) = \frac{\Big[\mu^* h_0(x) + \lambda^*\big(f(x|a_h)K_h - f(x|a_l)K_l\big)\Big] v'\big(w(x)\big) - f(x)}{\sqrt{\int \Big\{ v'\big(w(x)\big)\big[\mu^* h_0(x) + \lambda^*\big(f(x|a_h)K_h - f(x|a_l)K_l\big)\big] - f(x) \Big\} dx}},$$

which is proportional to $L(x, \lambda^*, \nu^*)$.

Up to now, we have shown that $\hat{\ell}^*$ solves the dual problem. To show it solves the primal problem given in (PP), we will now establish that strong duality holds. Denote the optimal value of the primal by Π^*. First, by weak duality, we have that $\mathcal{L}(\lambda^*, \nu^*) \geq \Pi^*$. Second, it is straightforward to check that $\hat{\ell}^*$ is feasible for problem (PP), and that λ^* and ν^* are strictly positive if and only if the respective (primal) constraint binds; meaning that the complementary slackness conditions are satisfied. This implies that $\mathcal{L}(\lambda^*, \nu^*) \leq \Pi^*$. Therefore, $\mathcal{L}(\lambda^*, \nu^*) = \Pi^*$, which proves that strong duality holds, and $\hat{\ell}^*$ solves (PP). $\square$

Proof of Corollary 1. Note that (NHS) and (9) imply that

$$\left[\int [x - w(x)][f(x|a_h) - f(x|a_l)]dx + \gamma'\big(q(w)\big) \right] \mathcal{D}q(w, \hat{\ell}^*) \geq 0.$$

Hence,

$$\mathcal{D}\pi(w, \hat{\ell}^*) \geq I^{\hat{\ell}^*}(w) - \int \hat{\ell}^*(x)[q(w)f(x|a_h) + (1 - q(w))f(x|a_l)]dx \geq 0,$$

where the final inequality stems from $\hat{\ell}(x) = 0$ for all $x \in X$ being feasible in problem (PP). Therefore, $\mathcal{D}\pi(w, \hat{\ell}^*) \geq 0$.

It remains to show that if the status quo contract w is optimal, we have that $\hat{\ell}^*(x) = 0$ for all $x \in X$.

Suppose $w \in \underset{\tilde{w}}{argmax} \, \{\pi(\tilde{w})\}$. As w maximizes π, it must be that $\mathcal{D}^{\hat{\ell}}_{\pi(w)} = 0$ for any $\hat{\ell}$. As a consequence, any $\hat{\ell}$ that satisfies (NHS) must be such that

$$I^{\hat{\ell}}(w) - \int \hat{\ell}(x)[q(w)f(x|a_h) + (1 - q(w))f(x|a_l)]dx \leq 0.$$

Therefore, $\hat{\ell}^*(x) = 0$ for all $x \in X$ solves problem (PP). $\square$