# Incentives and Selection*

Henrique Castro-Pires† and George Georgiadis‡

February 4, 2025

Click here for the latest version

**Abstract**

Performance pay schemes influence not only workers' incentives to exert effort but also the composition of the workforce in terms of skill. Conventional wisdom suggests that steepening incentives should attract higher-skilled workers. However, we show that this is not universally true: under certain conditions, steeper incentives reduce the average skill level of the workforce. We identify sufficient conditions on observables such that a marginal adjustment to the pay scheme improves the skill composition of the workforce. Building on these insights, we determine the optimal adjustment to incentives that enhances performance without compromising worker selection.

*We are grateful to Eduardo Azevedo, Arjada Bardhi, Rahul Deb, 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, and the Game Theory & Information Economics Conference at EPGE/FGV-Rio de Janeiro.

†University of Miami

‡Northwestern University

# 1 Introduction

The design of appropriate incentives is fundamental to an organization's success. In addition to influencing workers' effort—what we refer to as the "incentive effect"—the structure of an incentive plan also shapes the pool of job applicants, which we term the "selection effect." Therefore, crafting an incentive scheme requires that firms account for both of these considerations.

In a seminal study, Lazear (2000) examined Safelite Glass Corporation's switch from hourly wages to piece rates for its key workers. This change led to a 44% increase in overall productivity, with half of this gain attributed to the firm attracting more productive workers. The mechanism is intuitive: a higher-powered incentive scheme is more attractive to higher-skilled workers, who are in a better position to increase their productivity, and thereby, their pay. We show that this logic need not always hold, and instead, steeper incentives may worsen the workforce's skill composition.

To introduce our main ideas, consider the following example. Suppose you manage a firm and pay workers according to a linear contract $w(x) = \beta + \alpha x$, where $\beta$ is a fixed salary, $\alpha$ is a piece rate, and $x$ is output. Once on the job, workers choose how much effort to exert, where effort raises output but is costly. Suppose also that workers differ in their skill level, with high-skilled workers having a lower marginal cost of effort than low-skilled workers. You post a job search announcement, and prospective workers apply for the job if they prefer it to their outside option. Then you implement a screening test that high types pass with higher probability than low types, and you hire randomly among the applicants who passed the test. You are considering raising $\alpha$ and want to know what will be the effect on the *quality* of your workforce; i.e., the share of high types among the hired workers.

As high types have a lower marginal cost of effort, they benefit more than low types from the steepening of incentives, so you might conclude that this will improve the quality of your workforce. However, increasing $\alpha$ improves the contract for both types, and as a result, the job posting now attracts both more high-skilled and more low-skilled applicants. Moreover, depending on the distribution of the outside option for each type, it is possible that more additional low types than high types apply, in which case the quality of the workforce worsens. The example demonstrates that the sign of the selection effect following a steepening of incentives depends not only on who benefits the most, but also on the shape of the outside option distributions.

We extend our results beyond linear contracts by considering how local perturbations to any given contract in an arbitrary—possibly nonlinear—direction affects efforts, the share of high types in the workforce, and profits. This approach allows us to define a notion of "steepening of incentives" for contracts other than linear ones. We say that a given change in the incentive scheme is a *steepening of incentives* if it increases the incentives for effort for both worker types.

Our first main result, Theorem 1, shows that for any *status quo* contract and perturbation that steepens incentives, there always exists outside option distributions that worsen selection. In Theorem 2, we also establish a partial converse result: we provide a condition over the distribution of outside options such that any steepening of incentives must improve selection. This condition requires that the reverse hazard rate (i.e., ratio between the density and cumulative distribution) 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.

Although useful, the previous results hinge on knowledge of the outside option distributions. We identify sufficient statistics over plausibly observable objects that allow us to test whether a given perturbation to a status quo contract has improved or harmed selection. Theorem 3 shows that selection strictly improves (worsens) if and only if the total mass of applicants increases in percentage terms more than the total mass of rejected applicants.[1] Intuitively, as high types pass the screening test more frequently than low types, a stronger response on the total mass of applicants compared to the mass of rejected applicants implies that the contract change has attracted proportionally more high types than low types, and hence, selection must have improved.

Theorem 3 enables one to test whether perturbing the contract in a given direction improves or harms selection, but it is silent about other perturbations. Theorem 4 shows that by observing the outcomes of a single directional change in the contract—which we refer to as a single *experiment*—the firm can characterize *all* perturbations that (weakly) improve selection.

We then turn to the profit-maximizing perturbation subject to not harming selection. To do so, we decompose the effect of locally perturbing an incentive scheme on profits into three parts:

$$\text{Total effect} \;=\; \text{Incentives Effect} \;+\; \text{Selection Effect} \;+\; \text{Direct Effect}.$$

The incentive effect captures the profit change due to workers adjusting their efforts while keeping the workforce composition fixed. The selection effect accounts for the change in the workforce composition, while holding efforts fixed. Finally, the direct effect accounts for the cost due to changing the contract holding efforts and selection fixed.

Ideally, the firm would like to adjust its incentive scheme in the profit-maximizing direction; i.e., in the direction in which the "total effect" is the largest. However, fully characterizing each of these effects requires substantial information about the environment, including features that are difficult to observe such as the distribution of outside options

[1] Equivalently, selection improves (worsens) if the elasticity of the total mass of applicants (with respect to incentives) is larger (smaller) than that of the rejected applicants.

and the workers' effort cost functions. Instead, we consider maximizing the sum of the incentive and direct effects subject to not harming selection. Doing so increases workers' efforts without worsening the composition of the workforce, which may have important long-run consequences on firm performance.

Our final result—Theorem 5—shows that when the output distribution is affine in effort, the firm only needs to observe two experiments to characterize the non-harmful-selection profit-maximizing direction of contract adjustment. The first experiment is an arbitrary perturbation in the original contract, allowing the firm to construct all directions that do not harm selection. In the second experiment, the firm offers additional performance bonuses to workers after they have been hired so that it can back out their effort responses without affecting selection into the firm. With the characterization of all directions that do not harm selection and the respective effort responses at hand, we construct the optimal direction of contract adjustment.

Our paper is related to the literature studying the role of monetary incentives in the productivity of organizations. Several papers document productivity gains following an increase in incentives, for example, Lazear (2000, 2018), Shearer (2004), Bandiera et al. (2005), and Friebel et al. (2017). However, others also document monetary incentives being ineffective or even backfiring (e.g., Leuven et al. (2010), Fryer (2011, 2013), and Alfitian et al. (2024)). Many theoretical explanations for the backfiring of performance pay have been raised in the literature, such as 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)), social norms (e.g., Gneezy and Rustichini (2000), Sliwka (2007)), and social preferences/peer pressure (e.g., Hamilton et al. (2003), Ashraf and Bandiera (2018)). We contribute to this literature by showing that the negative effect on firm performance can also stem from worsening the firm's workforce composition, even in a setting with canonical rational agents.

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 evidence is mixed. Following increases in financial rewards, Dal Bó et al. (2013) find evidence of improved selection of civil servants in Mexico, Guiteras and Jack (2018) document no improvement in the context of informal labor in rural Malawi, while Deserranno (2019) finds a negative selection among health-promoters in Uganda. We contribute to 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 absent intrinsic motivation considerations, selection might be harmed depending on the shape of the workers' outside option distributions. Second and most importantly, we provide a simple test for improved selection. A major challenge for studying selection is getting enough data to evaluate it. Our Theorem 3 provides a parsimonious test that could potentially be used in future empirical studies.

On a methodological ground, this paper is connected to the literature on sufficient statistics using a variational approach, which exploits envelope conditions from agents' optimization problems to characterize behavioral responses and optimal policies in terms of a few model parameters, typically elasticities; see Chetty (2009) for a review. This approach dates back to Harberger (1964) measuring deadweight losses of commodity taxes and was also used to study income-taxation (Saez (2001)), designing policies to fight corruption (Ortner and Chassang (2018)), welfare programs (Finkelstein and Notowidigdo (2019)) among other applications. The closest paper to ours is Georgiadis and Powell (2022), who bring these tools to analyze a moral hazard problem. Our main contributions are twofold: we incorporate the selection of new workers in the analysis, and we characterize sufficient statistics for improved (harmed) selection.

# 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 $\lambda > 0$ 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 a job.

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, $a \in [\underline{a}, \overline{a}] \subset \mathbb{R}_+$ 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:

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

Before we formulate the principal's payoff, we introduce some notations. For a given contract $w$, we denote by $A(w)$ the *total* mass of applicants, by $R(w)$ the mass of *rejected* applicants, and by $q(w)$ the share of high types among the hired agents. That is,

$$A(w) = pG_h(u_h(w)) + (1-p)G_l(u_l(w))$$

$$R(w) = r_h pG_h(u_h(w)) + r_l(1-p)G_l(u_l(w))$$

$$q(w) = \frac{(1-r_h)pG_h(u_h(w))}{A(w) - R(w)}$$

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

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

where $a_t(w)$ denotes the optimal effort of a type-$t$ agent, and $\gamma(\cdot)$ is a non-decreasing differentiable function that captures the continuation payoff of having share $q$ of high-ability workers.

We will study how the share of high types among the hired agents $q$ and the principal's payoff $\pi$ respond to changes in some arbitrary status quo contract $w$.

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., $c_l(a) \geq c_h(a)$ and $c_l'(a) > c_h'(a)$ for all $a \in [\underline{a}, \overline{a}]$.

A.3. The density function $f(\cdot|a)$ has full support over the output space $X \subseteq \mathbb{R}_+$, 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 effort is interpreted as an agent's expected output.

A.4. At the status quo contract $w$, strictly positive masses of both types apply for the job —i.e., $G_h\big(u_h(w)\big), G_l\big(u_l(w)\big) > 0$— and the mass of approved workers is strictly larger than the vacancies, i.e., $\lambda < A(w) - R(w)$.

We will say that a change in the contract causes *positive selection* if it causes $q(\cdot)$ to rise, and it causes *negative selection* otherwise.

## 2.1 A Useful Observation

We now establish a preliminary result that is informative of the key determinants of selection. Notice that the share of high types among the hired agents under the status quo contract $w$ can be rewritten as

$$\begin{aligned} q(w) =& \frac{(1-r_h)pG_h\big(u_h(w)\big)}{(1-r_h)pG_h\big(u_h(w))\big) + (1-r_l)(1-p)G_l\big(u_l(w)\big)} \\ =& \frac{(1-r_h)pG_h\big(u_h(w))\big)/G_l\big(u_l(w)\big)}{(1-r_h)pG_h\big(u_h(w)\big)/G_l\big(u_l(w)\big) + (1-r_l)(1-p)}, \end{aligned}$$

where the first equality stems from the definitions of $A(\cdot)$ and $R(\cdot)$, and the second from dividing numerator and denominator by $G_l(u_l(w)))$. We immediately obtain that a change in the contract leads to positive selection if and only if it causes $G_h(u_h(\cdot)))/G_l(u_l(\cdot)))$ to increase.

We use this observation to evaluate how a change in the status quo contract affects selection. It suffices to inspect how the change affects the ratio between the shares of high and low types who chose to apply, while the precision of the principal's screening test (i.e., $r_l$ and $r_h$) and the fraction of high types in the population are immaterial.

# 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.

### An Example

We consider a class of examples where agents are risk-neutral, have isoelastic effort costs, and the principal offers a linear contract. That is,

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

where $\theta_l \leq \theta_h$ and $\gamma > 0$.

Using that $a \equiv \mathbb{E}[x|a]$ and solving for the agent's optimal effort, it follows that[2]

$$u_t(\alpha) = \frac{\alpha\gamma}{1+\gamma} (\alpha\theta_t)^{1/\gamma} \quad \text{and} \quad \frac{du_t}{d\alpha} = (\alpha\theta_t)^{\frac{1}{\gamma}}. \tag{1}$$

Consider an increase of the slope $\alpha$ to some larger $\hat{\alpha}$. In response, the expected utility of both types increases, so more low and high types apply. Let $u_t$ and $\hat{u}_t$ denote the utility of type $t$ before and after the steepening, respectively. As $du_t/d\alpha$ is increasing in $\theta_t$, we have that high types benefit more from the steepening of incentives than the low types. However, a larger gain in utility by the high types is not sufficient to guarantee an

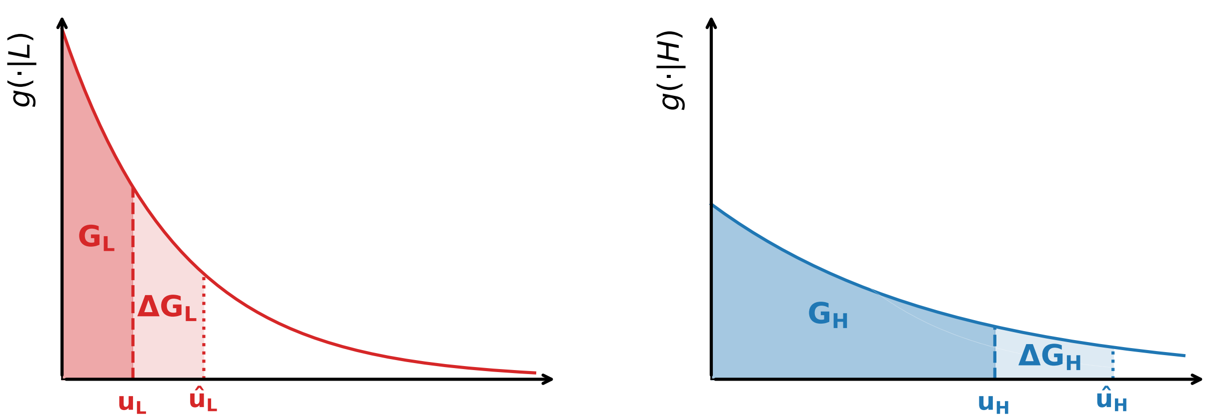
[2]We abuse notation and write $u_t(\alpha)$ to simply denote the expected utility of a type-$t$ agent when offered a linear contract with slope $\alpha$.

improvement in selection. As argued before, selection improves if and only if the ratio between the fraction of high types and low types that apply $G_h\big(u_h(\cdot)\big)/G_l\big(u_l(\cdot)\big)$ increases.

We illustrate this point in Figure (1). Subfigures (1a) and (1b) display each type's outside option density function, where the dark shaded areas represent $G(u_l)$ and $G(u_h)$ under the status quo contract, while $\Delta G_l$ and $\Delta G_h$ denote the additional mass of applicants stemming from the steepening.[3] The change in the contract will improve selection if and only if

$$\frac{G(u_h)}{G(u_l)} < \frac{G(u_h) + \Delta G(u_h)}{G(u_l) + \Delta G(u_l)}. \tag{2}$$

Note, however, that despite the gain in utility being larger for the high type, i.e., $\hat{u}_h - u_h > \hat{u}_l - u_l$, we may have $\Delta G(u_h) < \Delta G(u_l)$. Hence, (2) need not to hold. A smaller increase in the utility offered to low types than to high types might still attract a larger mass of the former, potentially worsening selection.




(a) Low type

(b) High type

Figure 1: Outside option distributions

In a nutshell, how a change in the contract affects selection depends not only on how much it benefits each type but also on the shape of the outside option distributions.

[3]In Figure 1, we assume $g_l(u) = exp(-u)$, $g_h(u) = (1/2)exp(-u/2)$, and $(u_l, \hat{u}_l, u_h, \hat{u}_h) = (0.6, 1.2, 2, 3)$; i.e., the outside option distribution of high types first-order stochastically dominates that of low types.

If we further assume type-independent outside options, i.e. $G_h \equiv G_l := G$, the contract change will generate negative selection if

$$\frac{d}{d\alpha}\frac{G(u_h(\alpha))}{G(u_l(\alpha))} =_s \frac{g(u_h)}{G(u_h)}\frac{du_h}{d\alpha} - \frac{g(u_l)}{G(u_l)}\frac{du_l}{d\alpha} =_s u_h\rho(u_h) - u_l\rho(u_l) < 0, \tag{3}$$

where $\rho := g/G$ is the reverse hazard rate function of $G$.[4] This derivative is negative if (but *not* only if) $u\rho(u)$ decreases in $u$. Thus, increasing the slope of the contract leads to *negative* selection if the reverse hazard rate function, $\rho$, has elasticity smaller than $-1$.[5]

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

Intuitively, if effort were inelastic, then both types would obtain the same amount of extra utils from an increase in the slope of the contract, and hence selection would be (unambiguously) improved if $G$ were log-convex, and it would be (unambiguously) harmed if $G$ were log-concave.[6] In general, however, high types obtain a larger amount of extra utils than low types. If the reverse hazard rate is increasing, then this steepening of the contract improves selection. Otherwise, selection is harmed whenever the reverse hazard rate decreases sufficiently fast that it offsets the larger increase in $u_h$.

## General Framework

The insights in the simple example above extend beyond the setting with linear contracts, risk-neutral utility, type-independent outside option distributions, and isoelastic costs. In

[4]We use the symbol "$=_s$" to indicate that the objects on either side have strictly the same sign.
[5]Notice that $u\rho(u)$ is decreasing if and only if $u\rho'(u)/\rho(u) < -1$.
[6]If $G$ were neither log-convex nor log-concave, then the sign of (3) could be either positive or negative.

this section, we generalize our results to find conditions under which a marginal change in a 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 $\ell$ by

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

where the direction of adjustment $\ell : X \to \mathbb{R}$ is the difference between an adjustment contract $\hat{w}$ and the status quo contract $w$.[7] Our main interest is how $q(w)$ varies as we change the status quo contract $w$ in direction $\ell$, i.e., whether (and when) $\mathcal{D}^{\ell}_{q(w)}$ is positive.

**Definition.** An adjustment of $w$ in direction $\ell$ *improves* (*harms*) selection if $\mathcal{D}^{\ell}_{q(w)} > (<)\ 0$.

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

$$\rho_h\big(u_h(w)\big) \cdot \mathcal{D}^{\ell}_{u_h(w)} > (<)\ \rho_l\big(u_l(w)\big) \cdot \mathcal{D}^{\ell}_{u_l(w)} \tag{4}$$

Lemma 1 allows us to find conditions over $G_h$ and $G_l$ under which steepening incentives improves or harms selection. When contracts are linear, the definition of "steepening incentives" corresponds to increasing the contract's slope. The following definition extends the notion of steepening incentives to arbitrary contracts.

**Definition.** A marginal change in direction $\ell$ *steepens incentives* if

$$\int v'\big(w(x)\big)\ell(x)\big[f(x|\hat{a}) - f(x|\tilde{a})\big]dx > 0 \quad \forall \hat{a} > \tilde{a};$$

[7]Equivalently, we could define the Gateau derivative from stemming from a convex combination between contracts $w$ and $\hat{w}$ with the weight on $\hat{w}$ being $\varepsilon$ and converging from above to zero. Note that as any $w$ and $\hat{w}$ are bounded and upper semicontinuous, so it is $\ell$.

i.e., if the greater the effort, the larger the monetary gain in utility.

Equivalently, a marginal change in direction $\ell$ steepens incentives if and only if it increases marginal incentives; i.e., $\int v'\big(w(x)\big)\ell(x)f_a(x|a)dx > 0$ for all $a$. For instance, in our example with linear contracts and risk-neutrality, steepening incentives is equivalent to increasing the slope $\alpha$.

**Theorem 1.** *Consider an arbitrary status quo contract $w$ and a steepening of incentives in direction $\ell$, where $\mathcal{D}^{\ell}_{u_l(w)} \cdot \mathcal{D}^{\ell}_{u_h(w)} > 0$. Then, there exists $G_h(\cdot) \succ_{FOSD} G_l(\cdot)$ for which a local adjustment in direction $\ell$ harms selection.*

Theorem 1 extends the insights of our example by showing that negative selection can arise under any steepening of incentives and for any status quo contract as long as the change affects both types' utilities in the same direction. 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., $\mathcal{D}^{\ell}_{u_h(w)} \geq \mathcal{D}^{\ell}_{u_l(w)}$), 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 1, 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 (for instance, a strict Pareto improvement), there exist outside option distributions under which such steepening harms selection.[8] 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 func-

[8]We say *strict Pareto improvement* to refer to changes that strictly increase the utility of both types.

tions.

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.,

$$\int v'\big(w(x)\big)\ell(x)f(x|a)dx > 0 \; \forall \; a \in [\underline{a}, \overline{a}]. \tag{5}$$

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 it.

**Theorem 2.** *Consider an arbitrary status quo contract $w$ and a steepening of incentives in direction $\ell$ that satisfies* (5)*. A local adjustment in direction $\ell$ improves selection for all effort cost functions if and only if $\rho_h(\tilde{u}) \geq \rho_l(\hat{u})$ for all $\hat{u}$ and $\tilde{u} \geq \hat{u}$.*

Theorem 2 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

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

which by Lemma 1 implies negative selection.

*Remark* 1. Theorems 1 and 2 require the local change in contracts to affect the utility of both types in the same direction. However, not all steepening of incentives do. Suppose one constructs a steepening of incentives for which the high type's utility increases while the low type's decreases. In that case, the firm attracts a larger mass of high and a smaller mass of low types, necessarily improving selection. One potential example of such a steepening is a rotation of payments, where the firm simultaneously increases the slope and reduces the baseline pay in the appropriate amount. The difficulty, however, might be finding the proper baseline pay reduction. If it is too small, then the rotation would benefit both types, while if it is too large, it would harm them both, and our results would apply. The appropriate reduction in the baseline pay would directly depend on each type's effort cost function, which might be challenging to observe.

## A sufficient statistic for improved selection

Theorem 2 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.

**Theorem 3.** *A local adjustment of $w$ in direction $\ell$ improves (harms) selection if and only if*

$$\frac{\mathcal{D}^{\ell}_{A(w)}}{A(w)} - \frac{\mathcal{D}^{\ell}_{R(w)}}{R(w)} > (<) 0. \tag{6}$$

Theorem 3 shows that selection improves (worsens) if and only if the mass of total applicants responds more (less) relative to its initial size than the mass of rejected applicants. Intuitively, if a change in the contract generates a larger response in the mass of total applicants than in the mass of rejected applicants, then it must have attracted more (fewer) high types than low types since high types are less likely to be rejected.

The terms $\mathcal{D}^{\ell}_{A(w)}/A$ and $\mathcal{D}^{\ell}_{R(w)}/R$ can be interpreted as semi-elasticities, where we measure how the mass of total applicants and the mass of rejected applicants vary relative to their initial size. Theorem 3 implies that these semi-elasticities suffice to answer whether a given local contract change improves or harms 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.

## Finding all directions that do not harm selection

Theorem 3 provides a test of whether an adjustment *in a given direction* $\ell$ 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 $A(w)$ and $R(w)$.

We consider data from one experiment (Experiment 1) where the firm marginally changes the status quo contract in direction $\ell$. Upon conducting such an experiment, the firm observes how the total and rejected masses of applicants change; that is, it observes $\mathcal{D}^{\ell}_{A(w)}$ and $\mathcal{D}^{\ell}_{R(w)}$. 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

$$\mathcal{D}^{\hat{\ell}}_{u_t(w)} = \int v'\big(w(x)\big)\hat{\ell}(x) f(x|a_t)dx \quad \forall t \in \{l, h\}. \tag{7}$$

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 4.** *Consider the data generated by Experiment 1, and suppose that $\mathcal{D}^{\ell}_{u_t(w)} \neq 0$ for all $t \in \{l, h\}$. Then, a local change in direction $\hat{\ell}$ does not harm selection if and only if*

$$\mathcal{D}^{\hat{\ell}}_{u_h(w)} \underbrace{\frac{r_l \mathcal{D}^{\ell}_{A(w)} - \mathcal{D}^{\ell}_{R(w)}}{(r_l A(w) - R(w))\mathcal{D}^{\ell}_{u_h(w)}}}_{:= K_h} \geq \mathcal{D}^{\hat{\ell}}_{u_l(w)} \underbrace{\frac{\mathcal{D}^{\ell}_{R(w)} - r_h \mathcal{D}^{\ell}_{A(w)}}{(R(w) - r_h A(w))\mathcal{D}^{\ell}_{u_l(w)}}}_{:= K_l}. \tag{8}$$

Note that $K_h$ and $K_l$ depend only on $\ell$ 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, (8) 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:[9]

$$\mathcal{D}^{\hat{\ell}}_{\pi(w)} = \underbrace{\int [x - w(x)][q(w) f_a(x|a_h)\mathcal{D}^{\hat{\ell}}_{a_h(w)} + (1-q(w)) f_a(x|a_l)\mathcal{D}^{\hat{\ell}}_{a_l(w)}]dx}_{\text{Incentive effect}}$$

$$+ \underbrace{\left[\int [x - w(x)][f(x|a_h) - f(x|a_l)]dx + \gamma'\big(q(w)\big)\right]\mathcal{D}^{\hat{\ell}}_{q(w)}}_{\text{Selection effect}}$$

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

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 $\mathcal{D}^{\hat{\ell}}_{\pi(w)}$ 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:

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

[9]We soon impose a sufficient condition for effort responses to be differentiatiable. We present the argument in this order for expositional purposes.

subject to

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

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

Recall that Theorem 4 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, allowing us to extrapolate how the marginal incentives change irrespective of the initial 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 $\ell$ changes the agent's effort by

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

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

$$\mathcal{D}^{\hat{\ell}}_{a_t(w)} = \underbrace{\frac{\mathcal{D}^{\ell}_{a_t(w)}}{\int \ell(x)v'(w(x))h_0(x)dx}}_{\text{Does not depend on } \hat{\ell}} \cdot \int \hat{\ell}(x)v'(w(x))h_0(x)dx. \tag{9}$$

Equation (9) 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 $\ell^+$, where $\ell^+$ 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 $\ell^+$ 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 $\ell^+$ while keeping $q$ fixed at the status quo. That is, upon running Experiment 2, the principal observes

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

and

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

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 $\mathcal{D}^{\ell^+}_{\bar{a}(w)}$ but also the function $h_0(x) = \mathcal{D}^{\ell^+}_{\bar{f}(w)}(x)/\mathcal{D}^{\ell^+}_{\bar{a}(w)}$. However, even with the data generated by Experimetn 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 5.** *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 $\hat{\ell}^*$ is proportional to $L(x, \lambda^*, \mu^*)$, where*

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

Theorem 5 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 in 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 $\lambda^*$ and $\mu^*$, 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*

$$\int [x - w(x)][f(x|a_h) - f(x|a_l)]dx + \gamma'\big(q(w)\big) \geq 0. \tag{10}$$

*Then $\mathcal{D}^{\hat{\ell}^*}_{\pi(w)} \geq 0$. 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 (10)). 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.

# A Ommitted Proofs

**Proof of Lemma 1.** The first step is to show that $\mathcal{D}^{\ell}_{u_t(w)}$ 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* $\{\partial\psi(a,\cdot)/\partial\varepsilon\}_{a\in[\underline{a},\overline{a}]}$ *is equidifferentiable at* $\varepsilon_0 \in [0,1]$ *and* $\underset{a\in[\underline{a},\overline{a}]}{sup} |\partial\psi(a,\varepsilon_0)/\partial\varepsilon| < +\infty$.

**Proof.** Note that as $v$ is continuous and $w$ and $\ell$ 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 2.** $\mathcal{D}^{\ell}_{u_t(w)}$ *exists and*

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

**Proof.** Claim 1 shows that the conditions of Theorem 3 in Milgrom and Segal (2002) are satisfied. Hence, the right-hand 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 $\varepsilon_0 = 0$

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

$\square$

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 $\ell$ delivers the result. $\square$

**Proof of Theorem 1.** Fix an arbitrary status quo contract $w$ and a steepening of incentives $\ell$ for which $\mathcal{D}^{\ell}_{u_l(w)} \cdot \mathcal{D}^{\ell}_{u_h(w)} > 0$. We then construct two distributions $G_h(\cdot) \succ_{FOSD} G_l(\cdot)$ such that adjusting $w$ in direction $\ell$ 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

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

By Lemma 1, a local adjustment in direction $\ell$ harms selection if and only if

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

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: $\mathcal{D}^{\ell}_{u_h(w)} > 0$ or $\mathcal{D}^{\ell}_{u_h(w)} < 0$.

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

Let $G(y|l) = exp\big(-[y+b]^{-\gamma}\big)$ and $G(y|h) = \big[G(y|l)\big]^{\delta}$, where $b > -u_l$, $\gamma > 0$ and $\delta > 1$. Note that as $\delta > 1$ and $G(y|l) < 1$ 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 $\gamma$ 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 $\gamma$ large enough

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

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

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 $G(y|h) < G(y|l)$. Hence, $G_h(\cdot)$ first-order stochastically dominates $G_l(\cdot)$. Moreover, $g_t/G_t = \lambda_t$. Therefore, for a sufficiently large $\lambda_h/\lambda_l$

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

$\square$

**Proof of Theorem 2.**

**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

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

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

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

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

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 $\ell$.

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

Let

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

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

where $\beta > 0$. Let $\beta_H > inf\{\beta \in \mathbb{R}_{++} : a^*(\beta) = \underline{a}\}$ and $\beta_L = \beta_H + \varepsilon$, where $\varepsilon > 0$. Then let

$$\gamma_t = \psi^*(\beta_t) - \beta_t \underline{a}^2 - \tilde{u}_t.$$

Note that for $\varepsilon$ sufficiently small, $\gamma_l > \gamma_h$. Finally, let $c_t(a) = \beta_t a^2 + \gamma_t$, which 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

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

Therefore,

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

Hence, by Lemma 1, an adjustment in direction $\ell$ harms selection. □

**Proof of Theorem 3.** Recall that

$$A(w) = pG_h(w) + (1-p)G_l(w)$$

$$R(w) = r_h p G_h(w) + r_l(1-p)G_l(w).$$

Hence,

$$G_l(w) = \frac{R(w) - r_h A(w)}{(1-p)(r_l - r_h)} \quad \text{and} \quad G_h(w) = \frac{r_l A(w) - R(w)}{p(r_l - r_h)}. \tag{11}$$

Also,

$$\mathcal{D}^{\ell}_{A(w)} = p g_h \mathcal{D}^{\ell}_{u_h(w)} + (1-p) g_l \mathcal{D}^{\ell}_{u_l(w)}$$

$$\mathcal{D}^{\ell}_{R(w)} = r_h p g_h \mathcal{D}^{\ell}_{u_h(w)} + r_l(1-p) g_l \mathcal{D}^{\ell}_{u_l(w)}.$$

Hence,

$$g_l \mathcal{D}^{\ell}_{u_l(w)} = \frac{\mathcal{D}^{\ell}_{R(w)} - r_h \mathcal{D}^{\ell}_{A(w)}}{(1-p)(r_l - r_h)} \quad \text{and} \quad g_h \mathcal{D}^{\ell}_{u_h(w)} = \frac{r_l \mathcal{D}^{\ell}_{A(w)} - \mathcal{D}^{\ell}_{R(w)}}{p(r_l - r_h)}. \tag{12}$$

Replacing (11) and (12) into (4), we get

$$
\begin{aligned}
\mathcal{D}^{\ell}_{q(w)} &=_s \frac{g_h}{G_h}\mathcal{D}^{\ell}_{u_h(w)} - \frac{g_l}{G_l}\mathcal{D}^{\ell}_{u_l(w)} \\
&= \frac{r_l\mathcal{D}^{\ell}_{A(w)} - \mathcal{D}^{\ell}_{R(w)}}{r_lA - R} - \frac{\mathcal{D}^{\ell}_{R(w)} - r_h\mathcal{D}^{\ell}_{A(w)}}{R - r_hA} \\
&= \frac{(r_l - r_h)AR}{(r_lA - R)(R - r_hA)}\left[\frac{\mathcal{D}^{\ell}_{A(w)}}{A(w)} - \frac{\mathcal{D}^{\ell}_{R(w)}}{R(w)}\right] \\
&=_s \frac{\mathcal{D}^{\ell}_{A(w)}}{A(w)} - \frac{\mathcal{D}^{\ell}_{R(w)}}{R(w)}.
\end{aligned}
$$

Where the last $=_s$ stems from $(r_lA - R)(R - r_hA) > 0$, which is a consequence of $G_l,\ G_h > 0$. □

**Proof of Theorem 4.** Recall that a marginal change in the contract in direction $\hat{\ell}$ improves selection if and only if

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

By equation (7), 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 (11) and (12), we have that

$$\frac{g_l}{G_l} = \frac{\mathcal{D}^{\ell}_{R(w)} - r_h\mathcal{D}^{\ell}_{A(w)}}{(R - r_hA)\mathcal{D}^{\ell}_{u_l(w)}}, \quad \text{and} \quad \frac{g_h}{G_h} = \frac{r_l\mathcal{D}^{\ell}_{A(w)} - \mathcal{D}^{\ell}_{R(w)}}{(r_lA - R)\mathcal{D}^{\ell}_{u_h(w)}},$$

which concludes the proof. □

**Proof of Theorem 5.** 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

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

From Experiment 1, the firm can reconstruct $K_h$ and $K_l$. From Experiment 2, the principal can recover $\mathcal{D}^{\ell+}_{\bar{a}(w)}$ 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

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

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 $\lambda$ and $\nu$ by solving the dual problem:

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

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

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

and

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

Thus, the optimal adjustment direction is

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

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 $\Pi^*$. 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 $\lambda^*$ and $\nu^*$ 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 ($PP$). □

**Proof of Corollary 1.** Note that ($NHS$) and (10) imply that

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

Hence,

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

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

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 $\pi$, 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 ($PP$). $\square$

# 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.

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.

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. 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.

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.