When Should a Principal Delegate to an Agent in Selection Processes?

Anonymous authors
Paper under double-blind review

Abstract

Decision-makers in high-stakes selection processes often face a fundamental choice: whether to make decisions themselves or to delegate authority to another entity whose incentives may only be partially aligned with their own. Such delegation arises naturally in settings like graduate admissions, hiring, or promotion, where a principal (e.g. a professor or manager) either reviews applicants personally and makes decisions or decisions are delegated to an agent (e.g. a committee or third-party or AI agent).

The principal has the expertise to conduct *holistic* evaluations of applicants (even accounting for factors like team fit), but incurs a cost for every application reviewed. In contrast, the agent can review a large volume of applications efficiently, greatly lowering the principal's costs. However the agent's evaluation is on the basis of a signal that is only *correlated* with the principal's metric but may be potentially misaligned, diminishing the expected quality of selected applicants. We study this fundamental trade-off in a stylized selection model with noisy signals. Our goal is to characterize when delegation is beneficial versus when decision-making should remain with the principal. We compare these regimes along three dimensions: (i) the principal's utility; (ii) the quality of the selected applicants according to the principal's metric; and (iii) the fairness of selection outcomes under disparate signal qualities.

1 Introduction

With the advent of automated decision making, software as a service, and agentic AI, there are many potential benefits to delegating to another decision maker. Third parties can offer AI and other tools that promise cost-efficient and accurate decision making at scale, such as for auto-bidding in online auctions (Aggarwal et al., 2024), self-driving cars (Forum, 2025), algorithmic pricing (Taylor, 2025), and resume-screening agents in hiring (NPR, 2025). Similar tools are also available now for facilitating efficient selection processes (where the goal is to choose from a pool of applicants), for example, in hiring and admissions.

Importantly, many of these decisions are made under time, cost, expertise, and informational constraints and inefficiencies, providing incentives for a decision maker to delegate to a third-party agent or AI tool. When the decision maker, whom we will call the *principal*, delegates a decision-making process to another, the *agent*, the agent can specialize in that process and take advantage of more data, larger scale, and higher efficiency, while the principal reaps the benefits of the decision. The agent may seek control over decision making anyway, for the sake of centralization & standardization, enforcing constraints like fairness considerations, or ensuring that the decisions benefit themselves (Brynjolfsson & Ng, 2023; Kapor et al., 2024; Lupia, 2015). However, delegation comes with its own set of challenges. The principal and the agent may not be aligned in terms of objectives or preferences (Alur et al., 2024; Candrian & Scherer, 2022; Lupia, 2015), leading to lower quality outcomes for the principal (Sliwka, 2001), a loss of agency (Lupia, 2015), and an opportunity for the agent to take advantage of their control over the process (Jensen & Meckling, 1976). Delegation to AI hiring tools, for example, may ignore cultural fit, reduce interaction between candidates and hiring managers, and fail to properly evaluate candidate's attributes (Aizenberg et al., 2025; Kim, 2025). Brokers who do not select suitable investments for their clients risk violating legal requirements (Anderson

& Winslow, 1992-1993). Delegation may also lead to issues when it comes to responsible and fair decision-making: for example, Amazon's AI hiring tool was found to overwhelmingly favor male applicants to female applicants purely because it had been trained on historical data of successful applicant resumes which were predominantly male (Dastin, 2018), failing to correct for biases in its own data.

This naturally motivates the following general question: ***when should a principal delegate decisions to an agent?*** To address this question, we study delegation specifically in the context of *selection processes*. Selection problems arise frequently in hiring and admissions and play a key role in organizational decision-making. We consider a model of a selection process where applicant quality can only be evaluated imperfectly using noisy signals. There is a principal and an agent that may use different metrics for evaluation, which are correlated (but may still be only partially aligned). The key consideration for the principal is: should they delegate the process to the agent and obtain applicants of potentially lower quality at very little to no cost; or should they retain decision-making authority to ensure better control over the quality of selected applicants but at high cost. We investigate the impact of each of these decisions (delegation versus non-delegation) on the outcomes of the selection process in terms of efficiency and fairness, which provides essential insights on when it might be beneficial for the principal to delegate.

Summary of Contributions. We summarize our main contributions as follows:

1. In Section 2, we introduce our model where a principal (they/them) must decide which applicants to select from a pool of applicants, but can delegate the decision to an agent (it). Our model applies to all types of agents, i.e., *strategic* (like hiring outsourced to a third-party firm) and *non-strategic* (like AI or LLM-based hiring tools).

2. We provide theoretical characterizations of the utility achieved *by the principal*, both when they make decisions themselves (Section 4) and when they delegate to the agent (Section 3). We show that directly comparing utilities across the two scenarios is sufficient for the principal to make a delegation decision.

3. We expand our analysis to selecting applicants from *multiple* populations. In particular, in our model, the populations are *intrinsically* identical; they differ *only* in the accuracy of the noisy signal observed about each population. We consider two models of such disparity: i) one where applicants in the disadvantaged groups have a *negatively biased* signal of their quality and ii) one where they have a *noisier* signal, compared to the advantaged group. We study the fairness implications of signal disparity in the context of demographic parity, under the the delegated model in Section 3 and the non-delegated model in Section 4.

4. Finally, in Section 5, we synthesize our results and provide insights on when delegation is beneficial, comparing the outcomes of both decisions across different metrics: i) quality of selected applicants, ii) the principal's utility (including both quality *and* selection costs), and iii) fairness defined as demographic parity. We make several interesting (and counter-intuitive) observations: there may be non-monotonicities in the principal's preferences for delegating vs not delegating in terms of how correlated the principal and the agent's metrics are. Further, what is preferable for fairness heavily depends on whether signal disparities come from i) systematic under-estimation of one population vs ii) disparate informativeness (e.g., variance) of signals across populations.

Related Work. Our work is closely related to an extensive body of research in economics on principal-agent problems (Ross (1973)). In traditional principal-agent problems, a principal and an agent usually have asymmetric information access and misaligned interests; the goal for the principal is then to design a mechanism or contract so as to incentivize the strategic agent to act in the principal's interest. Our work falls is similar in spirit to a special class of principal-agent problems called *strategic delegation* (Vickers, 1985; Bester & Krähmer, 2008; Alonso & Matouschek, 2007; Baik & Kim, 1997). In strategic delegation problems, the principal makes the decision of whether they should allow the agent (or delegate) to act on their behalf (for example, hiring a lawyer for litigation). The principal grants decision-making authority to the agent because the agent is better at the task, or enables the principal to lower costs. Beyond standard strategic delegation problems where an agent is typically strategic, our framework also encompasses settings

with a non-strategic agent: for example, an AI agent or automated decision-making tool aids the principal in decision making. Unlike previous work, our work considers delegation in the context of selection processes, which feature an agent not necessarily incentivized to use the same metric to evaluate candidates as the principal (e.g. the agent uses AI that is not trained for hiring candidates that are a "fit" with the particular firm). This means that the delegation decision depends not just on the effect of the agent's hidden efforts, but also on its choices of metric and selection threshold, which we focus on in this work. Perhaps closest to our work is Bester & Krähmer (2008) which deals with task/project selection in the context of delegation (which is analogous to threshold selection in our model). However, our model is more general (in terms of the space of thresholds and utility functions). We also consider fairness via the impact of selection processes on the candidates. We specify our model of strategic delegation of selection processes in Section 2. For a detailed review on other aspects of generalized principal-agent problems like contract design, information design, learning etc., we refer readers to (Dütting et al., 2024; Dughmi, 2017; Lin & Chen, 2024).

We focus on selection problems, with motivating examples arising in hiring and admissions. Designing efficient and high quality hiring processes has been the focus of long-standing research. Prominent directions include analysis of popular heuristics like hiring 'above the mean' or 'above the median' (Broder et al., 2010; Helmi & Panholzer, 2013), algorithmic hiring under uncertainty (Purohit et al., 2019; Li et al., 2025), screening under sequential or pipeline settings (Cohen et al., 2023; Epstein & Ma, 2024), to name a few. Our work contributes to this broad area, and specifically studies the problem of delegation in the context of hiring and selection problems where the principal has the option to outsource the selection process entirely to the agent.

With the recent proliferation of AI tools that are capable of sophisticated decision-making, outsourcing or *delegation* is becoming increasingly common. In turn, recent works have focused on whether decision tasks in high-stakes settings should be assigned to 'algorithmic delegates' that can act reliably without human oversight (Milewski & Lewis, 1997; Lubars & Tan, 2019), and more broadly on the design of synergistic systems with both human and AI collaborators (Bansal et al., 2019; 2021; Lai et al., 2022; De Toni et al., 2024; Donahue et al., 2024). Recent work has also looked at the task of selecting the most optimal delegate among many (Greenwood et al., 2025; Raghavan, 2024). Our focus is slightly different, mostly centering on delegation in *selection processes*.

Finally, our work has implications for responsible AI and fairness. Since the origins of the theory of statistical discrimination (Arrow, 1971; Phelps, 1972), researchers have contributed across multiple dimensions of fairness, including but not limited to the challenges of algorithmic fairness (Kleinberg et al., 2018; Green, 2022; Corbett-Davies et al., 2017; Kleinberg, 2018), fairness auditing (Kallus et al., 2022), fairness in ML (Hardt et al., 2016; Karimi-Haghighi & Castillo, 2021; Mary et al., 2019), fairness for better societal outcomes (Mouzannar et al., 2019; Pitoura et al., 2022; Zehlike et al., 2020; Bower et al., 2017) etc. Fairness has also been studied extensively in the context of ranking and selection problems (Zehlike et al., 2017; Cohen et al., 2019; Blum et al., 2022; Celis et al., 2016; 2024; Kleinberg & Raghavan, 2018; Emelianov et al., 2020; Garg et al., 2021). Most of these works either look at the problem of fair selection in the presence of different forms of bias or consider the fairness implications of various selection strategies in the context of single-stage or multi-stage/sequential selection. On the contrary, we look at a different problem: we explore whether organizational decisions like who is conducting the selection process (principal or agent) can also have major fairness implications.

2 Model

This section provides our model: the applicants, the principal, and the two different types of selection processes (one where the principal makes all decisions and the other where the principal delegates to the agent). We include an extension to a multi-group setting (with an "advantaged" and a"disadvantaged" group) to understand the fairness implications of selection processes with and without delegation. A key feature of our model is that applicants can only be evaluated using noisy or imperfect signals.

2.1 Applicant Characteristics

Applicants Each applicant is described by a set of attributes $(s, \tilde{s}, t, \tilde{t})$. The tuple (s, t) indicates the applicant's private type. While the agent wants to evaluate applicants on the basis of s alone, t is the preferred choice of evaluation metric for the principal. We assume that (s, t) are jointly Gaussian mean-zero random variables with variances σ_s^2 and σ_t^2 respectively and $\rho \geq 0$ is the degree of correlation between them. However, (s, t) is *unobservable*. Instead, $(\tilde{s}, \tilde{t})$ represents noisy signals about s and t that are observed by the agent and the principal respectively. We assume that $\tilde{s} = s + \epsilon_s$ and $\tilde{t} = t + \epsilon_t$ where ϵ_s and ϵ_t are mean-zero Gaussian noise terms ($\epsilon_s \sim \mathcal{N}(0, \sigma_{es}^2)$ and $\epsilon_t \sim \mathcal{N}(0, \sigma_{et}^2)$) and $\epsilon_s \perp s$, $\epsilon_t \perp t$ and $\epsilon_s \perp \epsilon_t$. This implies that $\tilde{s}$ and $\tilde{t}$ are also jointly correlated mean-zero Gaussian random variables with variances $\sigma_{\tilde{s}}^2$ and $\sigma_{\tilde{t}}^2$ respectively, where $\sigma_{\tilde{s}}^2 = \sigma_s^2 + \sigma_{es}^2$ and $\sigma_{\tilde{t}}^2 = \sigma_t^2 + \sigma_{et}^2$. Note that Gaussian assumptions are relatively standard when it comes to modeling population distributions. Please refer to Kannan et al. (2019); Garg et al. (2020); Liu & Garg (2021) for a few examples of such works.

We can now decompose t as a linear combination of s and a signal f representing the independent information about t that is not captured by s:

Claim 1. *Let (s, t) be jointly Gaussian random variables with $\mathbb{E}[s] = \mathbb{E}[t] = 0$ and $Var(s) = \sigma_s^2$, $Var(t) = \sigma_t^2$ and $Corr(s, t) = \rho \geq 0$. Then, there exists random variable $f \sim \mathcal{N}(0, \sigma_f^2)$ with $\sigma_f^2 = (1 - \rho^2)\sigma_t^2$ and $f \perp s$, such that t can be expressed as $t \triangleq \gamma_1 s + \gamma_2 f$, where $\gamma_1, \gamma_2 \in \mathbb{R}$.*

The proof may be found in Appendix A.1. This shows that no matter how the principal's metric t is (positively) correlated with s, it can always be decomposed as a linear combination of the agent's metric s and an independent metric f. The same argument extends to $\tilde{t}$ which can be decomposed as a linear combination of $\tilde{s}$ and some $\tilde{f}$ where $\tilde{f}$ is a mean-zero Gaussian random variable (which is expressable as the sum of f from earlier and an independent mean-zero noise term ϵ_f) and $\tilde{f} \perp \tilde{s}$.

Without loss of generality, the decomposition can also be taken to be convex[1]. Therefore, from now on, we write:

$$t \triangleq \alpha f + (1 - \alpha)s, \tag{1}$$

for some $\alpha \in [0, 1]$ (similarly, $\tilde{t} \triangleq \alpha\tilde{f} + (1 - \alpha)\tilde{s}$ with the same α where $\tilde{f} \sim \mathcal{N}(0, \sigma_{\tilde{f}}^2)$, $\tilde{s} \sim \mathcal{N}(0, \sigma_{\tilde{s}}^2)$ and $\sigma_{\tilde{t}}^2 = \alpha^2\sigma_{\tilde{f}}^2 + (1 - \alpha)^2\sigma_{\tilde{s}}^2$).). This alternate characterization is useful because it allows us to think about s and f as two independent attributes of an applicant that have standalone semantic importance. E.g., s may be a measure of an applicant's *skills*, as for a job, while f could be a measure of *fit* with the hirer[2], i.e., how well-matched the applicant is to the principal. This also grants the quantity α an interpretation as the relative importance the principal places on fit versus skill (besides its role as measuring how correlated s and t are). Therefore, t can be perceived as a measure of *overall applicant quality* from the principal's perspective or simply *quality* for brevity, with σ_t^2 now being interpreted as the extent of diversity in quality in the applicant pool. Note that we call s "skill" and f "fit" purely for expositional purposes, they can always be used to refer to other independent applicant attributes.

If f is not derived from Claim 1, but rather represents some other pre-defined metric, f and s may be correlated with each other. However, all our main results go through even in that case. For details, see Appendix D.

2.2 Types of Selection Processes

We consider two types of selection processes in our model: *one where the principal delegates to the agent* and *the other where the principal makes all decisions themselves*. For example, in the context of graduate admissions, the first type of selection involves all admission decisions being made by a central admissions committee without the direct involvement of the professor (therefore the professor does not incur any costs,

[1]Note that $\gamma_1 \geq 0$ and $\gamma_2 > 0$. Then we can re-normalize the signal t by $\gamma_1 + \gamma_2$ to obtain a re-normalized t as a convex combination of s and f.

[2]Team fit is often a very important aspect of selection, which is separate from how skilled the candidate is (Sekiguchi & Huber, 2011).

but their utility depends on the average quality of students hired by the committee). In the latter type of selection, the professor themselves are tasked with the responsibility of reviewing applications (incurring a cost for every application reviewed) and deciding which students to hire. We now elaborate on each type of selection process in the general setting.

Selection Process Delegated to Agent. In this setting, the agent makes decisions unilaterally on which applicants to select. The agent does not see the fit score $\tilde{f}$ and bases its evaluation entirely on the perceived preparedness of the applicant, $\tilde{s}$. The decision rule is straightforward: *admit applicants with* $\tilde{s} \geq \tau_1$, where τ_1 is a pre-determined threshold. In the context of graduate admissions, this could be, for example, a GPA or GRE score requirement for admissions.

The principal's overall *ex-ante* expected utility when k applicants are hired by the agent is then given by:

$$U_{dg}(\tau_1) = k \cdot \mathbb{E}\left[t \mid \tilde{s} \geq \tau_1\right]. \tag{2}$$

Note here that *ex-ante*, the principal only knows that the hired applicants would be above the threshold and may have no additional information about them. Importantly, when the selection process is delegated, the principal themselves incur no time or effort cost, but the downside is that some of these hired applicants may be poorly matched to the job thereby reducing the principal's overall expected utility.

Our framework allows us to model two types of agents:

- *Non-strategic agents:* this, for example, could be seen as a general-purpose tool or an external AI agent that the principal is using to make decisions in their stead. The AI agent, for example, an initial resume screening tool for hiring, makes decisions based on overall application quality (encoded by s), but may not be able to tailor to specific internal team preferences (encoded by f)[3]. This non-strategic agent may also be a centralized university admissions committee, who is only interested in holding students to a certain bar (here, τ_1) based on their standardized test scores (here, $\tilde{s}$).

- *Strategic agents:* in this case, the agent is a strategic entity that may have their own utility or cost function, given by $c(s, \tau_1)$ if they use metric s with threshold τ_1. In this case, the agent can choose their τ_1 to maximize their own utility. This could be seen, for example, as a company outsourcing their interviewing decisions to third-party recruiters. A recruiter's understanding of a qualified candidate (encoded by s) may not be fully aligned with a company's preferences (encoded by $\alpha f + (1-\alpha)s$, where f measures the misalignment in objectives as characterized above). As τ_1 increases, a recruiter must spend more time and effort weeding out candidates, also increasing their cost. In this situation, the principal may need to compensate the agent for the dis-utility they incur, as per Section 3.1.

Selection Process under No Delegation. In this setting, the principal is tasked with the responsibility of selecting applicants themselves and wants to hire at most k applicants in expectation. Reviewing each application directly reveals the noisy estimate $\tilde{t}$ of the applicant's quality, where $\tilde{t} = \alpha\tilde{f} + (1-\alpha)\tilde{s}$, but it also incurs a fixed cost of c_{rev}.

The principal has to strategically make two decisions: i) what endogenous selection threshold $\tilde{\tau}$ to use (an applicant gets hired only if their perceived quality exceeds this threshold), and ii) how many applications n_{rev} to review—note that the principal cannot deterministically control how many candidates they hire, so they reason in expectation. If they review n_{rev} applications, they end up selecting $n_{rev} \cdot \mathbb{P}[\tilde{t} \geq \tilde{\tau}]$ applicants in expectation and each hired applicant earns them an expected utility of $E[t \mid \tilde{t} \geq \tilde{\tau}]$. Therefore, the principal's overall expected utility from the selection process when they make decisions themselves is given by:

$$U_{ndg}(\tilde{\tau}, n_{rev}) = n_{rev} \cdot \mathbb{P}[\tilde{t} \geq \tilde{\tau}] \cdot \mathbb{E}[t \mid \tilde{t} \geq \tilde{\tau}] - n_{rev} \cdot c_{rev}, \tag{3}$$

[3]In particular, situations where individuals pass interviews based on their qualifications but end up not being matched to a team are relatively common in practice

where $n_{rev} \cdot \mathbb{P}[\tilde{t} \geq \tilde{\tau}] \leq k$. We consider the constraint to be "no more than k applicants in expectation" for model generality. We show later in Section 4 that this constraint is equivalent, in our setting, to the equality constraint $n_{rev} \cdot \mathbb{P}[\tilde{t} \geq \tilde{\tau}] = k$, under the condition that it is viable for the principal to review applications (so, exactly k candidates are hired in expectation). Before concluding this segment, we specify the following assumption that we make throughout:

Assumption 1. *Our model always operates in the regime where the applicant pool is sufficiently large. This ensures that even for high thresholds τ_1 and $\tilde{\tau}$, sufficiently many candidates can always be found above the threshold.*

We expect this assumption to hold true in many real-world settings like graduate admissions at top US universities or top industry companies which receive hundreds to thousands of applications every year for a handful of positions.

2.3 Fairness under Multiple Groups

In order to explore the effects of the choice of selection process on fairness, beyond the general setting with a single homogeneous group of applicants, we also consider a multi-group setting. In particular, we assume that there are two groups A and B[4] with demographic ratios $\Lambda_A = \lambda$ and $\Lambda_B = 1 - \lambda$ respectively for some $\lambda \in (0, 1)$, i.e. a uniformly randomly chosen applicant from the applicant pool belongs to group A with probability λ.

Assumption 2. *Both groups have the same distribution of true types (s, f).*

This assumption is based on the *we are all equal* (WAE) worldview introduced in the seminal work of Friedler et al. (2021). This enables us to study group-level disparate outcomes in the selection process without a difference in true types as a possible cause of those disparate outcomes.

In our setting, applicants from Group B are disadvantaged compared to those from Group A. The disadvantage is primarily with respect to how the noisy signal $\tilde{s}$ about applicant skill is perceived (the disadvantage can also manifest in $\tilde{f}$ but our main insights still go through unchanged) and can be of one of the two following forms:

- *Signal $\tilde{s}$ with biased mean.* The signal $\tilde{s}$ for group A applicants is drawn from $\mathcal{N}(0, \sigma_{\tilde{s}}^2)$ while for B applicants $\tilde{s}$ is drawn from $\mathcal{N}(-\beta, \sigma_{\tilde{s}}^2)$ for some $\beta > 0$. Consequently, the signal distribution for group A stochastically dominates the signal distribution for group B, i.e., given any threshold, the probability of finding an applicant from group A above the threshold is higher than the corresponding probability for an applicant from group B. Such disparities are frequently observed in practice and are very well-documented, for example, disparities in SAT scores between high-income and low-income students (Zwick, 2013) or gender gaps in math and verbal subject scores on standardized tests (Griselda, 2024).

- *Signal $\tilde{s}$ with disparate variance.* In this case, applicants from Group B are less well-understood than those from group A. We model this as receiving higher variance signals in population B, in line with previous works like Garg et al. (2020); Kannan et al. (2019). The lack of understanding of population B is modeled as having $\sigma_{\tilde{s},B} > \sigma_{\tilde{s},A}$. This may be, for example, because they have been historically marginalized in academia or the workplace, or because applicants are applying with backgrounds that decision-makers have little prior experience with.

We aim to investigate how the type of selection process affects the group-wise composition of selected applicants—in particular, if the choice of whether to delegate or not has outsized impacts on disparity between the two groups.

[4]Our results qualitatively extend beyond two groups, but for the sake of notational ease, our exposition assumes there are exactly two groups.

3 A Selection Process Delegated to the Agent

In this section, we consider the setting where the principal delegates the task of selecting applicants to the agent. We explore the effect that delegation has on principal utilities in the single group setting: we focus on understanding how said utility evolves in the parameters of the problem, including the selection threshold τ_1 and the importance of fit controlled by α. We then consider the multi-group setting, in particular observing how the composition of admitted applicants in a selection process under delegation looks like and how it changes with the selection threshold and the extent of advantage one group already has over the other.

3.1 An Unifying Framework for Strategic and Non-Strategic Agents

In this section, we show that the case of a strategic agent and non-strategic agent can be handled under a single unifying framework. In particular, given that the agent uses threshold τ_1 (which may be fixed or chosen strategically), we show that under classical contracts, the principal delegates decisions to the agent if and only if $U_{dg}(\tau_1) \geq U_{ndg}(\tilde{\tau}, n_{rev})$. This is exactly the same as the case of a non-strategic agent, where there is no monetary transfers between the two parties and $U_{dg}(\tau_1)$ and $U_{ndg}(\tilde{\tau}, n_{rev})$ are the actual principal utilities under delegation and no delegation, respectively.

The principal-agent contract in the strategic case In particular, we consider a principal and a strategic agent who form a contract. If the principal decides to delegate decisions to the agent, as seen in Section 2, the agent incurs a cost of $c(s, \tau_1)$ when using selection threshold τ_1. In this case, the principal must provide the agent with a performance-based payment $P(\tau_1, \tilde{\tau}, n_{rev})$ as is standard in the contracting literature (Sappington, 1991; Grossman & Hart, 1992). In particular, the agent gets utility[5]

$$P(\tau_1, \tilde{\tau}, n_{rev}) - c(s, \tau_1).$$

We consider the following class of linear contracts:

Definition 1 (Surplus-Sharing Contracts)**.** *Let $\eta \in [0, 1]$. A surplus-sharing contract at rate η is one where, if the principal decides to delegate decisions to the agent, they pay the agent a fraction η of their total surplus from delegation, i.e.*

$$P(\tau_1, \tilde{\tau}, n_{rev}) = \eta(U_{dg}(\tau_1) - U_{ndg}(\tilde{\tau}, n_{rev})).$$

If the principal does not delegate, then $P(\tau_1, \tilde{\tau}, n_{rev}) = 0$.

Note that this class of contracts is standard in the contracting and in the principal-agent literature. In particular, linear contracts are both simple and known to be optimal and robust in many settings (Yu & Kong, 2020; Rogerson, 1987; Carroll, 2015; Dütting et al., 2019)).

When does a principal prefer delegating to an agent? We now show how surplus-sharing contracts allow us to reduce our framework to the non-strategic case. In particular, we note that:

Claim 2. *The principal delegates to the strategic agent under a surplus-sharing contract at any rate $\eta \in [0, 1]$ if and only if $U_{dg}(\tau_1) \geq U_{ndg}(\tilde{\tau}, n_{rev})$.*

Proof. The principal delegates if and only if they gain increased utility from doing so, i.e. if and only if $(1 - \eta)U_{dg}(\tau_1) + \eta U_{ndg}(\tilde{\tau}, n_{rev}) \geq U_{ndg}(\tilde{\tau}, n_{rev})$, which immediately reduces to the desired condition. □

I.e., the principal's decision in the presence of a strategic agent is exactly identical to the case of a non-strategic agent: delegate if and only if their utility from delegation (not including payments) U_{dg} is higher than their utility from making their own decisions U_{ndg}. As such, in the rest of the paper, we reduce the principal's delegation decision to comparing $U_{dg}(\tau_1)$ and $U_{ndg}(\tilde{\tau}, n_{rev})$, whether the agent is strategic or not.

[5]We assume that the cost of the agent is such that there exists τ_1 for which the principal's surplus exceeds the agent's cost. Otherwise, we are in the trivial case where no contract is possible in the first place, as it is impossible for both the principal to obtain positive surplus and the agent to obtain positive utility.

Remark 1. *Note that in our setting, a strategic agent may use a threshold τ_1 imposed by the principal. It may also pick the* optimal *threshold that maximizes its own utility, i.e. $\tau_1^* = \arg\max P(\tau_1, \tilde{\tau}, n_{rev}) - c(s, \tau_1)$. Our goal in this work is not to characterize the optimal choice τ_1^* of τ_1 for the agent, which depends on its cost c, and the choice η^* of η for the principal, which may be selected via bargaining. We assume that these quantities can be anticipated by both parties and are set to be the equilibrium quantities, and treat them as exogenous. Rather, the goal of our work is to understand* whether *and* when *a principal prefers to defer to an agent, assuming the contract above. We provide results for all choices of τ_1, irrespective of how they were chosen.*

3.2 Single Group Setting: A Utilitarian View

Recall that the principal's expected *ex-ante* utility per hired applicant is given by $\mathbb{E}[t \mid \tilde{s} \geq \tau_1]$. Our first main result expresses this utility in closed form in terms of key problem parameters.

Lemma 1. *When the selection process is delegated to the agent, the expected ex-ante utility per hired applicant earned by a principal who puts weight $\alpha \in (0, 1)$ on applicant fit, is given by:*

$$\mathbb{E}[t \mid \tilde{s} \geq \tau_1] = \frac{(1-\alpha)\sigma_s^2}{\sigma_{\tilde{s}}} \cdot H\left(\frac{\tau_1}{\sigma_{\tilde{s}}}\right),$$

where τ_1 is the agent's selection threshold and $H(\cdot)$ is the hazard rate function of a standard normal random variable.

The proof of the lemma can be found in Appendix B.1. A few direct consequences of this lemma are as follows:

Corollary 1. *When the selection process is delegated to the agent, the principal's ex-ante utility from each admitted applicant is monotonically increasing in the agent's selection threshold τ_1.*

This follows from Lemma 1: the hazard rate function of the standard normal random variable is monotonically increasing in its argument (proof in Appendix E.2). The corollary shows that when the agent uses a higher selection threshold, the principal's utility increases as: i) the average ability of accepted applicants increases and ii) the principal does not incur any selection cost. We also note that the principal's utility is monotonic and decreasing in α:

Corollary 2. *When the selection process is delegated to the agent, the principal's ex-ante utility diminishes monotonically in α which is the principal's preferential weight on applicant fit.*

This follows from Lemma 1—increasing α decreases the principal's expected utility from each hired applicant since the principal's preferences increasingly diverge from the preferences of the agent they delegate to.

3.3 Fairness Implications in Multi-Group Settings

When the applicant pool consists of applicants from different groups, an immediate follow-up question is: what is the average group-wise composition of the set of applicants hired through a selection process which has been delegated to the agent? The answer to this question has important fairness implications. For example, if groups A and B are identical in all respects and group A has a demographic ratio of λ, one *fair outcome* might be that on average, λ fraction of the admitted applicants come from group A. This subscribes to the well-known concept of *demographic parity*[6] in the fairness literature (Dwork et al., 2012). In our setting, however, despite being intrinsically identical, the observed signals about the two groups are not identical—in particular, group B is disadvantaged in some way (either their signals are noisier (higher variance) or the distribution mean is negatively biased compared to Group A). Our goal in this section is to investigate how the extent of this disparity between the groups manifests in the outcome of the selection process.

[6]Note here that because our populations have identical distributions of true ability and fit, demographic parity is a natural notion of fairness.

Importantly, *the agent uses the same selection standard (τ_1) for everyone, irrespective of group identity*—our agent is group-blind[7]. In scenarios where the agent is strategic, choosing a single threshold offers the benefit of simplicity which often helps to lower costs. In other cases, this may also be enforced explicitly by the principal as part of the contract, either due to legal constraints or to enable the principal to verify if the terms of the contract are being upheld by the agent. A key difference of this with the setting where the principal does not delegate is that in the latter, the principal has the flexibility to customize group-specific selection standards—for example, in the context of graduate admissions, individual professors can still make their own decisions about who to hire.

Composition of selected applicants Let $\Phi_A(\cdot)$ and $\Phi_B(\cdot)$ indicate the normal CDFs of the signal ($\tilde{s}$) distributions for groups A and B respectively. Now, let $\Phi_M(\cdot)$ indicate the CDF of the Gaussian mixture distribution where a sample belongs to group A with probability λ and group B with probability $1-\lambda$:

$$\Phi_M(x) = \lambda \cdot \Phi_A(x) + (1-\lambda) \cdot \Phi_B(x), \quad \forall\, x \in \mathbb{R}.$$

Similarly, $\bar{\Phi}_A(\cdot)$, $\bar{\Phi}_B(\cdot)$ and $\bar{\Phi}_M(\cdot)$ indicate the corresponding complementary CDFs. We now present the following result (whose proof can be found in Appendix B.2):

Lemma 2. *Consider a selection process which is delegated to the agent with a selection threshold of τ_1. For any applicant hired through this process from a mixed applicant pool, the probability that said applicant belongs to group i is given by:*

$$\frac{\Lambda_i \cdot \bar{\Phi}_i(\tau_1)}{\bar{\Phi}_M(\tau_1)} \quad \forall\, i \in \{A, B\},$$

where $\Lambda_A = \lambda$ and $\Lambda_B = 1-\lambda$.

When do significant disparities arise? In order to understand how *fair* the realized composition of selected applicants is, we introduce the following fairness metric:

$$\mathcal{D} = \frac{Y_A}{\Lambda_A} - \frac{Y_B}{\Lambda_B}, \tag{4}$$

where Y_i is the realized proportion of admits that belong to group i (random variable) with $\mathbb{E}[Y_i] = \frac{\Lambda_i \cdot \bar{\Phi}_i(\tau_1)}{\bar{\Phi}_M(\tau_1)}$ (as shown in Lemma 2) and Λ_i is the demographic ratio of group i as before. The ratio Y_i/Λ_i indicates whether group i is over- or under-represented in the pool of hired applicants compared to what is *demographically fair*. Thus, the metric $\mathcal{D}$ (which looks at the difference of these ratios) is a measure of the extent and the direction of unfairness in the composition of hires across the two groups. Note that when groups A and B have identical signal distributions, $\bar{\Phi}_A(\tau_1) = \bar{\Phi}_B(\tau_1) = \bar{\Phi}_M(\tau_1)$ which implies that $\mathbb{E}[\mathcal{D}] = 0$—this corresponds to the case where we have a *perfectly fair* group composition in expectation. A larger absolute value of $\mathbb{E}[\mathcal{D}]$ indicates more unfairness, with positive and negative values indicating unfairness in favor and against the advantaged group (A) respectively. Our next result characterizes some of the statistical properties of the fairness metric $\mathcal{D}$:

Theorem 1. *Consider a selection process which has been delegated to the agent and uses a selection threshold of τ_1 to hire applicants from a mixed applicant pool consisting of groups A and B. In that case, the group fairness metric $\mathcal{D}$ satisfies:*

$$\mathbb{E}[\mathcal{D}] = \frac{\bar{\Phi}_A(\tau_1) - \bar{\Phi}_B(\tau_1)}{\bar{\Phi}_M(\tau_1)}, \quad \textit{and} \quad \frac{\bar{\Phi}_A(\tau_1) + \bar{\Phi}_B(\tau_1)}{\bar{\Phi}_M(\tau_1)} \geq \mathbb{E}[|\mathcal{D}|] \geq \frac{|\bar{\Phi}_A(\tau_1) - \bar{\Phi}_B(\tau_1)|}{\bar{\Phi}_M(\tau_1)}.$$

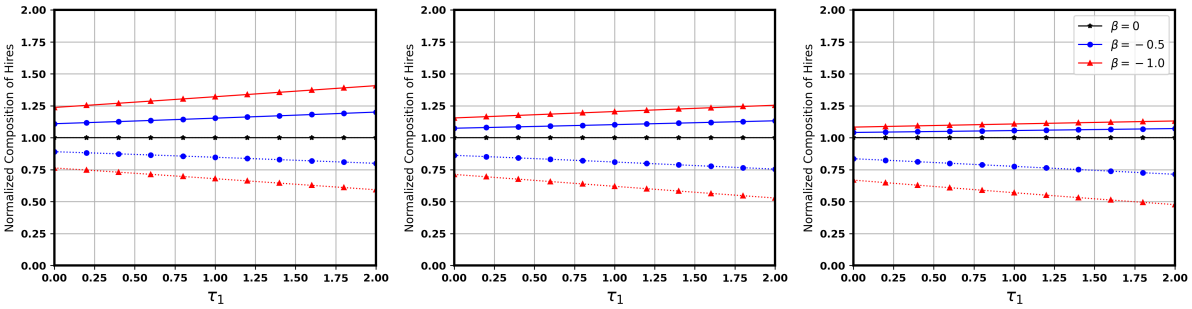
The proof of the theorem can be found in Appendix B.3. Firstly observe that when Group B is disadvantaged, $\bar{\Phi}_A(\tau_1) \neq \bar{\Phi}_B(\tau_1)$ which implies that the lower bound on $\mathbb{E}[|\mathcal{D}|]$ is non-trivial. This implies that if the signal

[7]For example, an automated hiring tool that relies on ostensibly neutral keywords may apply a uniform decision rule across applicants, while failing to account for the fact that such keywords can be disparately associated with different demographic groups; see for example Amazon's recent failure in using automated AI decision-making tools for hiring (Dastin, 2018). Or a central university admission committee that makes admissions decisions without taking into account the disparate meaning of standardized scores across different populations.

distributions are relatively different and one group has a larger tail than the other (for example, due to large additive bias), then significant disparities will arise between the groups on average. The expression for $\mathbb{E}[\mathcal{D}]$ also gives us insights about the direction of the expected *unfairness*, whether it is in favor of group A or group B. In particular:

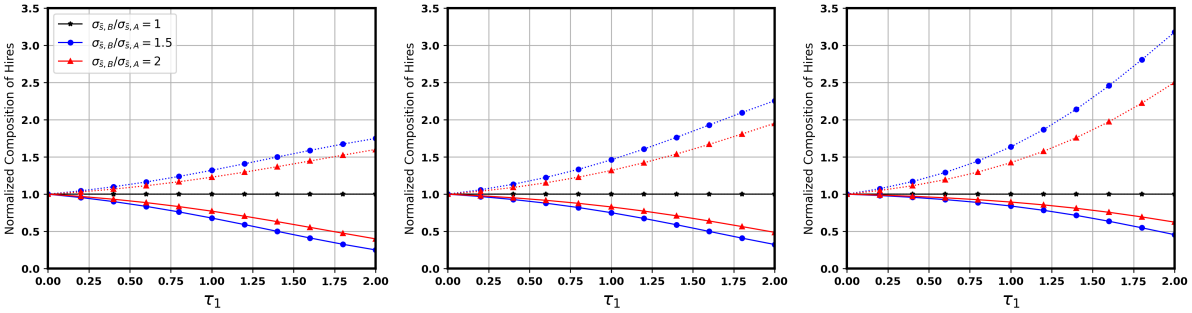
Corollary 3. *Suppose,* $\tilde{s}_A \sim \mathcal{N}\left(0, \sigma^2_{\tilde{s},A}\right)$ *and* $\tilde{s}_B \sim \mathcal{N}\left(-\beta, \sigma^2_{\tilde{s},B}\right)$ *with* $\frac{\sigma_{\tilde{s},B}}{\sigma_{\tilde{s},A}} = r$. *Then,*

$$\mathbb{E}[\mathcal{D}] > 0 \quad \textit{iff } (r-1)\tau_1 - \beta < 0, \quad \textit{and} \quad \mathbb{E}[\mathcal{D}] < 0 \quad \textit{iff } (r-1)\tau_1 - \beta > 0.$$



(a) $\Lambda_A = 0.50$ (b) $\Lambda_A = 0.65$ (c) $\Lambda_A = 0.80$

Figure 1: We plot the expected fraction of hires (normalized by demographic ratios) from group A (solid lines) and group B (dashed lines) respectively as a function of the agent's selection threshold τ_1 for different levels of bias β on the mean of group B's observed score distribution ($\tilde{s}$) and different levels of population skew (Λ_A). As τ_1 increases, the gap between the groups grows uniformly indicating that the disadvantaged group suffers as the selection process becomes more selective (from blue outwards to red). The same trend is observed in the magnitude of bias (β). However, as the prevalence of the majority group in the population (Λ_A) increases, the leading group has little scope for gaining additional advantage, so the extent of disparities actually diminishes.



(a) $\Lambda_A = 0.50$ (b) $\Lambda_A = 0.65$ (c) $\Lambda_A = 0.80$

Figure 2: We plot the expected fraction of hires (normalized by demographic ratios) from group A (solid lines) and group B (dashed lines) respectively as a function of the agent's selection threshold τ_1 (> 0) for different levels of the ratio $r = \sigma_{\tilde{s},B}/\sigma_{\tilde{s},A}$ and different levels of population skew (Λ_A). The ratio r captures how much more noisy group B's signals are with respect to group A. In this case, group B becomes increasingly more favored for selection and unfairness grows in favor of group B as r and τ_1 increase.

Now let us see what happens in the two special models of disadvantage we introduced earlier:

Case I: Negatively biased signal mean ($\beta > 0$, $r = 1$). In this case, we see that $\mathbb{E}[\mathcal{D}] > 0$, i.e., *the unfairness is in favor of group A*. The negatively biased signal mean for group B implies that $\Phi_A(\cdot)$ stochastically dominates $\Phi_B(\cdot)$, so given any threshold τ_1, the probability of any selected applicant to belong to group A is strictly larger. Figure 1, where we plot $\mathbb{E}[Y_A/\Lambda_A]$ and $\mathbb{E}[Y_B/\Lambda_B]$ from Equation equation 4 as a function of τ_1 for different levels of bias β and different levels of population skew Λ_A, shows this conclusively.

Case II: Higher signal variance ($\beta = 0$, $r > 1$). In this case, the direction of unfairness depends on the sign of τ_1. If $\tau_1 > 0$, $\mathbb{E}[\mathcal{D}] < 0$, i.e., *unfairness actually grows in favor of group B*. Due to the

higher variance, the distribution of $\tilde{s}_B$ has fatter tails, which means that the likelihood of a randomly picked applicant, who passes the threshold, to be from group B increases. This can be seen in Figure 2 (the dashed lines representing group B's normalized hire composition always lie above 1, indicating that they are hired at rates much higher than what is demographically fair, at the expense of group A)[8]. Over-selection from the group with larger signal variance under group-blind selection rules has also been observed in previous work, such as Emelianov et al. (2022).

The above discussion highlights that when the principal delegates to agent and the agent uses a fixed threshold-based selection policy, the selection outcomes can be *unfair*. However, which group benefits from the unfairness is more nuanced and depends on how their signal distributions are related to each other. When it comes to Case II above, it is important to highlight that the disadvantaged group getting favored for selection in expectation is not necessarily a 'good outcome'. In fact, we can show that the expected quality of a selected applicant decays monotonically in the level of noise in signal $\tilde{s}$ (as per Figure 5 in Appendix F), i.e., applicants from group B who get selected are of strictly lower expected quality compared to their group A counterparts. This is because the excessive noise in the signal leads to many 'bad' selections from group B. This, in turn, can actually reinforce negative stereotypes about group B in future iterations of the selection process, only harming them in the long term; this also creates unfairness *within* group B, where less qualified candidates routinely get selected over more qualified ones. In the real world, both types of biases are often present simultaneously (Picault, 2017; Emelianov et al., 2022; Phelps, 1972) and the nature and extent of unfairness depends on which type dominates. The benefit of a structured model framework like ours is that it allows us to isolate the causal effects of each bias type.

4 A Selection Process without Delegation

In this section, we consider a selection process where the principal makes selection decisions themselves without delegating to the agent. This is different from the previous setting in the sense that the principal can now decide which applicants to hire albeit at a high cost. In the first part of the section, we characterize how this trade-off plays out and what decisions the principal makes in order to maximize their net utility. Interestingly and surprisingly, we demonstrate how the principal's utility in this setting can actually be non-monotonic in key problem parameters, unlike the setting with delegation. Finally, we conclude the section with an analysis of the multi-group setting, showing that when the principal has the flexibility to set the selection criteria for each group and decide who to hire, the hiring outcomes are completely different from the setting with delegation, leading to significant implications for fairness.

4.1 Principal's Optimal Decisions

Recall that the principal has to choose i) how many applications to review n_{rev} and ii) what threshold $\tilde{\tau}$ on perceived quality ($\tilde{t}$) to use to hire applicants in a way that maximizes their overall utility. The overall utility is given by:

$$U_{ndg}(\tilde{\tau}, n_{rev}) = n_{rev} \cdot \mathbb{P}[\tilde{t} \geq \tilde{\tau}] \cdot \mathbb{E}[t \mid \tilde{t} \geq \tilde{\tau}] - n_{rev} \cdot c_{rev},$$

where c_{rev} is the cost of reviewing each additional application. Therefore, the principal's optimization is as follows:

$$\max_{n_{rev} \geq 0, \tilde{\tau}} \quad U_{ndg}(\tilde{\tau}, n_{rev}) \quad \text{s.t.} \quad n_{rev} \cdot \mathbb{P}[\tilde{t} \geq \tilde{\tau}] \leq k. \tag{5}$$

Our first result highlights that the solution to optimization program equation 5 has an important dependence on the cost c_{rev}. Further, it shows that the optimal decisions n^*_{rev} and $\tilde{\tau}^*$ can be decoupled and computed efficiently.

Theorem 2. *The optimal threshold $\tilde{\tau}^*$ can be obtained as:*

$$\tilde{\tau}^* = \arg\max_{\tilde{\tau}} \quad v(\tilde{\tau}) \quad \textit{where } v(\tilde{\tau}) = \frac{\sigma_{\tilde{t}}^2}{\sigma_{\tilde{t}}} \cdot \frac{\phi\left(\tilde{\tau}/\sigma_{\tilde{t}}\right)}{\Phi^c\left(\tilde{\tau}/\sigma_{\tilde{t}}\right)} - \frac{c_{rev}}{\Phi^c\left(\tilde{\tau}/\sigma_{\tilde{t}}\right)}, \tag{6}$$

[8]The sign of $\mathbb{E}[\mathcal{D}]$ does flip when $\tau_1 < 0$, but $\tau_1 < 0$ is generally not a useful case.

where $\phi(\cdot)$ and $\Phi^c(\cdot)$ denote the PDF and complementary CDF of the standard normal random variable. $\tilde{\tau}^*$ is unique and can be computed efficiently. Once we solve for $\tilde{\tau}^*$, the optimal n^*_{rev} can be obtained as follows:

- *If $\frac{\sigma_t^2}{\sigma_{\tilde{t}}} \cdot \frac{1}{\sqrt{2\pi}} > c_{rev}$, then $n^*_{rev} = \frac{k}{\mathbb{P}[\tilde{t} \geq \tilde{\tau}^*]}$;*
- *Else, $n^*_{rev} = 0$.*

The proof for the theorem can be found in Appendix C.1. A key insight is that the principal's optimal overall utility depends on the cost c_{rev}. If the cost is too high, from the principal's point of view there exists no choice of $(n^*_{rev}, \tilde{\tau}^*)$ that can provide strictly positive utility. In this case, the only reasonable option is to choose $n^*_{rev} = 0$. In particular, we have:

Corollary 4. *A selection process without delegation is viable*[9] *for the principal if and only if $\frac{\sigma_t^2}{\sigma_{\tilde{t}}} \cdot \frac{1}{\sqrt{2\pi}} > c_{rev}$.*

For large costs, the principal earns zero (trivial) utility, making it unviable to conduct selections on their own.

4.2 Single Group Setting: A Utilitarian View

Using the characterization of the optimal decisions in the previous section, we can express the overall expected utility earned by the principal in the selection process without delegation. Note that we are operating in the regime where the cost c_{rev} is small enough ($< \frac{\sigma_t^2}{\sigma_{\tilde{t}}} \cdot \frac{1}{\sqrt{2\pi}}$) that such a selection process in viable in the first place.

Lemma 3. *Suppose that $c_{rev} < \frac{\sigma_t^2}{\sigma_{\tilde{t}}} \cdot \frac{1}{\sqrt{2\pi}}$ and $\tilde{\tau}^*$ is the optimal selection threshold used by the principal. Then the expected utility earned by the principal per hired applicant is equal to $\frac{\sigma_t^2}{\sigma_{\tilde{t}}^2} \cdot \tilde{\tau}^*$.*

The proof can be found in Appendix C.3. Our main goal here is to understand how the optimal overall expected utility depends on problem parameters. We are particularly interested in the dependence on c_{rev} and α. While c_{rev} clearly affects the average quality of hires, it also affects the overall cost. So it is not apriori clear how the net utility might be affected. On the other hand, α captures the degree to which the principal prioritizes applicant fit when measuring quality—as such it is an important intrinsic preference parameter for the principal and needs to be studied. We present two results below capturing these dependencies:

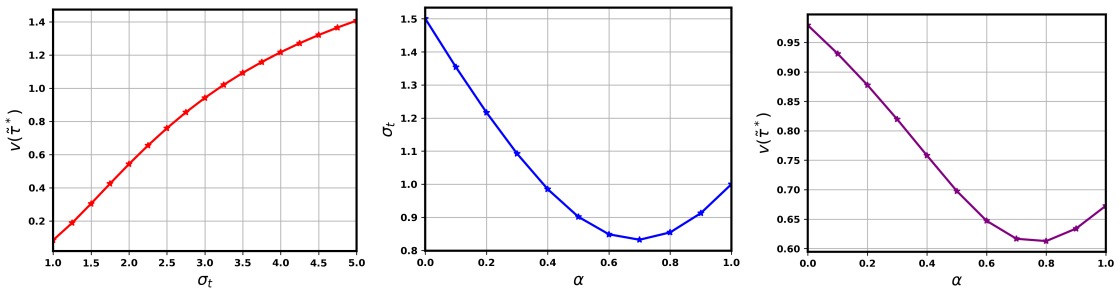
Claim 3. *In a selection process without delegation which is viable, the optimal overall expected utility for the principal is monotonically decreasing in the cost c_{rev} incurred in reviewing each application.*

The proof for this claim follows from Lemma 3 once we make the observation that a higher c_{rev} lowers the optimal selection threshold $\tilde{\tau}^*$ (Claim 10 in Appendix C.4). The intuition is the following: higher reviewing cost means that the principal wants to fill all available positions by reviewing only a small number of applications—this necessitates lowering of the selection standard.

Claim 4. *In a selection process without delegation which is viable, the optimal overall expected utility for the principal is **non-monotonic** in the parameter α that measures their weight on applicant fit.*

Figure 3 shows that while the optimal overall expected utility for the principal is monotonically increasing in σ_t (a more diverse applicant pool offers better utility on average, follows from Claim 11 in Appendix C.4), σ_t itself is non-monotonic in α. In particular, σ_t is U-shaped as α increases from 0 to 1 (since $\sigma_t = \sqrt{\alpha^2\sigma_f^2 + (1-\alpha)^2\sigma_s^2}$ which is convex in α and attains its minimum somewhere in $(0, 1)$). This leads to the overall optimal expected utility to be non-monotonic in α. Basically, the intuition is that at intermediate levels of α (when the principal puts approximately equal weights on both applicant fit and ability), finding candidates who are good enough on both metrics becomes harder leading to decreased expected utilities for the principal. Note that this trend is in sharp contrast to the setting with delegation where we showed that principal utilities are monotonically decreasing in α.

[9] There are other equivalent versions of Corollary 4 that can characterize the viability condition for selection processes without delegation in terms of $\tilde{\tau}^*$. For example, we can show that $\frac{\sigma_t^2}{\sigma_{\tilde{t}}} \cdot \frac{1}{\sqrt{2\pi}} > c_{rev} \iff n^*_{rev} > 0 \iff \tilde{\tau}^* > 0$. For details, refer to Appendix C.2.



(a) $v(\tilde{\tau}^*)$ vs σ_t (b) σ_t vs α (c) $v(\tilde{\tau}^*)$ vs α

Figure 3: The principal's net expected utility per hired applicant $v(\tilde{\tau}^*)$ is non-monotonic in α when the principal does not delegate. Parameter combinations for sub-figures: (a) $\sigma_e = 2$, $c_{rev} = 0.1$. (b) $\sigma_s = 1.5$, $\sigma_f = 1$. (c) $\sigma_s = 1.5$, $\sigma_f = 1$, $\sigma_{ef} = \sigma_{es} = 0.5$, $c_{rev} = 0.1$.

4.3 Multi-Group Setting: A Fairness Approach

Finally, we consider a setting where the applicant pool is mixed and consists of applicants from groups A and B. Recall that groups A and B are inherently identical, but Group B is disadvantaged in the sense that either their signals are noisier (with higher variance, i.e., $\sigma_{\tilde{s},A} < \sigma_{\tilde{s},B}$) or the distribution mean is negatively biased with respect to Group A (bias of $-\beta$). Note that a fundamental difference of this setting with the previous one (with delegation) is that *the principal has complete freedom to set individually optimized selection thresholds for each group in order to maximize utility* and does not suffer from the "naivety" of the agent.

We are again interested in the group-wise composition of hires in order to understand the fairness implications of not delegating. To that end, we first show that the disadvantaged group will end up providing strictly worse expected utility per hired applicant for the principal if the disadvantage manifests in the form of higher signal variance. However, if the bias is additive and *fully known* to the principal, it can be accounted for and a principal can simultaneously maximize utility and satisfy demographic parity.

Lemma 4. *Consider a selection process where the principal does not delegate and which is viable. Further, suppose that the principal is allowed to optimize selection thresholds $\tilde{\tau}_i^*$ individually for each group $i \in \{A, B\}$, with $v_i(\tilde{\tau}_i^*) = \mathbb{E}\left[t_i \mid \tilde{t}_i \geq \tilde{\tau}_i^*\right] - \frac{c_{rev}}{\mathbb{P}\left[\tilde{t}_i \geq \tilde{\tau}_i^*\right]}$ representing the principal's optimal net expected utility for every hired applicant from group i. In this case, for $\alpha \in (0, 1)$, we have that*

$$\sigma_{\tilde{s},A} = \sigma_{\tilde{s},B} \text{ and } \beta > 0 \implies v_A(\tilde{\tau}_A^*) = v_B(\tilde{\tau}_B^*), \qquad \text{(Biased mean model)}$$

$$\sigma_{\tilde{s},A} < \sigma_{\tilde{s},B} \text{ and } \beta \geq 0 \implies v_A(\tilde{\tau}_A^*) > v_B(\tilde{\tau}_B^*). \qquad \text{(Disparate variance model)}$$

Further, in the first case (with negatively biased signal mean for group B, but no variance disparity), group B's optimal threshold $\tilde{\tau}_B^*$ satisfies $\tilde{\tau}_B^* = \tilde{\tau}_A^* - (1 - \alpha)\beta$.

The proof for the above lemma can be found in Appendix C.5. This result shows that selecting from the group, which is not well-understood, is less lucrative from the principal's point of view —a high degree of noise leads to inaccurate evaluations and hence poor selection decisions. However, any negative additive biases in the signal mean can be handled as long as the magnitude of the bias can be learnt. We now see what consequences this has on fairness.

Consider the modified optimization problem for the principal when the applicant pool is mixed. Here, the principal is not restricted to selecting a certain number of applicants from either population. Instead, they decide how to simultaneously hire from both populations to fill *shared capacity* k.

$$\max_{n_{rev}(A), n_{rev}(B), \tilde{\tau}_A, \tilde{\tau}_B} \quad U_{ndg}^{(A)}(\tilde{\tau}_A, n_{rev}(A)) + U_{ndg}^{(B)}(\tilde{\tau}_B, n_{rev}(B)) \quad \text{s.t.}$$

$$n_{rev}(A) \cdot \mathbb{P}[\tilde{t}_A \geq \tilde{\tau}_A] + n_{rev}(B) \cdot \mathbb{P}[\tilde{t}_B \geq \tilde{\tau}_B] \leq k, \quad n_{rev}(A), n_{rev}(B) \geq 0. \qquad (7)$$

In light of Lemma 4, we expect the principal to hire more applicants from population A than from population B when group B's signals are noisier (bias only exists in the variance). However, our main result here shows that the reality is even worse, providing dire news for fairness in selection processes without delegation:

Theorem 3. *Suppose that group B is disadvantaged because of a noisier signal $\tilde{s}$ (biased signal variance, but no bias in the signal mean) and the cost c_{rev} is low enough that the selection process without delegation is viable. Then, at the optimal solution to Problem 7, $n_{rev}(B)^* = 0$, i.e., no applicant from Group B is hired.*

The proof for the theorem can be found in Appendix C.6. Returning to our graduate admissions example, PhD applicants from foreign or lesser-known educational backgrounds often face a disadvantage because their transcripts/GPA/letters may not be interpretable in the same way as their peers from well-known institutions. When faced with such noisy signals, faculty often choose to err on the side of caution, hiring students overwhelmingly from backgrounds they are familiar with. Our model and results in fact highlight effects observed empirically. One such example is faculty hiring in the U.S. where a small number of elite universities supply the overwhelming majority of faculty hires (Wapman et al., 2022; Nowogrodzki, 2022). In particular, the seminal work of Hu & Chen (2018) shows that this can be in part explained by reputational effects where evaluators have worse evaluations or lower confidence in applicants from certain populations or institutions, aligning with the insights of our own work. This also follows a long line of work in the dynamics of collective reputation which itself draws from theories of statistical discrimination (Arrow, 1971; Coate & Loury, 1993; Tirole, 1996; Levin et al., 2009).

The outcome, however, is completely different for the setting where a negative bias exists in group B's signal mean for $\tilde{s}$. Since additive biases can be corrected for, no group disparities arise:

Theorem 4. *Suppose that group B is disadvantaged because of a negatively biased mean for signal $\tilde{s}$ (but no bias in signal variance) and the cost c_{rev} is low enough that the selection process without delegation is viable. Then, there exists an optimal solution to Problem 7 where the group-wise composition of hires satisfies demographic parity.*

This again follows directly from Lemma 4. When the bias β is known, the principal can correct for the bias by augmenting the signal values $\tilde{s}$ for all group B applicants by amount β. As a result, the distribution of quality for both groups become virtually indistinguishable, thereby leading to the above outcome.

5 To Delegate or Not: Efficiency & Fairness Implications

We conclude this paper by conducting a comparative analysis between the two selection processes discussed so far and investigating when it can be beneficial for a principal to delegate to an agent. We focus on three primary dimensions of comparison: i) the utility of the principal, ii) the average quality of selected applicants from the principal's perspective, and iii) the fairness of hiring outcomes. While the first two pertain to the quality and efficiency of the hiring process, ensuring that hiring processes are *fair* is also of paramount importance. This is particularly crucial in light of many real-world instances of biased or discriminatory hiring practices (Dastin, 2018; Kline et al., 2022).

Principal Utilities & Expected Quality of Selected Applicants. First, we explore conditions under which it is beneficial for the principal to delegate to the agent, if the goal is to obtain i) higher expected quality for selected applicants; or ii) higher expected utilities per selected applicant. Figure 4 shows the difference in net principal utilities per selected applicant ($\Delta_{utility}$) and the difference in expected quality of a selected applicant ($\Delta_{quality}$) between the settings with and without delegation, as a function of key problem parameters. Formally, we define:

$$\Delta_{quality} = \mathbb{E}\left[t \mid \tilde{s} \geq \tau_1\right] - \mathbb{E}\left[t \mid \tilde{t} \geq \tilde{\tau}^*\right], \quad \text{and}$$

$$\Delta_{utility} = \left(\mathbb{E}\left[t \mid \tilde{s} \geq \tau_1\right]\right) - \left(\mathbb{E}\left[t \mid \tilde{t} \geq \tilde{\tau}^*\right] - \frac{c_{rev}}{\mathbb{P}\left[\tilde{t} \geq \tilde{\tau}^*\right]}\right).$$

A positive sign for $\Delta_{quality}$ and $\Delta_{utility}$ identifies preferences towards delegation. We make the following observations:

(1) *Dependence on c_{rev}.* For all else fixed, increasing c_{rev} decreases the principal's optimal threshold $\tilde{\tau}^*$ (they cannot afford to review too many applications). This implies that for every problem instance, there exists a threshold value of cost beyond which it is always better to delegate (Figure 10 in Appendix F).

(2) *Dependence on τ_1.* At any fixed $\alpha \in (0, 1)$, the expected quality of applicants selected by the agent increases monotonically in the selection threshold τ_1, as shown in Corollary 1; at the same time, the choice of τ_1 does not affect the case where the principal does not delegate. This implies that $\Delta_{quality}$ and $\Delta_{utility}$ are both *monotonically increasing* in τ_1. Therefore, for any given α, there exists a $\tau_1(\alpha)$ such that delegation is strictly better for the principal in terms of both expected quality of hire and net expected utility per hire for all $\tau_1 \geq \tau_1(\alpha)$.

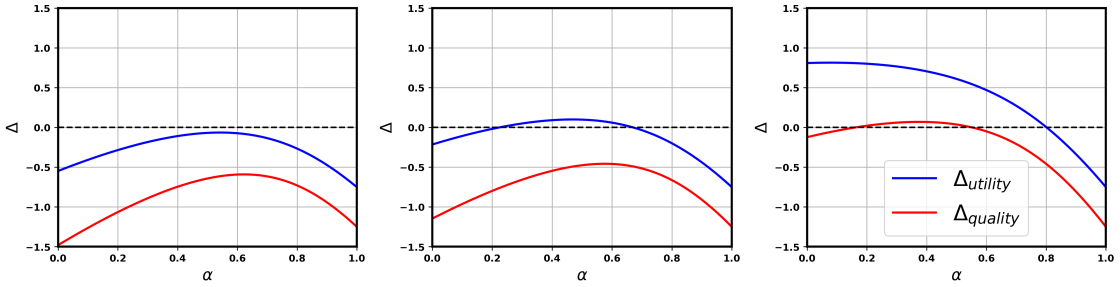
(3) *Dependence on α.* At fixed τ_1, whether it is beneficial to delegate or not can be **non-monotonic** in α. This can be seen in Figure 4—note that $\Delta_{utility}$ and $\Delta_{quality}$ are both inverted U-shaped and there exists intermediate regimes of α where delegation would be beneficial (see subfigures b) and c)). Intuitively, at intermediate values of α, finding a good candidate may be costly for the principal (the candidate needs to be good along both dimensions of ability and fit and hence more applications need to be reviewed), and delegation to the agent can be beneficial. The agent can compensate for applicants doing poorly on fit by ensuring that they do exceptionally well on ability. Actually, α influences the outcome through two distinct pathways: by affecting the principal's variance σ_t and also the correlation ρ between s and t. We isolate and study these individual effects in Appendix F (Figures 7, 8, 9).

However, given any τ_1, there always exists some $\alpha(\tau_1) \in (0, 1)$ such that for $\alpha \geq \alpha(\tau_1)$, delegating is never beneficial. Intuitively, once the principal and agent preferences become sufficiently misaligned, it is always better for the principal to retain control over decision-making. As τ_1 becomes smaller, this $\alpha(\tau_1)$ also becomes smaller and delegation eventually may become non-beneficial for all α (for example, see subfigure 4a). This is the regime where the average ability of applicants selected by the agents is not sufficient to compensate for the lack of selection based on fit.

In general, monotonicity of the delegation decision depends on the size of τ_1 compared to $\tilde{\tau}^*_{\alpha=0}$, the threshold the principal would choose if $\alpha = 0$.

- If $\tau_1 \geq \tilde{\tau}^*_{\alpha=0}$, the delegation decision is monotonic in α. For α up to some threshold, the principal always delegates, and for α larger than the threshold, the principal never delegates.
- If $\tau_1 << \tilde{\tau}^*_{\alpha=0}$, again the decision is fully monotonic, but the principal never delegates.
- If $\tau_1 < \tilde{\tau}^*_{\alpha=0}$, but these two thresholds are sufficiently close, the delegation decision may be non-monotonic. Roughly, when α is close to $1/2$ and the agent uses a threshold close to but smaller than the optimal threshold for metric s, the principal may prefer to delegate (also refer to Figure 6 in Appendix F for a graphical illustration).

Fairness in Multi-Group Settings. Building up on Sections 3 and 4, we note that neither process can achieve *fairness* (in terms of demographic parity) across all settings. However, the modes of failure differ across the two processes. Importantly, *the nature of disadvantage faced by a group directly determines whether unfairness caused by this disadvantage can be mitigated or not.* For example, when the disadvantage exists purely as additive biases in the signal mean, a decision-maker (like the principal) can correct for it through group-aware selection policies, without incurring any dis-utility; in this case, not delegating leads to significantly better outcomes for fairness. However, there are some other forms of disadvantage that cannot be corrected for without sacrificing process efficiency. For example, in the disparate variance setting, a rational principal would choose to completely exclude the disadvantaged group (because signals are unreliable) to minimize the risk of hiring poor quality applicants and lowering their utility. In this case, neither of our selection processes is capable of achieving fair outcomes, but delegating to the agent at least ensures selections from both groups (albeit at unfair rates). On a broader note, this highlights the importance of i) understanding the exact type of disparities faced by different populations; and ii) the extent of flexibility a decision maker has while designing selection processes. There is no one-size-fits-all solution.



(a) $\tau_1 = 0.5$ (b) $\tau_1 = 1.0$ (c) $\tau_1 = 2.4$

Figure 4: We plot $\Delta_{quality}$ and $\Delta_{utility}$ as a function of α for different levels of agent's threshold τ_1. Positive values of Δ indicate that the principal prefers delegation. Parameter combination: $c_{rev} = 0.1$, $\sigma_f = 1$, $\sigma_s = 2$, $\sigma_{\tilde{f}} = 1.12$, $\sigma_{\tilde{s}} = 2.06$. For reference, the principal's threshold at $\alpha = 0$ for these parameters is $\tilde{\tau}^* = 1.24$. At $\alpha = 0$, when the principal and agent's metrics are perfectly aligned, comparing $\tilde{\tau}^*$ with τ_1 directly tells us if delegation is better. For a more comprehensive graphical representation of regimes where delegation is beneficial, see Appendix F.

References

Gagan Aggarwal, Ashwinkumar Badanidiyuru, Santiago R Balseiro, Kshipra Bhawalkar, Yuan Deng, Zhe Feng, Gagan Goel, Christopher Liaw, Haihao Lu, Mohammad Mahdian, et al. Auto-bidding and auctions in online advertising: A survey. *ACM SIGecom Exchanges*, 22(1):159–183, 2024.

Evgeni Aizenberg, Matthew J Dennis, and Jeroen van den Hoven. Examining the assumptions of ai hiring assessments and their impact on job seekers' autonomy over self-representation. *AI & Society*, 40(2): 919–927, 2025.

Ricardo Alonso and Niko Matouschek. Relational delegation. *The RAND Journal of Economics*, 38(4): 1070–1089, 2007.

Rohan Alur, Manish Raghavan, and Devavrat Shah. Human expertise in algorithmic prediction. *Advances in Neural Information Processing Systems*, 37:138088–138129, 2024.

Seth C. Anderson and Donald Arthur Winslow. Defining suitability. *Kentucky Law Journal*, 81:105, 1992-1993.

Kenneth Arrow. The theory of discrimination. 1971.

Kyung Hwan Baik and In-Gyu Kim. Delegation in contests. *European Journal of Political Economy*, 13(2): 281–298, 1997.

Gagan Bansal, Besmira Nushi, Ece Kamar, Walter S Lasecki, Daniel S Weld, and Eric Horvitz. Beyond accuracy: The role of mental models in human-ai team performance. In *Proceedings of the AAAI conference on human computation and crowdsourcing*, volume 7, pp. 2–11, 2019.

Gagan Bansal, Besmira Nushi, Ece Kamar, Eric Horvitz, and Daniel S Weld. Is the most accurate ai the best teammate? optimizing ai for teamwork. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 11405–11414, 2021.

Helmut Bester and Daniel Krähmer. Delegation and incentives. *The RAND Journal of Economics*, 39(3): 664–682, 2008.

Avrim Blum, Kevin Stangl, and Ali Vakilian. Multi stage screening: Enforcing fairness and maximizing efficiency in a pre-existing pipeline. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, pp. 1178–1193, 2022.

Amanda Bower, Sarah N Kitchen, Laura Niss, Martin J Strauss, Alexander Vargas, and Suresh Venkatasubramanian. Fair pipelines. *arXiv preprint arXiv:1707.00391*, 2017.

Andrei Z Broder, Adam Kirsch, Ravi Kumar, Michael Mitzenmacher, Eli Upfal, and Sergei Vassilvitskii. The hiring problem and lake wobegon strategies. *SIAM Journal on Computing*, 39(4):1233–1255, 2010.

Erik Brynjolfsson and Andrew Ng. Big ai can centralize decision-making and power, and that's a problem. In *Missing Links in AI Governance*, pp. 65–88. UNESCO/Mila – Québec Institute of Artificial Intelligence, 2023.

Cindy Candrian and Anne Scherer. Rise of the machines: Delegating decisions to autonomous ai. *Computers in Human Behavior*, 134:107308, 2022. ISSN 0747-5632. doi: https://doi.org/10.1016/j.chb.2022.107308. URL https://www.sciencedirect.com/science/article/pii/S0747563222001303.

Gabriel Carroll. Robustness and linear contracts. *American Economic Review*, 105(2):536–563, 2015.

L Elisa Celis, Amit Deshpande, Tarun Kathuria, and Nisheeth K Vishnoi. How to be fair and diverse? *arXiv preprint arXiv:1610.07183*, 2016.

L Elisa Celis, Amit Kumar, Nisheeth K Vishnoi, and Andrew Xu. Centralized selection with preferences in the presence of biases. In *International Conference on Machine Learning*, pp. 5934–5981. PMLR, 2024.

Stephen Coate and Glenn C Loury. Will affirmative-action policies eliminate negative stereotypes? *The American Economic Review*, pp. 1220–1240, 1993.

Lee Cohen, Zachary C Lipton, and Yishay Mansour. Efficient candidate screening under multiple tests and implications for fairness. *arXiv preprint arXiv:1905.11361*, 2019.

Lee Cohen, Saeed Sharifi-Malvajerdi, Kevin Stangl, Ali Vakilian, and Juba Ziani. Sequential strategic screening. In *International Conference on Machine Learning*, pp. 6279–6295. PMLR, 2023.

Sam Corbett-Davies, Emma Pierson, Avi Feller, Sharad Goel, and Aziz Huq. Algorithmic decision making and the cost of fairness. In *Proceedings of the 23rd acm sigkdd international conference on knowledge discovery and data mining*, pp. 797–806, 2017.

Jeffrey Dastin. Insight - amazon scraps secret ai recruiting tool that showed bias against women. https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/, 2018.

Giovanni De Toni, Nastaran Okati, Suhas Thejaswi, Eleni Straitouri, and Manuel Rodriguez. Towards human-ai complementarity with prediction sets. *Advances in Neural Information Processing Systems*, 37: 31380–31409, 2024.

Chuong Do. More on multivariate gaussians. 2008.

Kate Donahue, Sreenivas Gollapudi, and Kostas Kollias. When are two lists better than one?: Benefits and harms in joint decision-making. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 10030–10038, 2024.

Shaddin Dughmi. Algorithmic information structure design: a survey. *ACM SIGecom Exchanges*, 15(2): 2–24, 2017.

Paul Dütting, Tim Roughgarden, and Inbal Talgam-Cohen. Simple versus optimal contracts. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pp. 369–387, 2019.

Paul Dütting, Michal Feldman, and Inbal Talgam-Cohen. Algorithmic contract theory: A survey. *Foundations and Trends® in Theoretical Computer Science*, 16(3-4):211–411, 2024.

Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. Fairness through awareness. In *Proceedings of the 3rd innovations in theoretical computer science conference*, pp. 214–226, 2012.

Vitalii Emelianov, Nicolas Gast, Krishna P Gummadi, and Patrick Loiseau. On fair selection in the presence of implicit variance. In *Proceedings of the 21st ACM Conference on Economics and Computation*, pp. 649–675, 2020.

Vitalii Emelianov, Nicolas Gast, Krishna P Gummadi, and Patrick Loiseau. On fair selection in the presence of implicit and differential variance. *Artificial Intelligence*, 302:103609, 2022.

Boris Epstein and Will Ma. Selection and ordering policies for hiring pipelines via linear programming. *Operations Research*, 72(5):2000–2013, 2024.

World Economic Forum. Which countries are ahead in the global autonomous vehicle race? `https://www.weforum.org/stories/2025/05/autonomous-vehicles-technology-future/`, 2025.

Sorelle A Friedler, Carlos Scheidegger, and Suresh Venkatasubramanian. The (im) possibility of fairness: Different value systems require different mechanisms for fair decision making. *Communications of the ACM*, 64(4):136–143, 2021.

Nikhil Garg, Hannah Li, and Faidra Monachou. Dropping standardized testing for admissions trades off information and access. *arXiv preprint arXiv:2010.04396*, 2020.

Nikhil Garg, Hannah Li, and Faidra Monachou. Standardized tests and affirmative action: The role of bias and variance. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pp. 261–261, 2021.

Ben Green. Escaping the impossibility of fairness: From formal to substantive algorithmic fairness. *Philosophy & Technology*, 35(4):90, 2022.

Sophie Greenwood, Karen Levy, Solon Barocas, Hoda Heidari, and Jon Kleinberg. Designing algorithmic delegates: The role of indistinguishability in human-ai handoff. In *Proceedings of the 26th ACM Conference on Economics and Computation*, pp. 306–336, 2025.

Silvia Griselda. Gender gap in standardized tests: What are we measuring? *Journal of Economic Behavior & Organization*, 221:191–229, 2024.

Sanford J Grossman and Oliver D Hart. An analysis of the principal-agent problem. In *Foundations of insurance economics: Readings in economics and finance*, pp. 302–340. Springer, 1992.

Moritz Hardt, Eric Price, and Nati Srebro. Equality of opportunity in supervised learning. *Advances in neural information processing systems*, 29, 2016.

Ahmed Helmi and Alois Panholzer. Analysis of the "hiring above the median" selection strategy for the hiring problem. *Algorithmica*, 66(4):762–803, 2013.

Lily Hu and Yiling Chen. A short-term intervention for long-term fairness in the labor market. In *Proceedings of the 2018 World Wide Web Conference*, pp. 1389–1398, 2018.

Michael C. Jensen and William H. Meckling. Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, 3(4):305–360, 1976. ISSN 0304-405X. doi: https://doi.org/10.1016/0304-405X(76)90026-X. URL `https://www.sciencedirect.com/science/article/pii/0304405X7690026X`.

Nathan Kallus, Xiaojie Mao, and Angela Zhou. Assessing algorithmic fairness with unobserved protected class using data combination. *Management Science*, 68(3):1959–1981, 2022.

Sampath Kannan, Aaron Roth, and Juba Ziani. Downstream effects of affirmative action. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pp. 240–248, 2019.

Adam Kapor, Mohit Karnani, and Christopher Neilson. Aftermarket frictions and the cost of off-platform options in centralized assignment mechanisms. *Journal of Political Economy*, 132(7):2346–2395, 2024.

Marzieh Karimi-Haghighi and Carlos Castillo. Enhancing a recidivism prediction tool with machine learning: effectiveness and algorithmic fairness. In *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law*, pp. 210–214, 2021.

Sungdoo Kim. Ai-driven hiring: a boon or a barrier to finding the right talent? *AI & Society*, pp. 1–8, 2025.

Jon Kleinberg. Inherent trade-offs in algorithmic fairness. In *Abstracts of the 2018 ACM International Conference on Measurement and Modeling of Computer Systems*, pp. 40–40, 2018.

Jon Kleinberg and Manish Raghavan. Selection problems in the presence of implicit bias. *arXiv preprint arXiv:1801.03533*, 2018.

Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan, and Ashesh Rambachan. Algorithmic fairness. In *Aea papers and proceedings*, volume 108, pp. 22–27. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, 2018.

Patrick Kline, Evan K Rose, and Christopher R Walters. Systemic discrimination among large us employers. *The Quarterly Journal of Economics*, 137(4):1963–2036, 2022.

Vivian Lai, Samuel Carton, Rajat Bhatnagar, Q Vera Liao, Yunfeng Zhang, and Chenhao Tan. Human-ai collaboration via conditional delegation: A case study of content moderation. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pp. 1–18, 2022.

Jonathan Levin et al. The dynamics of collective reputation. *The BE Journal of Theoretical Economics*, 9 (1):1–25, 2009.

Danielle Li, Lindsey Raymond, and Peter Bergman. Hiring as exploration. *Review of Economic Studies*, pp. rdaf040, 2025.

Tao Lin and Yiling Chen. Generalized principal-agent problem with a learning agent. *arXiv preprint arXiv:2402.09721*, 2024.

Zhi Liu and Nikhil Garg. Test-optional policies: Overcoming strategic behavior and informational gaps. In *Proceedings of the 1st ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, pp. 1–13, 2021.

Brian Lubars and Chenhao Tan. Ask not what ai can do, but what ai should do: Towards a framework of task delegability. *Advances in neural information processing systems*, 32, 2019.

Arthur Lupia. Delegation of power: Agency theory. In James D. Wright (ed.), *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)*, pp. 58–60. Elsevier, Oxford, second edition edition, 2015. ISBN 978-0-08-097087-5. doi: https://doi.org/10.1016/B978-0-08-097086-8.93029-0. URL `https://www.sciencedirect.com/science/article/pii/B9780080970868930290`.

Jérémie Mary, Clément Calauzenes, and Noureddine El Karoui. Fairness-aware learning for continuous attributes and treatments. In *International Conference on Machine Learning*, pp. 4382–4391. PMLR, 2019.

Allen E Milewski and Steven H Lewis. Delegating to software agents. *International Journal of Human-Computer Studies*, 46(4):485–500, 1997.

Hussein Mouzannar, Mesrob I Ohannessian, and Nathan Srebro. From fair decision making to social equality. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pp. 359–368, 2019.

Anna Nowogrodzki. Most us professors are trained at same few elite universities. *Nature*, 609(7929):887, 2022.

NPR. Ai is screening your resume. here's how to make it past the bots. https://www.npr.org/2025/10/03/nx-s1-5534959/are-ai-hiring-tools-any-good-this-journalist-found-widespread-bias-and-bugs, 2025.

Edmund S Phelps. The statistical theory of racism and sexism. *The american economic review*, 62(4): 659–661, 1972.

Julien Picault. Risk-averse managers, labour market structures, public policies and discrimination. *The BE Journal of Theoretical Economics*, 17(1):20150079, 2017.

Evaggelia Pitoura, Kostas Stefanidis, and Georgia Koutrika. Fairness in rankings and recommendations: an overview. *The VLDB Journal*, pp. 1–28, 2022.

Manish Purohit, Sreenivas Gollapudi, and Manish Raghavan. Hiring under uncertainty. In *International Conference on Machine Learning*, pp. 5181–5189. PMLR, 2019.

Manish Raghavan. Competition and diversity in generative ai. *arXiv preprint arXiv:2412.08610*, 2024.

William P Rogerson. On the optimality of menus of linear contracts. Technical report, Discussion paper, 1987.

Stephen A Ross. The economic theory of agency: The principal's problem. *The American economic review*, 63(2):134–139, 1973.

David E M Sappington. Incentives in principal-agent relationships. *Journal of economic Perspectives*, 5(2): 45–66, 1991.

Tomoki Sekiguchi and Vandra L Huber. The use of person–organization fit and person–job fit information in making selection decisions. *Organizational behavior and human decision processes*, 116(2):203–216, 2011.

Dirk Sliwka. On the costs and benefits of delegation in organizations. *Journal of Institutional and Theoretical Economics (JITE) / Zeitschrift für die gesamte Staatswissenschaft*, 157(4):568–590, 2001. ISSN 09324569. URL http://www.jstor.org/stable/40752295.

Cody Taylor. The case for algorithmic pricing: Consumer welfare, market efficiency, and policy missteps. https://www.mercatus.org/research/policy-briefs/case-algorithmic-pricing-consumer-welfare-market-efficiency-and-policy, 2025.

Jean Tirole. A theory of collective reputations (with applications to the persistence of corruption and to firm quality). *The review of economic studies*, 63(1):1–22, 1996.

John Vickers. Delegation and the theory of the firm. *The Economic Journal*, 95(Supplement):138–147, 1985.

K Hunter Wapman, Sam Zhang, Aaron Clauset, and Daniel B Larremore. Quantifying hierarchy and dynamics in us faculty hiring and retention. *Nature*, 610(7930):120–127, 2022.

Yimin Yu and Xiangyin Kong. Robust contract designs: Linear contracts and moral hazard. *Operations Research*, 68(5):1457–1473, 2020.

Meike Zehlike, Francesco Bonchi, Carlos Castillo, Sara Hajian, Mohamed Megahed, and Ricardo Baeza-Yates. Fa* ir: A fair top-k ranking algorithm. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pp. 1569–1578, 2017.

Meike Zehlike, Philipp Hacker, and Emil Wiedemann. Matching code and law: achieving algorithmic fairness with optimal transport. *Data Mining and Knowledge Discovery*, 34(1):163–200, 2020.

Rebecca Zwick. Is the sat a "wealth test?" the link between educational achievement and socioeconomic status. In *Rethinking the SAT*, pp. 203–216. Routledge, 2013.

A Proofs for Section 2

A.1 Proof of Claim 1

We know that s and t are jointly gaussian mean-zero random variables with variances σ_s^2 and σ_t^2 respectively and a correlation coefficient $\rho \geq 0$. We decompose t as follows:

$$t = \underbrace{\frac{Cov(t,s)}{V(s)}}_{\gamma_1} \cdot s + \underbrace{t - \frac{Cov(t,s)}{V(s)} \cdot s}_{f}.$$

We now claim that $f \perp s$. In order to prove this, we need to show that: i) (s, f) are also jointly Gaussian, and ii) $Cov(s, f) = 0$. The first statement follows from the fact that (s, f) can be expressed as a linear transformation of (s, t) which are themselves jointly Gaussian:

$$\begin{bmatrix} s \\ f \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -\frac{Cov(t,s)}{V(s)} & 1 \end{bmatrix} \begin{bmatrix} s \\ t \end{bmatrix},$$

therefore (s, f) must also be jointly Gaussian. We now compute $Cov(s, f)$ directly:

$$Cov(s,f) = Cov\left(s, t - \frac{Cov(t,s)}{V(s)} \cdot s\right) = Cov(s,t) - \frac{Cov(t,s)}{V(s)} \cdot Cov(s,s) = Cov(s,t) - \frac{Cov(s,t)}{V(s)} \cdot V(s) = 0.$$

Two jointly gaussian random variables also being uncorrelated automatically implies independence. Therefore, f must also be Gaussian. Further,

$$\mathbb{E}[f] = \mathbb{E}\left[t - \frac{Cov(t,s)}{V(s)} \cdot s\right] = \mathbb{E}[t] - \frac{Cov(t,s)}{V(s)} \cdot \mathbb{E}[s] = 0.$$

$$\begin{aligned} V(f) &= V\left(t - \frac{Cov(t,s)}{V(s)} \cdot s\right) \\ &= V(t) + \left(\frac{Cov(t,s)}{V(s)}\right)^2 \cdot V(s) - 2\left(\frac{Cov(t,s)}{V(s)}\right) \cdot Cov(t,s) \\ &= \sigma_t^2 + \frac{\rho^2\sigma_t^2}{\sigma_s^2} \cdot \sigma_s^2 - 2 \cdot \frac{(\rho\sigma_t\sigma_s)^2}{\sigma_s^2} \\ &= \sigma_t^2 + \rho^2\sigma_t^2 - 2\rho^2\sigma_t^2 \\ &= (1-\rho^2)\sigma_t^2. \end{aligned}$$

Thus, we have shown that t can be expressed as a linear combination $\gamma_1 s + \gamma_2 f$ of s and another mean-zero Gaussian random variable f with variance $\sigma_f^2 = (1-\rho^2)\sigma_t^2$ and $f \perp s$. In this case, $\gamma_1 = \frac{Cov(t,s)}{V(s)} = \frac{\rho\sigma_t}{\sigma_s} \geq 0$ and $\gamma_2 = 1$. This concludes the proof of the claim.

B Proofs for Section 3

B.1 Proof of Lemma 1

Recall that the principal's expected utility from a applicant hired through a selection process which has been delegated to the agent is given by $\mathbb{E}[t \mid \tilde{s} \geq \tau_1]$ Before we proceed with the main proof, we need to introduce some technical results:

Claim 5. *If two random variables X and Y are such that $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$, $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ and Σ is the covariance between X and Y, then the conditional distribution of X conditional on $Y = y$ is given by:*

$$X \mid (Y = y) \sim \mathcal{N}\left(\mu_X + \frac{\Sigma}{\sigma_Y^2} \cdot (y - \mu_Y), \sigma_X^2 - \frac{\Sigma^2}{\sigma_Y^2}\right)$$

Proof. For a detailed proof of the above claim, please refer to these notes Do (2008). □

Noting that $s \sim \mathcal{N}(0, \sigma_s^2)$, $\tilde{s} \sim \mathcal{N}(0, \sigma_{\tilde{s}}^2)$ and $Cov(s, \tilde{s}) = \sigma_s^2$, we can use the above claim to conclude that:

$$\mathbb{E}[s \mid \tilde{s} = x] = \frac{\sigma_s^2}{\sigma_{\tilde{s}}^2} x.$$

We will now use this result to prove the following technical result that links $\mathbb{E}[s \mid \tilde{s} \geq \tau_1]$ to $\mathbb{E}[\tilde{s} \mid \tilde{s} \geq \tau_1]$:

Claim 6. *We must have:* $\mathbb{E}[s \mid \tilde{s} \geq \tau_1] = \left(\frac{\sigma_s}{\sigma_{\tilde{s}}}\right)^2 \cdot \mathbb{E}[\tilde{s} \mid \tilde{s} \geq \tau_1]$.

Proof. The proof follows directly from Claim 5 and the law of total expectation.

$$\begin{aligned}
\mathbb{E}[s \mid \tilde{s} \geq \tau_1] &= \mathbb{E}\left[\mathbb{E}[s \mid \tilde{s}] \mid \tilde{s} \geq \tau_1\right] \quad \text{(law of total expectation)} \\
&= \mathbb{E}\left[\frac{\sigma_s^2}{\sigma_{\tilde{s}}^2} \tilde{s} \mid \tilde{s} \geq \tau_1\right] \quad \text{(by Claim 5)} \\
&= \frac{\sigma_s^2}{\sigma_{\tilde{s}}^2} \cdot \mathbb{E}[\tilde{s} \mid \tilde{s} \geq \tau_1].
\end{aligned}$$

□

Now, if we can compute $\mathbb{E}[\tilde{s} \mid \tilde{s} \geq \tau_1]$, we are done. To this end, we present technical Claim 7 (without proof) which provides the closed form expression for the mean of a truncated normal random variable.

Claim 7. *Let $X \sim N(\mu_X, \sigma_X^2)$. Suppose, the lower tail of X is truncated at a. Then,*

$$\mathbb{E}[X \mid X \geq a] = \mu_X + \sigma_X \cdot H\left(\frac{a - \mu_X}{\sigma_X}\right),$$

where $H(y) = \frac{\phi(y)}{1 - \Phi(y)}$ represents the hazard rate of a standard normal random variable.

Proof. For completeness, we provide a detailed proof in Appendix E.1. □

Noting that $\tilde{s} \sim \mathcal{N}(0, \sigma_{\tilde{s}}^2)$ and using Claim 7, we have $\mathbb{E}[\tilde{s} \mid \tilde{s} \geq \tau_1] = \sigma_{\tilde{s}} \cdot H\left(\frac{\tau_1}{\sigma_{\tilde{s}}}\right)$. Putting everything together, we have:

$$\begin{aligned}
\mathbb{E}[t \mid \tilde{s} \geq \tau_1] &= \mathbb{E}[\alpha f + (1 - \alpha) s \mid \tilde{s} \geq \tau_1] \\
&= \alpha \cdot \mathbb{E}[f \mid \tilde{s} \geq \tau_1] + (1 - \alpha) \cdot \mathbb{E}[s \mid \tilde{s} \geq \tau_1] \\
&= \alpha \cdot \mathbb{E}[f] + (1 - \alpha) \cdot \mathbb{E}[s \mid \tilde{s} \geq \tau_1] \quad (\text{since } f \perp s, f \perp \epsilon_s) \\
&= (1 - \alpha) \cdot \mathbb{E}[s \mid \tilde{s} \geq \tau_1] \quad (\text{since } \mathbb{E}[f] = 0) \\
&= (1 - \alpha) \cdot \frac{\sigma_s^2}{\sigma_{\tilde{s}}^2} \cdot \mathbb{E}[\tilde{s} \mid \tilde{s} \geq \tau_1] \quad \text{(Claim 6)} \\
&= (1 - \alpha) \cdot \frac{\sigma_s^2}{\sigma_{\tilde{s}}^2} \cdot \sigma_{\tilde{s}} \cdot H\left(\frac{\tau_1}{\sigma_{\tilde{s}}}\right) \quad \text{(Claim 7)} \\
&= \frac{(1 - \alpha)\sigma_s^2}{\sigma_{\tilde{s}}} \cdot H\left(\frac{\tau_1}{\sigma_{\tilde{s}}}\right).
\end{aligned}$$

B.2 Proof of Lemma 2

Suppose, groups A and B have demographic ratios of λ and $1 - \lambda$ respectively. Let $\Phi_A(\cdot)$ and $\Phi_B(\cdot)$ indicate the CDFs of the test score distributions of each group. The test score distribution $\Phi_M(\cdot)$ of the mixture distribution is therefore given as follows:

$$\Phi_M(x) = \lambda \cdot \Phi_A(x) + (1 - \lambda) \cdot \Phi_B(x) \quad \forall\, x \in \mathbb{R},$$

and the corresponding complementary CDF is given by:

$$\bar{\Phi}_M(x) = \lambda \cdot \bar{\Phi}_A(x) + (1-\lambda) \cdot \bar{\Phi}_B(x) \quad \forall\, x \in \mathbb{R}$$

Now,

$$\begin{aligned}
&\mathbb{P}\left[\text{randomly picked applicant belongs to Gr A} \mid \text{test score} \geq \tau_1\right] \\
=&\frac{\mathbb{P}\left[\text{randomly picked applicant belongs to Gr A, test score} \geq \tau_1\right]}{\mathbb{P}\left[\text{a random draw from } \Phi(\cdot) \text{ has a value} \geq \tau_1\right]} \\
=&\frac{\mathbb{P}\left[\text{randomly picked applicant belongs to Gr A}\right] \cdot \mathbb{P}\left[\text{test score} \geq \tau_1 \mid \text{applicant belongs to Gr A}\right]}{\mathbb{P}\left[\text{a random draw from } \Phi_M(\cdot) \text{ has a value} \geq \tau_1\right]} \\
=&\frac{\lambda \cdot \bar{\Phi}_A(\tau_1)}{\bar{\Phi}_M(\tau_1)}.
\end{aligned}$$

Similarly, for group B,

$$\mathbb{P}\left[\text{randomly picked applicant belongs to Gr B} \mid \text{test score} \geq \tau_1\right] = \frac{(1-\lambda) \cdot \bar{\Phi}_B(\tau_1)}{\bar{\Phi}_M(\tau_1)}.$$

This concludes the proof of the lemma.

B.3 Proof of Theorem 1

Suppose that K applicants have been hired from the mixed applicant pool at threshold τ_1 through a selection process delegated to the agent. Let X_i be the indicator random variable that takes value 1 if hired applicant i belongs to Group A, 0 otherwise. Lemma 2 shows that $X_i \sim \text{Ber}\left(\frac{\Lambda_A \cdot \bar{\Phi}_A(\tau_1)}{\bar{\Phi}_M(\tau_1)}\right)$. Note that can express the fairness metric $\mathcal{D}$ in terms of X_i's as follows:

$$\mathcal{D} = \frac{1}{K} \sum_{i=1}^{K} \underbrace{\left(\frac{X_i}{\Lambda_A} - \frac{1-X_i}{\Lambda_B}\right)}_{Z_i}.$$

Therefore, for $i \in [K]$, Z_i has the following pmf:

$$Z_i = \begin{cases} \frac{1}{\lambda} & \text{w.p.} \quad \frac{\lambda \cdot \bar{\Phi}_A(\tau_1)}{\bar{\Phi}_M(\tau_1)}, \\ -\frac{1}{1-\lambda} & \text{w.p.} \quad \frac{(1-\lambda) \cdot \bar{\Phi}_B(\tau_1)}{\bar{\Phi}_M(\tau_1)}. \end{cases}$$

Then, we have $\mathbb{E}[Z_i] = \frac{\bar{\Phi}_A(\tau_1) - \bar{\Phi}_B(\tau_1)}{\bar{\Phi}_M(\tau_1)}$. Further note that Z_i's are mutually independent. Now,

$$\mathbb{E}[|\mathcal{D}|] = \mathbb{E}\left[\frac{1}{K}\left|\sum_{i=1}^{K} Z_i\right|\right] = \frac{1}{K}\mathbb{E}\left[\left|\sum_{i=1}^{K} Z_i\right|\right] \leq \frac{1}{K}\mathbb{E}\left[\sum_{i=1}^{K} |Z_i|\right] = \mathbb{E}[|Z_i|] = \frac{\bar{\Phi}_A(\tau_1) + \bar{\Phi}_B(\tau_1)}{\bar{\Phi}_M(\tau_1)}.$$

The inequality used above follows from triangle inequality. We can also derive a lower bound using Jensen' inequality for convex functions as follows:

$$\mathbb{E}[|\mathcal{D}|] = \mathbb{E}\left[\frac{1}{K}\left|\sum_{i=1}^{K} Z_i\right|\right] = \frac{1}{K}\mathbb{E}\left[\left|\sum_{i=1}^{K} Z_i\right|\right] \geq \frac{1}{K}\left|\mathbb{E}[\sum_{i=1}^{K} Z_i]\right| = |\mathbb{E}[Z_i]| = \frac{|\bar{\Phi}_A(\tau_1) - \bar{\Phi}_B(\tau_1)|}{\bar{\Phi}_M(\tau_1)}.$$

C Proofs for Section 4

C.1 Proof of Theorem 2

We will complete the proof in 3 parts: in the first part, we will compute the conditional expectation term so that we can rewrite the objective function in problem equation 5 in closed form. In the second part, we will argue how to solve the optimization problem itself. Finally, in part 3, we will argue about properties of the optimal threshold that enable efficient computation.

Part 1: Computing $\mathbb{E}[t \mid \tilde{t} \geq \tilde{\tau}]$

$$\begin{aligned}\mathbb{E}[t \mid \tilde{t} \geq \tilde{\tau}] &= \frac{\sigma_t^2}{\sigma_{\tilde{t}}^2} \cdot \mathbb{E}[\tilde{t} \mid \tilde{t} \geq \tilde{\tau}] \quad \text{(identical to Claim 6)} \\ &= \frac{\sigma_t^2}{\sigma_{\tilde{t}}^2} \cdot \sigma_{\tilde{t}} \cdot H\left(\frac{\tilde{\tau}}{\sigma_{\tilde{t}}}\right) \quad \text{(identical to Claim 7)} \\ &= \frac{\sigma_t^2}{\sigma_{\tilde{t}}} \cdot \frac{\phi\left(\tilde{\tau}/\sigma_{\tilde{t}}\right)}{\Phi^c\left(\tilde{\tau}/\sigma_{\tilde{t}}\right)}.\end{aligned}$$

Part 2: Solving the optimization problem

Now, we define:

$$\begin{aligned}O^* = \max_{n_{rev} \geq 0, \tilde{\tau}} \quad & n_{rev} \cdot \mathbb{P}[\tilde{t} \geq \tilde{\tau}] \cdot \mathbb{E}[t \mid \tilde{t} \geq \tilde{\tau}] - n_{rev} \cdot c_{rev} \\ s.t. \quad & n_{rev} \cdot \mathbb{P}[\tilde{t} \geq \tilde{\tau}] \leq k.\end{aligned}$$

Now, note that any solution $(n, \tilde{\tau})$ of the form $(0, \tilde{\tau})$ is feasible and produces an objective value of 0. Therefore, $O^* \geq 0$. There are 2 cases:

1. Case 1 ($O^* > 0$): Suppose, $(n_{rev}^*, \tilde{\tau}^*)$ is the optimal solution which produces O^*. Clearly, $n_{rev}^* > 0$ and $\mathbb{P}[\tilde{t} \geq \tilde{\tau}^*] \cdot \mathbb{E}[t \mid \tilde{t} \geq \tilde{\tau}^*] - c_{rev} > 0$. We claim that the constraint: $n_{rev} \cdot \mathbb{P}[\tilde{t} \geq \tilde{\tau}] \leq k$ must be active at $(n_{rev}^*, \tilde{\tau}^*)$. This follows directly from the fact that the objective function is increasing in n_{rev} at the optimal solution $\tilde{\tau}^*$. Therefore, we must have:

$$n_{rev}^* = \frac{k}{\mathbb{P}[\tilde{t} \geq \tilde{\tau}^*]} \quad \text{if } \mathbb{P}[\tilde{t} \geq \tilde{\tau}^*] \cdot \mathbb{E}[t \mid \tilde{t} \geq \tilde{\tau}^*] > c_{rev}.$$

   In this case, we can reduce the optimization problem to the following form:

$$\max_{\tilde{\tau}} \quad \frac{k}{\mathbb{P}[\tilde{t} \geq \tilde{\tau}]} \cdot \mathbb{P}[\tilde{t} \geq \tilde{\tau}] \cdot \mathbb{E}[t \mid \tilde{t} \geq \tilde{\tau}] - c_{rev} \cdot \frac{k}{\mathbb{P}[\tilde{t} \geq \tilde{\tau}]}$$

   or equivalently,

$$\tilde{\tau}^* = arg\max_{\tilde{\tau}} \quad \mathbb{E}[t \mid \tilde{t} \geq \tilde{\tau}] - \frac{c_{rev}}{\mathbb{P}[\tilde{t} \geq \tilde{\tau}]}$$

   Substituting the expressions for $\mathbb{E}[t \mid \tilde{t} \geq \tilde{\tau}]$ and $\mathbb{P}[\tilde{t} \geq \tilde{\tau}]$, we obtain:

$$\tilde{\tau}^* = \ arg\max_{\tilde{\tau}} \quad \underbrace{\frac{\sigma_t^2}{\sigma_{\tilde{t}}} \cdot \frac{\phi\left(\tilde{\tau}/\sigma_{\tilde{t}}\right)}{\Phi^c\left(\tilde{\tau}/\sigma_{\tilde{t}}\right)} - \frac{c_{rev}}{\Phi^c\left(\tilde{\tau}/\sigma_{\tilde{t}}\right)}}_{v(\tilde{\tau})}$$

2. Case 2 ($O^* = 0$): From the analysis in Case 1, it is clear that $O^* = 0$ if and only if $\mathbb{P}[\tilde{t} \geq \tilde{\tau}] \cdot \mathbb{E}[t \mid \tilde{t} \geq \tilde{\tau}] - c_{rev} \leq 0$ for all $\tilde{\tau}$ or equivalently, $\mathbb{E}[t \mid \tilde{t} \geq \tilde{\tau}] - \frac{c_{rev}}{\mathbb{P}[\tilde{t} \geq \tilde{\tau}]} \leq 0$ for all $\tilde{\tau}$ (including $\tilde{\tau}^*$ obtained in Case 1). In this case, $n_{rev}^* = 0$.

So far, we have shown that $n_{rev}^* > 0$ if and only if $v^* = \frac{\sigma_t^2}{\sigma_{\tilde{t}}} \cdot \frac{\phi(\tilde{\tau}^*/\sigma_{\tilde{t}})}{\Phi^c(\tilde{\tau}^*/\sigma_{\tilde{t}})} - \frac{c_{rev}}{\Phi^c(\tilde{\tau}^*/\sigma_{\tilde{t}})} > 0$. We now need to show the following:

$$v^* > 0 \iff c_{rev} < \frac{1}{\sqrt{2\pi}} \cdot \frac{\sigma_t^2}{\sigma_{\tilde{t}}}.$$

For the forward direction ($\implies$), we have:

$$\begin{aligned} v^* > 0 &\implies \frac{\sigma_t^2}{\sigma_{\tilde{t}}} \cdot \phi(\tilde{\tau}^*/\sigma_{\tilde{t}}) - c_{rev} > 0 \quad (\text{since for finite } \tilde{\tau}^*,\ \frac{1}{\Phi^c(\tilde{\tau}^*/\sigma_{\tilde{t}})} > 0) \\ &\implies c_{rev} < \frac{\sigma_t^2}{\sigma_{\tilde{t}}} \cdot \phi(\tilde{\tau}^*/\sigma_{\tilde{t}}) \\ &\implies c_{rev} < \frac{\sigma_t^2}{\sigma_{\tilde{t}}} \cdot \phi(0) \quad (\phi(x) \text{ is maximized at } x = 0) \\ &\implies c_{rev} < \frac{1}{\sqrt{2\pi}} \cdot \frac{\sigma_t^2}{\sigma_{\tilde{t}}}. \end{aligned}$$

For the other direction ($\impliedby$), we first consider the principal's net utility per applicant hired at threshold $\tilde{\tau}$, given by:

$$v(\tilde{\tau}) = \frac{1}{\Phi^c(\tilde{\tau}/\sigma_{\tilde{t}})} \left[\frac{\sigma_t^2}{\sigma_{\tilde{t}}} \cdot \phi(\tilde{\tau}/\sigma_{\tilde{t}}) - c_{rev} \right].$$

Now,

$$\begin{aligned} c_{rev} < \frac{1}{\sqrt{2\pi}} \cdot \frac{\sigma_t^2}{\sigma_{\tilde{t}}} &\implies \frac{\sigma_t^2}{\sigma_{\tilde{t}}} \cdot \phi(0) - c_{rev} > 0 \\ &\implies v(0) > 0 \\ &\implies v^* > 0 \quad (\text{since } v^* = \max_{\tilde{\tau}}\ v(\tilde{\tau})). \end{aligned}$$

Finally, using the simplified condition, we can characterize the optimal solution of the principal's optimization problem as follows:

- Solve for $\tilde{\tau}^* = arg\max_{\tilde{\tau}} \quad \frac{\sigma_t^2}{\sigma_{\tilde{t}}} \cdot \frac{\phi(\tilde{\tau}/\sigma_{\tilde{t}})}{\Phi^c(\tilde{\tau}/\sigma_{\tilde{t}})} - \frac{c_{rev}}{\Phi^c(\tilde{\tau}/\sigma_{\tilde{t}})}$.
- If $c_{rev} < \frac{1}{\sqrt{2\pi}} \cdot \frac{\sigma_t^2}{\sigma_{\tilde{t}}}$, choose $n^*_{rev} = \frac{k}{\mathbb{P}[\tilde{t} \geq \tilde{\tau}^*]}$.
- Else, choose $n^*_{rev} = 0$.

Part 3: Showing that $\tilde{\tau}^*$ is unique and can be computed efficiently.

In order to complete the last part of the proof, we will introduce the following claim:

Claim 8. *$\tilde{\tau}^*$, which maximizes the objective in (6), is the unique solution to the following non-linear equation:*

$$g\left(\frac{\tilde{\tau}}{\sigma_{\tilde{t}}}\right) := \frac{\sigma_t^2}{\sigma_{\tilde{t}}} \cdot \left[\phi\left(\frac{\tilde{\tau}}{\sigma_{\tilde{t}}}\right) - \frac{\tilde{\tau}}{\sigma_{\tilde{t}}} \cdot \Phi^c\left(\frac{\tilde{\tau}}{\sigma_{\tilde{t}}}\right) \right] - c_{rev} = 0. \tag{8}$$

In particular, $g(\cdot)$ is monotonically decreasing and $\tilde{\tau}^*$ is finite and can be obtained easily using binary search.

Proof. Recall that:

$$\tilde{\tau}^* = arg\max_{\tilde{\tau}} \quad \underbrace{\frac{\frac{\sigma_t^2}{\sigma_{\tilde{t}}} \cdot \phi(\tilde{\tau}/\sigma_{\tilde{t}}) - c_{rev}}{\Phi^c(\tilde{\tau}/\sigma_{\tilde{t}})}}_{v(\tilde{\tau})},$$

Define $a = \frac{\sigma_t^2}{\sigma_{\tilde{t}}}$ and $z = \frac{\tilde{\tau}}{\sigma_{\tilde{t}}}$. Using this transformation, we have rewrite $v(\tilde{\tau})$ as $w(z)$ where:

$$w(z) = \frac{a \cdot \phi(z) - c_{rev}}{\Phi^c(z)},$$

Now, if $z^* = arg\max_z\ w(z)$, it should be immediately clear that $\tilde{\tau}^* = \sigma_{\tilde{t}} \cdot z^*$. Therefore, our goal now is to show that z^* is the unique solution to $g(z) := a \cdot \phi(z) - a \cdot z \cdot \Phi^c(z) - c_{rev} = 0$. Note that $w(z)$ is differentiable in z, therefore using the first order conditions, we conclude that any local extremum should satisfy:

$$\begin{aligned}
w'(z) = 0 &\implies \frac{a \cdot \Phi^c(z) \cdot \phi'(z) + \phi(z)\left(a \cdot \phi(z) - c_{rev}\right)}{\left(\Phi^c(z)\right)^2} = 0 \\
&\implies \frac{-a \cdot z \cdot \Phi^c(z) \cdot \phi(z) + \phi(z)\left(a \cdot \phi(z) - c_{rev}\right)}{\left(\Phi^c(z)\right)^2} = 0 \quad (\text{since } \phi'(z) = -z \cdot \phi(z)) \\
&\implies \frac{\phi(z)}{\left(\Phi^c(z)\right)^2} \cdot \left[-a \cdot z \cdot \Phi^c(z) + a \cdot \phi(z) - c_{rev}\right] = 0 \\
&\implies \frac{\phi(z)}{\left(\Phi^c(z)\right)^2} \cdot g(z) = 0.
\end{aligned}$$

We claim that $\pm\infty$ cannot be the maximizers of $w(z)$. This follows from the fact that $\lim_{z\to+\infty} w(z) = -\infty$ and $\lim_{z\to-\infty} w(z) = -c_{rev}$, but $u(0) > -c_{rev}$. Therefore, the maximizer of $w(z)$ must be finite which immediately implies that the only possible maximizer of $w(z)$ must be a solution of $g(z) = 0$.

Now, $g(z)$ is continuous in z and $\lim_{z\to\infty} g(z) > 0$ and $\lim_{z\to-\infty} g(z) < 0$. Therefore, by the intermediate value theorem, there must be a finite solution to $g(z) = 0$. Additionally, we have:

$$\begin{aligned}
g'(z) &= a \cdot \phi'(z) - a \cdot \Phi^c(z) - a \cdot z \cdot (-\phi(z)) \\
&= -a \cdot z \cdot \phi(z) - a \cdot \Phi^c(z) + a \cdot z \cdot \phi(z) \quad (\text{using } \phi'(z) = -z \cdot \phi(z)) \\
&= -a \cdot \Phi^c(z) < 0.
\end{aligned}$$

This means that $g(z)$ is monotonically decreasing in z which implies the following:

- $g(z) = 0$ has a unique solution z^*;
- $w'(z) > 0$ for $z < z^*$ and $w'(z) < 0$ for $z > z^*$, showing that $w(z)$ is concave and z^* is the unique maximizer of $w(z)$; and
- z^* can be obtained using binary search.

This concludes the proof of the claim. □

C.2 Equivalent Characterizations of Corollary 4

Claim 9. *The following statements are equivalent:*

a) $\tilde{\tau}^* > 0$,

b) $n^*_{rev} > 0$,

c) $c_{rev} < \frac{1}{\sqrt{2\pi}} \cdot \frac{\sigma_t^2}{\sigma_{\tilde{t}}}$.

Proof. In order to complete the proof, we will show that $c \implies b$, $b \implies a$ and $a \implies c$ in that order.

1. The first part (showing $c \implies b$) follows directly from the statement of Theorem 2.
2. For the second part ($b \implies a$), suppose that $n^*_{rev} > 0$. Therefore, it must be that $v^* = v(\tilde{\tau}^*)$ must be > 0. This implies:
$$\frac{1}{\Phi^c(\tilde{\tau}^*/\sigma_{\tilde{t}})} \cdot \left[\frac{\sigma_t^2}{\sigma_{\tilde{t}}} \cdot \phi(\tilde{\tau}^*/\sigma_{\tilde{t}}) - c_{rev}\right] > 0,$$

which implies that $\frac{\sigma_t^2}{\sigma_{\tilde{t}}} \cdot \phi(\tilde{\tau}^*/\sigma_{\tilde{t}}) - c_{rev} > 0$. But, we know that $g(\tilde{\tau}^*/\sigma_{\tilde{t}}) = 0$ which means that:

$$\frac{\sigma_t^2}{\sigma_{\tilde{t}}} \cdot \phi(\tilde{\tau}^*/\sigma_{\tilde{t}}) - c_{rev} = \frac{\sigma_t^2}{\sigma_{\tilde{t}}} \cdot \frac{\tilde{\tau}^*}{\sigma_{\tilde{t}}} \cdot \Phi^c(\tilde{\tau}^*/\sigma_{\tilde{t}}).$$

Then, $\tilde{\tau}^* \cdot \Phi^c(\tilde{\tau}^*/\sigma_{\tilde{t}}) > 0$ which clearly implies that $\tilde{\tau}^* > 0$.

3. Finally, the first part of the proof ($a \implies c$), note that $\tilde{\tau}^* > 0$ implies that $g\left(\frac{\tilde{\tau}^*}{\sigma_{\tilde{t}}}\right) < g(0) = \frac{1}{\sqrt{2\pi}} \cdot \frac{\sigma_t^2}{\sigma_{\tilde{t}}} - c_{rev}$ since $g(\cdot)$ is monotonically decreasing in its argument. But, $g\left(\frac{\tilde{\tau}^*}{\sigma_{\tilde{t}}}\right) = 0$. Therefore, it must be that $c_{rev} < \frac{1}{\sqrt{2\pi}} \cdot \frac{\sigma_t^2}{\sigma_{\tilde{t}}}$. This concludes the proof.

□

C.3 Proof of Lemma 3

The proof of the lemma follows directly from Claim 8. Recall that if $\tilde{\tau}^*$ is the optimal threshold for the principal, then it must satisfy $g(\tilde{\tau}^*/\sigma_{\tilde{t}}) = 0$. This implies:

$$\begin{aligned}
&\frac{\sigma_t^2}{\sigma_{\tilde{t}}} \cdot \phi\left(\frac{\tilde{\tau}^*}{\sigma_{\tilde{t}}}\right) - \frac{\sigma_t^2}{\sigma_{\tilde{t}}^2} \cdot \tilde{\tau}^* \cdot \Phi^c\left(\frac{\tilde{\tau}^*}{\sigma_{\tilde{t}}}\right) - c_{rev} = 0 \\
\implies &\frac{\sigma_t^2}{\sigma_{\tilde{t}}} \cdot \phi\left(\frac{\tilde{\tau}^*}{\sigma_{\tilde{t}}}\right) - c_{rev} = \frac{\sigma_t^2}{\sigma_{\tilde{t}}^2} \cdot \tilde{\tau}^* \cdot \Phi^c\left(\frac{\tilde{\tau}^*}{\sigma_{\tilde{t}}}\right) \\
\implies &\frac{\sigma_t^2}{\sigma_{\tilde{t}}} \cdot \frac{\phi\left(\tilde{\tau}^*/\sigma_{\tilde{t}}\right)}{\Phi^c\left(\tilde{\tau}^*/\sigma_{\tilde{t}}\right)} - \frac{c_{rev}}{\Phi^c\left(\tilde{\tau}^*/\sigma_{\tilde{t}}\right)} = \frac{\sigma_t^2}{\sigma_{\tilde{t}}^2} \cdot \tilde{\tau}^*.
\end{aligned}$$

The LHS represents the principal's utility per hired applicant at threshold $\tilde{\tau}^*$. This concludes the proof of the lemma.

C.4 Monotonicity Results

Claim 10. *If $c_{rev,1} > c_{rev,2}$, then $\tilde{\tau}_1^* < \tilde{\tau}_2^*$ and $n_{rev,1}^* \leq n_{rev,2}^*$.*

Proof. The order of optimal thresholds follows directly from noting that the function $g\left(\frac{\tilde{\tau}}{\sigma_{\tilde{t}}}\right)$ is monotonically decreasing in $\tilde{\tau}$. Now, there are 3 cases:

1. $0 < \tilde{\tau}_1^* < \tilde{\tau}_2^* \implies n_{rev,1}^* < n_{rev,2}^*$ since $n_{rev}^* = \frac{k - n_1}{\mathbb{P}[\tilde{t} \geq \tilde{\tau}^*]}$.
2. $\tilde{\tau}_1^* \leq 0 < \tilde{\tau}_2^* \implies 0 = n_{rev,1}^* < n_{rev,2}^*$ (by Claim 9).
3. $\tilde{\tau}_1^* < \tilde{\tau}_2^* \leq 0 \implies n_{rev,1}^* = n_{rev,2}^* = 0$.

This concludes the proof. □

Claim 11. *Suppose, c_{rev} is small enough that the optimal threshold is positive. Then, a higher value of σ_t leads to a higher optimal threshold $\tilde{\tau}^*$, i.e., $\sigma_{t,1} > \sigma_{t,2} \implies \tilde{\tau}_1^* > \tilde{\tau}_2^*$.*

Proof. Recall that $\tilde{\tau}^*$ is obtained as the unique solution to $g\left(\frac{\tilde{\tau}^*}{\sigma_{\tilde{t}}}\right) = 0$, which implies that:

$$\phi\left(\frac{\tilde{\tau}^*}{\sigma_{\tilde{t}}}\right) - \frac{\tilde{\tau}^*}{\sigma_{\tilde{t}}} \cdot \Phi^c\left(\frac{\tilde{\tau}^*}{\sigma_{\tilde{t}}}\right) = \frac{\sigma_{\tilde{t}}}{\sigma_t^2} \cdot c_{rev}.$$

As σ_t increases, the ratio $\frac{\sigma_{\tilde{t}}}{\sigma_t^2} = \frac{\sqrt{\sigma_t^2+\alpha^2\sigma_{ef}^2+(1-\alpha)^2\sigma_{es}^2}}{\sigma_t^2}$ decreases and therefore, the RHS of the above equation decreases. Since we know that the LHS is monotonically decreasing in the ratio $\frac{\tilde{\tau}^*}{\sigma_{\tilde{t}}}$, for the equality to still hold, the ratio should increase. Therefore, when $\sigma_{t,1} > \sigma_{t,2}$, we have:

$$\frac{\tilde{\tau}_1^*}{\sigma_{\tilde{t},1}} > \frac{\tilde{\tau}_2^*}{\sigma_{\tilde{t},2}} \implies \frac{\tilde{\tau}_1^*}{\tilde{\tau}_2^*} > \frac{\sigma_{\tilde{t},1}}{\sigma_{\tilde{t},2}} > 1 \implies \tilde{\tau}_1^* > \tilde{\tau}_2^*.$$

□

C.5 Proof of Lemma 4

We will complete the proof separately for the two different models of disadvantage for group B.

Negatively biased signal mean. We consider that Group B is disadvantaged in the sense that the mean of the distribution of signal $\tilde{s}$ for group B is negatively biased by amount β with respect to Group A. The bias in $\tilde{s}$ translates into a bias β' in the mean of the perceived quality ($\tilde{t}$) distribution for Group B with $\beta' = (1-\alpha)\beta$, i.e., $\tilde{t}$ for Group A follows $\mathcal{N}(0, \sigma_{\tilde{t}}^2)$, while for group B applicants, it follows $\mathcal{N}(-\beta', \sigma_{\tilde{t}}^2)$. We first derive the revised expression for the net expected utility for the principal for hiring an applicant from group B above some threshold $\tilde{\tau}$. We have:

$$\begin{aligned}
v_B(\tilde{\tau}) &= \mathbb{E}\left[t \mid \tilde{t} \geq \tilde{\tau}\right] - \frac{c_{rev}}{\mathbb{P}\left[\tilde{t} \geq \tilde{\tau}\right]} \\
&= \frac{\sigma_t^2}{\sigma_{\tilde{t}}^2} \cdot \mathbb{E}\left[\tilde{t} + \beta' \mid \tilde{t} \geq \tilde{\tau}\right] - \frac{c_{rev}}{\mathbb{P}\left[\tilde{t} \geq \tilde{\tau}\right]} \\
&= \frac{\sigma_t^2}{\sigma_{\tilde{t}}^2} \cdot \beta' + \frac{\sigma_t^2}{\sigma_{\tilde{t}}^2} \cdot \left(-\beta' + \sigma_{\tilde{t}} \cdot H\left(\frac{\tilde{\tau}+\beta'}{\sigma_{\tilde{t}}}\right)\right) - \frac{c_{rev}}{\Phi^c\left(\frac{\tilde{\tau}+\beta'}{\sigma_{\tilde{t}}}\right)} \\
&= \left[\frac{\sigma_t^2}{\sigma_{\tilde{t}}} \cdot H\left(\frac{\tilde{\tau}+\beta'}{\sigma_{\tilde{t}}}\right) - \frac{c_{rev}}{\Phi^c\left(\frac{\tilde{\tau}+\beta'}{\sigma_{\tilde{t}}}\right)}\right]
\end{aligned}$$

Now, suppose, $\tilde{\tau}_A^*$ maximizes $v_A(\tilde{\tau}) = \frac{\sigma_t^2}{\sigma_{\tilde{t}}} \cdot H\left(\frac{\tilde{\tau}}{\sigma_{\tilde{t}}}\right) - \frac{c_{rev}}{\Phi^c\left(\frac{\tilde{\tau}}{\sigma_{\tilde{t}}}\right)}$. This implies that $\tilde{\tau}_A^* - \beta'$ must maximize $v_B(\tilde{\tau})$, i.e., $\tilde{\tau}_B^* = \tilde{\tau}_A^* - \beta'$.
This immediately implies that:

$$v_B(\tilde{\tau}_B^*) = v_B(\tilde{\tau}_A^* - \beta') = v_A(\tilde{\tau}_A^*).$$

This concludes the proof for the first part.

Noisier signal distribution, i.e., $\sigma_{e,A} < \sigma_{e,B}$. The idea for this part of the proof is to use the simplified expression for the principal's optimal utility at equilibrium using Lemma 3. We have:

$$v^* = v(\tilde{\tau}^*) = \mathbb{E}[t \mid \tilde{t} \geq \tilde{\tau}^*] - \frac{c_{rev}}{\mathbb{P}[\tilde{t} \geq \tilde{\tau}^*]} = \frac{\sigma_t^2}{\sigma_{\tilde{t}}} \cdot \frac{\phi(\tilde{\tau}^*/\sigma_{\tilde{t}})}{\Phi^c(\tilde{\tau}^*/\sigma_{\tilde{t}})} - \frac{c_{rev}}{\Phi^c(\tilde{\tau}^*/\sigma_{\tilde{t}})} = (\sigma_t^2) \cdot \frac{1}{\sigma_{\tilde{t}}} \cdot \left(\frac{\tilde{\tau}^*}{\sigma_{\tilde{t}}}\right).$$

Now, recall that $\sigma_{t_A} = \sigma_{t_B}$. But, $\sigma_{e,A} < \sigma_{e,B} \implies \sigma_{\tilde{t}_A} < \sigma_{\tilde{t}_B} \implies \frac{1}{\sigma_{\tilde{t}_A}} > \frac{1}{\sigma_{\tilde{t}_B}}$. Now we will argue that $\frac{\tilde{\tau}_A^*}{\sigma_{\tilde{t}_A}} > \frac{\tilde{\tau}_B^*}{\sigma_{\tilde{t}_B}}$. From Claim 8, we know that $\frac{\tilde{\tau}^*}{\sigma_{\tilde{t}}}$ must satisfy:

$$\phi\left(\frac{\tilde{\tau}^*}{\sigma_{\tilde{t}}}\right) - \frac{\tilde{\tau}^*}{\sigma_{\tilde{t}}} \cdot \Phi^c\left(\frac{\tilde{\tau}^*}{\sigma_{\tilde{t}}}\right) = \frac{\sigma_{\tilde{t}}}{\sigma_t^2} \cdot c_{rev}.$$

For Group B, the RHS of the above equation is larger (due to larger $\sigma_{\tilde{t}}$). Since the LHS is monotonically decreasing in $\frac{\tilde{\tau}^*}{\sigma_{\tilde{t}}}$, it must be that $\frac{\tilde{\tau}_A^*}{\sigma_{\tilde{t}_A}} > \frac{\tilde{\tau}_B^*}{\sigma_{\tilde{t}_B}}$. Putting all parts together, we conclude that $v_A^* > v_B^*$.

C.6 Proof of Theorem 3

Note that the above optimization problem can be decoupled as follows: Pick r_A such that $0 \leq r_A \leq k$ and then pick $0 \leq r_B \leq k - r_A$. Then, we can decompose the constraint as follows:

$$n_{rev}(A) \cdot \mathbb{P}[\tilde{t}_A \geq \tilde{\tau}_A] \leq r_A; \quad n_{rev}(B) \cdot \mathbb{P}[\tilde{t}_B \geq \tilde{\tau}_B] \leq r_B,$$

which leads to the following independent optimization problems for groups A and B.

$$\begin{aligned} \max_{n_{rev}(A), \tilde{\tau}_A} \quad & n_{rev}(A) \cdot \mathbb{P}[\tilde{t}_A \geq \tilde{\tau}_A] \cdot \mathbb{E}[t_A \mid \tilde{t}_A \geq \tilde{\tau}_A] - c_{rev} n_{rev}(A) \quad s.t. \\ & n_{rev}(A) \cdot \mathbb{P}[\tilde{t}_A \geq \tilde{\tau}_A] \leq r_A, n_{rev}(A) \geq 0. \end{aligned}$$

$$\begin{aligned} \max_{n_{rev}(B), \tilde{\tau}_B} \quad & n_{rev}(B) \cdot \mathbb{P}[\tilde{t}_B \geq \tilde{\tau}_B] \cdot \mathbb{E}[t_B \mid \tilde{t}_B \geq \tilde{\tau}_B] - c_{rev} n_{rev}(B) \quad s.t. \\ & n_{rev}(B) \cdot \mathbb{P}[\tilde{t}_B \geq \tilde{\tau}_B] \leq r_B, n_{rev}(B) \geq 0. \end{aligned}$$

We know how to solve either of these problems in isolation. The assumption on c_{rev} guarantees that both problems have positive objective values at optimality. Let the optimal thresholds be given independently as $\tilde{\tau}_A^*$ and $\tilde{\tau}_B^*$. In that case, $\left(\frac{r_A}{\mathbb{P}[\tilde{t}_A \geq \tilde{\tau}_A^*]}, \frac{r_B}{\mathbb{P}[\tilde{t}_B \geq \tilde{\tau}_B^*]}, \tilde{\tau}_A^*, \tilde{\tau}_B^*\right)$ is a feasible solution to Problem 7. Using an identical argument as in the proof of Theorem 2, we can show that at optimality, $r_A + r_B = k$. Thus, the problem reduces to choosing r_A so that it maximizes the objective in 7. Therefore, we have:

$$\max_{r_A \leq k} \quad r_A \cdot \mathbb{E}[t_A \mid \tilde{t}_A \geq \tilde{\tau}_A^*] + (k - r_A) \cdot \mathbb{E}[t_B \mid \tilde{t}_B \geq \tilde{\tau}_B^*] - c_{rev} \cdot \left(\frac{r_A}{\mathbb{P}[\tilde{t}_A \geq \tilde{\tau}_A^*]} + \frac{k - r_A}{\mathbb{P}[\tilde{t}_B \geq \tilde{\tau}_B^*]}\right),$$

which by re-arranging terms, we can rewrite as follows:

$$\begin{aligned} \max_{r_A \leq k} \quad r_A \cdot & \underbrace{\left[\left(\mathbb{E}[t_A \mid \tilde{t}_A \geq \tilde{\tau}_A^*] - \frac{c_{rev}}{\Phi^c(\tilde{\tau}_A^*/\sigma_{\tilde{t}_A})}\right) - \left(\mathbb{E}[t_B \mid \tilde{t}_B \geq \tilde{\tau}_B^*] - \frac{c_{rev}}{\Phi^c(\tilde{\tau}_B^*/\sigma_{\tilde{t}_B})}\right)\right]}_{(v_A^* - v_B^*)} \\ & + \underbrace{k \cdot \mathbb{E}[t_B \mid \tilde{t}_B \geq \tilde{\tau}_B^*] - c_{rev} \cdot \frac{k}{\mathbb{P}[\tilde{t}_b \geq \tilde{\tau}_B^*]}}_{\text{terms independent of } r_A} \end{aligned}$$

From Lemma 4, we know that $v_A^* > v_B^*$ which immediately shows that co-efficient of r_A in the expression above is positive. Therefore, $r_A^* = k$ which implies that $r_B^* = 0$ or equivalently, $n_{rev}(B)^* = 0$. This concludes the proof.

D Theoretical Extension: s and f are correlated

We now consider the following extension: The metrics s and f chosen by the principal to measure applicant quality are no longer independent, rather they are jointly Gaussian with correlation coefficient r. Note that $r \in [-1, 1]$. The case $r = 0$ reduces to our current case where s and f are independent.

When s and f are correlated, f can be decomposed into a term that is perfectly correlated with s and an independent term. Therefore, the actual weight placed by the principal on s when measuring quality is different from $(1 - \alpha)$ due to contributions from f (it may be larger or smaller depending on the sign of the correlation between s and f). This means that we only need to closely re-inspect those results in Sections 3 and 4 where we observe trends with α, in particular, Lemma 1, Corollary 2 and Claim 4. All our other results will remain completely unchanged, because they deal directly with the distributions of s and $\tilde{s}$ as in Section 3 (which remains unaffected by the correlation) or the distributions of t and $\tilde{t}$ as in Section 4 (where the effect of correlation is limited to the variances of t and $\tilde{t}$).

D.1 Changes in Section 3

Generalized form of principal utility in Lemma 1. Here, we derive the generalized form of the principal's utility per hired applicant under delegation when s and f have correlation coefficient r. Recall that:

$$\mathbb{E}[t \mid \tilde{s} \geq \tau_1] = \alpha \cdot \mathbb{E}[f \mid \tilde{s} \geq \tau_1] + (1-\alpha) \cdot \mathbb{E}[s \mid \tilde{s} \geq \tau_1].$$

We already know what the second term evaluates to (Lemma 1). However, the first term no longer evaluates to 0 because s and f are correlated. Firstly, if $Corr(s, f) = r$, then $Cov(\tilde{s}, f) = Cov(s, f) = r\sigma_s\sigma_f$. Now,

$$f \mid \tilde{s} = \tilde{s} \sim \mathcal{N}\left(0 + \frac{r\sigma_s\sigma_f}{\sigma_{\tilde{s}}^2} \cdot (\tilde{s} - 0), \sigma_f^2 - \frac{r^2\sigma_s^2\sigma_f^2}{\sigma_{\tilde{s}}^2}\right).$$

Therefore, $\mathbb{E}[f \mid \tilde{s}] = \left(\frac{r\sigma_s\sigma_f}{\sigma_{\tilde{s}}^2}\right) \cdot \tilde{s}$. This implies:

$$\mathbb{E}[f \mid \tilde{s} \geq \tau_1] = \mathbb{E}\left[\mathbb{E}[f \mid \tilde{s}] \mid \tilde{s} \geq \tau_1\right] = \frac{r\sigma_s\sigma_f}{\sigma_{\tilde{s}}^2}\mathbb{E}[\tilde{s} \mid \tilde{s} \geq \tau_1] = \frac{r\sigma_s\sigma_f}{\sigma_{\tilde{s}}} \cdot H\left(\frac{\tau_1}{\sigma_{\tilde{s}}}\right).$$

Putting everything together, we obtain:

$$\mathbb{E}[t \mid \tilde{s} \geq \tau_1] = \frac{\sigma_s}{\sigma_{\tilde{s}}}\left[\alpha r\sigma_f + (1-\alpha)\sigma_s\right] \cdot H\left(\frac{\tau_1}{\sigma_{\tilde{s}}}\right) = \frac{\sigma_s}{\sigma_{\tilde{s}}}\left[\sigma_s - \alpha(\sigma_s - r\sigma_f)\right] \cdot H\left(\frac{\tau_1}{\sigma_{\tilde{s}}}\right).$$

Note that when $r = 0$, the expression reduces to the expression in Lemma 1. Further, it is still monotonic in α (Corollary 2), however, whether it is monotonically increasing or decreasing depends on the relative magnitudes of σ_s, σ_f and the degree and sign of the correlation coefficient r. **All other results in Section 3 (including the fairness results) remain unchanged.**

D.2 Changes in Section 4

As described earlier, the only result we need to inspect in Section 4 is Claim 4 which captures the non-monotonicity of the principal's utility per hired student under non-delegation as a function of α.

Recall that the principal's optimal expected utility per hired applicant is given by $\frac{\sigma_t^2}{\sigma_{\tilde{t}}^2}\tilde{\tau}^*$. We previously argued that this utility is monotonically increasing in σ_t — that result remains unchanged. We just need to check if the non-monotonicity of σ_t with α holds even when s and f are correlated. If $r \geq 0$ is the correlation co-efficient between s and f, we have:

$$\begin{aligned}\sigma_t^2 &= Var(\alpha f) + Var((1-\alpha)s) + 2Cov(\alpha f, (1-\alpha)s) \\ &= \alpha^2\sigma_f^2 + (1-\alpha)^2\sigma_s^2 + 2\alpha(1-\alpha)\rho\sigma_s\sigma_f.\end{aligned}$$

Therefore, σ_t clearly is quadratic in α which implies that in general, it is also non-monotonic in α. Putting everything together, the principal's average utility per hired applicant also continues to be non-monotonic in α. Therefore, **introducing correlation between s and f does not change any of our results in Section 4**.

D.3 Changes in Section 5

None of the broad insights change in terms of the principal's delegation decision even when s and f are correlated. We continue to have scenarios where the delegation decision may be non-monotonic in parameter α.

E Proofs of Supplementary Results

E.1 Proof of Claim 7

We know that $X \sim \mathcal{N}(\mu_X, \sigma_x^2)$. Now,

$$\begin{aligned}
\mathbb{E}[X \mid X \geq a] &= \int_a^\infty u \cdot \mathbb{P}[X = u \mid X \geq a] du \\
&= \int_a^\infty u \cdot \frac{\mathbb{P}[X = u, X \geq a]}{\mathbb{P}[X \geq a]} du \\
&= \frac{1}{\mathbb{P}[X \geq a]} \int_a^\infty u \cdot \mathbb{P}[X = u] du \\
&= \frac{1}{\Phi^c\left(\frac{a-\mu_X}{\sigma_X}\right)} \int_a^\infty u \cdot \frac{1}{\sigma_X \sqrt{2\pi}} \cdot \exp\left(-\frac{(u-\mu_X)^2}{2\sigma_X^2}\right) du.
\end{aligned}$$

Now we do a variable substitution. Define $w = \frac{u-\mu_X}{\sigma_X}$. Then, we can rewrite as follows:

$$\begin{aligned}
\mathbb{E}[X \mid X \geq a] &= \frac{1}{\Phi^c\left(\frac{a-\mu_X}{\sigma_X}\right)} \int_{(a-\mu_X)/\sigma_X}^\infty (\mu_X + \sigma_X \cdot w) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{w^2}{2}\right) dw \\
&= \frac{1}{\Phi^c\left(\frac{a-\mu_X}{\sigma_X}\right)} \left[\mu_X \int_{(a-\mu_X)/\sigma_X}^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{w^2}{2}\right) dw + \sigma_X \int_{(a-\mu_X)/\sigma_X}^\infty \frac{w}{\sqrt{2\pi}} \exp\left(-\frac{w^2}{2}\right) dw\right] \\
&= \frac{1}{\Phi^c\left(\frac{a-\mu_X}{\sigma_X}\right)} \left[\mu_X \cdot \Phi^c\left(\frac{a-\mu_X}{\sigma_X}\right) + \sigma_X \int_{(a-\mu_X)/\sigma_X}^\infty w\phi(w) dw\right] \\
&= \mu_X + \frac{\sigma_X}{\Phi^c\left(\frac{a-\mu_X}{\sigma_X}\right)} \int_{(a-\mu_X)/\sigma_X}^\infty w\phi(w) dw.
\end{aligned}$$

Note that the proof is completed if we can just show that the integrand above equals $\phi\left(\frac{a-\mu_X}{\sigma_X}\right)$.

$$\begin{aligned}
\int_{(a-\mu_X)/\sigma_X}^\infty w\phi(w) dw &= \int_{(a-\mu_X)/\sigma_X}^\infty -\phi'(w) dw \quad (\text{since } \phi'(w) = -w\phi(w)) \\
&= \int_\infty^{(a-\mu_X)/\sigma_X} \phi'(w) dw \\
&= \phi\left(\frac{a-\mu_X}{\sigma_X}\right) - \phi(0) \\
&= \phi\left(\frac{a-\mu_X}{\sigma_X}\right).
\end{aligned}$$

E.2 Monotonicity of the standard normal hazard rate function $H(x)$

Here, we show that hazard rate function of the standard normal random variable, given by $H(x) = \frac{\phi(x)}{\Phi^c(x)}$ is monotonically increasing in x. Observe that $H(x) \geq 0$ trivially for all x. First, we will show that $H(x) \geq x$

for all x. It suffices to show for $x > 0$ (since it holds trivially for $x \leq 0$).

$$\begin{aligned}
1 - \Phi(x) &= \int_{u=x}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du \\
&\leq \int_{u=x}^{\infty} \frac{1}{\sqrt{2\pi}} \frac{u}{x} \exp\left(-\frac{u^2}{2}\right) du \quad (\text{since } \frac{u}{x} \geq 1) \\
&= \frac{1}{x\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \\
&= \frac{\phi(x)}{x}.
\end{aligned}$$

Therefore, $H(x) = \frac{\phi(x)}{1-\Phi(x)} \geq x$. Now,

$$\begin{aligned}
H'(x) &= \frac{\Phi^c(x)\phi'(x) + (\phi(x))^2}{(\Phi^c(x))^2} \\
&= \frac{-x\phi(x)\Phi^c(x) + (\phi(x))^2}{(\Phi^c(x))^2} \quad (\text{substituting } \phi'(x) = -x\phi(x)) \\
&= -xH(x) + (H(x))^2 \\
&= H(x)(H(x) - x) \geq 0.
\end{aligned}$$

The last equality follows from the fact that $H(x) \geq 0$ and $H(x) \geq x$. This concludes the proof that $H(x)$ is increasing.

F Additional Figures

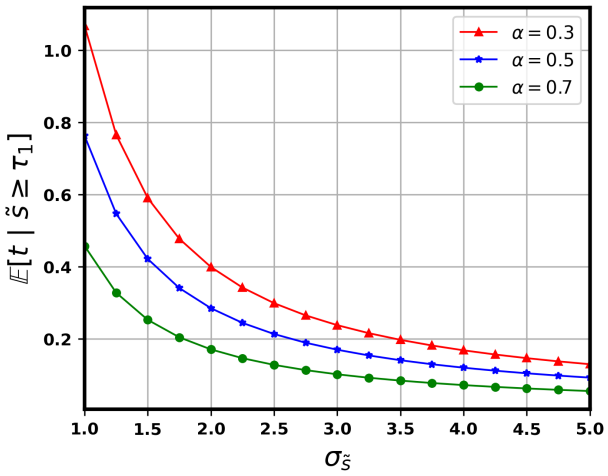
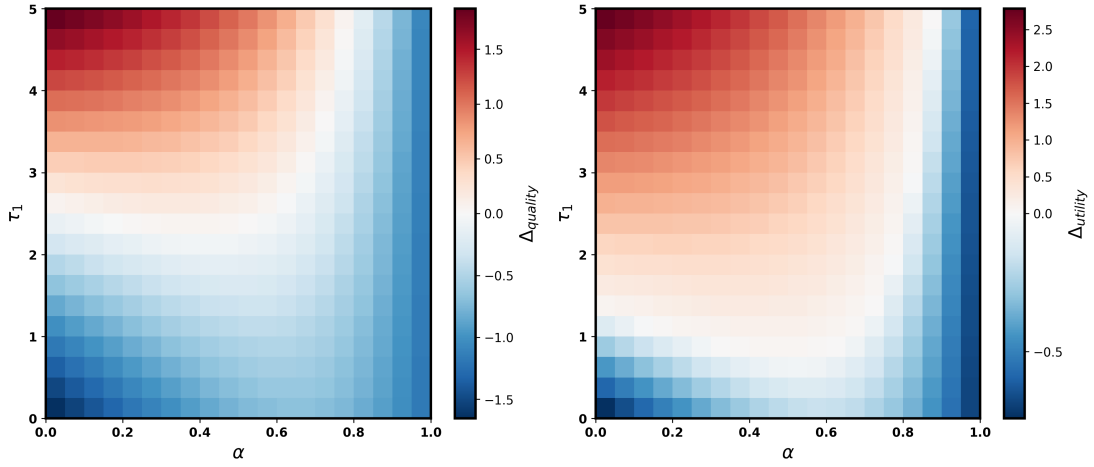


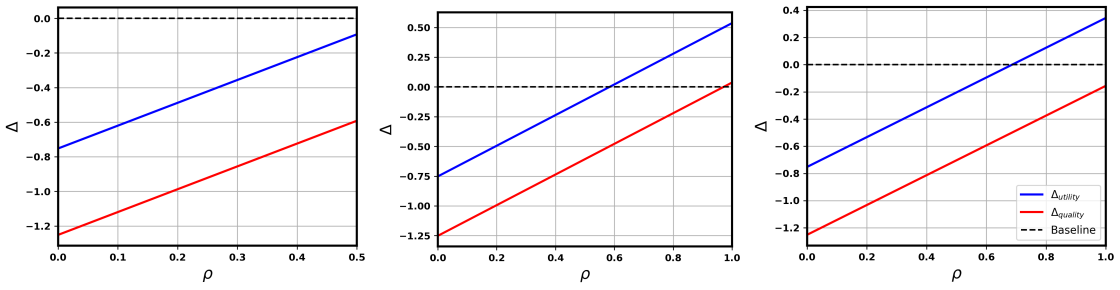
Figure 5: We plot the expected quality of a selected applicant (when the agent uses a selection threshold of τ_1) as a function of the variance $\sigma_{\tilde{s}}$ of the noisy signal $\tilde{s}$. As $\sigma_{\tilde{s}}$ increases, the expected quality decays monotonically for all α.

F.1 Factors Affecting Principal's Delegation Decision



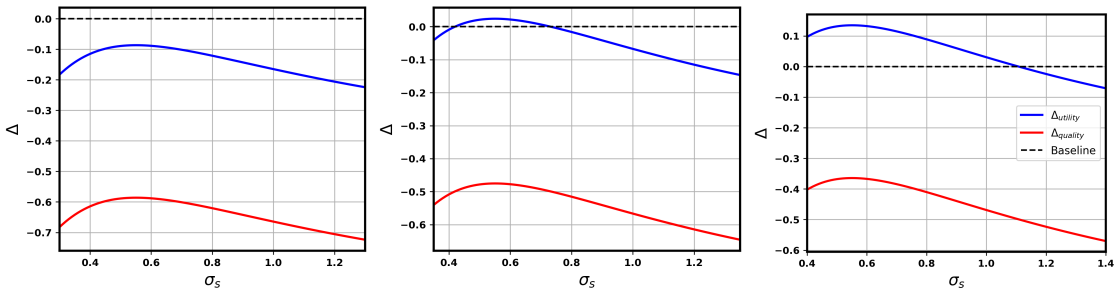
(a) (b)

Figure 6: We plot a heatmap of values of $\Delta_{quality}$ (left) and $\Delta_{utility}$ (right) over different combinations of the principal's weight α over the interval $[0, 1]$ and the agent's threshold τ_1 over the interval $[0, 5]$. Each heatmap consists of 400 grid points with each grid of the size 0.05×0.25. Values are evaluated at the centre of each grid. The red regions indicate locations where $\Delta > 0$ and blue regions indicate locations where $\Delta < 0$. The white regions indicate the transition boundary where $\Delta \approx 0$. The heatmaps validate our discussion about where delegation is beneficial for the principal (Section 5).



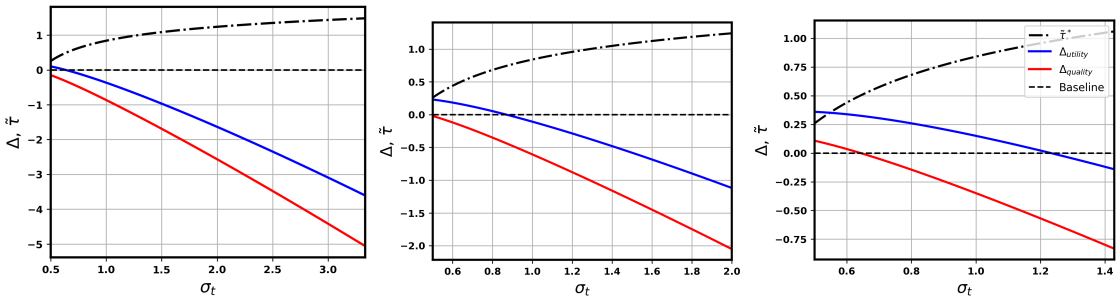
(a) $\sigma_s/\sigma_t = 0.5$ (b) $\sigma_s/\sigma_t = 1$ (c) $\sigma_s/\sigma_t = 2$

Figure 7: In this set of plots, we isolate the effect of the correlation ρ between the signals s and t and how it affects the delegation decision of the principal. In each of these plots, we fix the σ_s and σ_t and vary ρ within the permissible range $\min(1, \frac{\sigma_s}{\sigma_t})$ allowed by our model. Parameter combination: $c_{rev} = 0.1$, $\tau_1 = 1$, $\sigma_{es} = \sigma_{et} = 0.5$, $\sigma_t = 1$. We pick σ_s from the set: $\{0.5, 1, 2\}$. The principal's optimal threshold across all 3 plots comes out to be 0.84. We observe that Δ always grows linearly with growing ρ if σ_s and σ_t are fixed. This always leads to monotonic delegation decisions — higher correlation means that the principal's and agent's metric are closely aligned and the principal does not gain by doing their own evaluation, so it is better to delegate.



(a) $\rho = 0.30$ (b) $\rho = 0.35$ (c) $\rho = 0.40$

Figure 8: In this set of plots, we isolate the effect of changing the variance of the agent's metric σ_s and how it affects the delegation decision of the principal. In each of these plots, we fix the σ_t and ρ and vary σ_s within the permissible range allowed by our model. Parameter combination: $c_{rev} = 0.1$, $\tau_1 = 2$, $\sigma_{es} = \sigma_{et} = 0.5$, $\sigma_t = 1$. We pick ρ from the set: $\{0.30, 0.35, 0.40\}$. The principal's optimal threshold across all 3 plots comes out to be 0.84. In this case, we observe that the principal's delegation decision can indeed be non-monotonic, as we see for $\rho = 0.35$. This is driven by the fact that the principal's utility from a hired student when they delegate is non-monotonic in σ_s. At low levels of σ_s, the principal's utility actually improves with increasing σ_s — because the distribution of s has very light tails initially, those who cross τ_1 are some of the top candidates in the pool in terms of ability. However, once σ_s grows a bit larger, the distribution tails become heavier, so the screening based on τ_1 becomes less informative gradually. This explains why the principal's utility starts decreasing.



(a) $\rho = 0.30$ (b) $\rho = 0.50$ (c) $\rho = 0.70$

Figure 9: In this set of plots, we isolate the effect of changing the variance of the principal's metric σ_t and how it affects the delegation decision of the principal. In each of these plots, we fix the σ_s and ρ and vary σ_t within the permissible range allowed by our model. Parameter combination: $c_{rev} = 0.1$, $\tau_1 = 1$, $\sigma_{es} = \sigma_{et} = 0.5$, $\sigma_s = 1$. We pick ρ from the set: $\{0.30, 0.50, 0.70\}$. Firstly, the principal's optimal threshold $\tilde{\tau}^*$ is monotonically increasing in σ_t (as predicted by theory). But the principal's delegation decision is always monotonic in σ_t. This follows because increasing σ_t increases the diversity of quality in the main pool, which means that the principal can find good quality applicants quietly. So, higher σ_t incentivizes the principal to retain decision authority and not delegate.

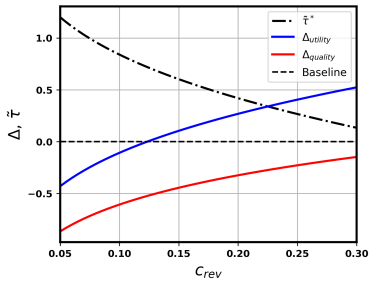


Figure 10: In this figure, we isolate the effect of the principal's cost c_{rev} on their delegation decision. Parameter combination: $\tau_1 = 1$, $\sigma_{es} = \sigma_{et} = 0.5$, $\sigma_s = \sigma_t = 1$, $\rho = 0.5$. Firstly, the principal's optimal threshold $\tilde{\tau}^*$ is monotonically decreasing in c_{rev} (as predicted by theory). Also, the principal's delegation decision is always monotonic in c_{rev}. Once the principal's cost grows too high, it is always better to delegate.