Selecting Applicants

Alex Frankel*

February 24, 2019

Abstract

A firm selects applicants to hire based on hard information, such as a test result, and soft information, such as a manager's evaluation of an interview. The contract that the firm offers to the manager can be thought of as a restriction on acceptance rates as a function of test results. I characterize optimal acceptance rate functions both when the firm knows the manager's mix of information and biases and when the firm is uncertain. These contracts may admit a simple implementation in which the manager can accept any set of applicants with a sufficiently high average test score.

KEYWORDS: Principal-agent contracting, Delegation, Hiring

1 Introduction

A firm making hiring decisions gets both "hard" and "soft" information about the quality of its job applicants. Hard information is directly observable to the firm. For instance, the applicants' education histories and years of experience at previous jobs are listed on their CVs, and the firm sees the applicants' results on any pre-employment tests that it administers. Soft information, by contrast, is reported to the firm by an agent: a hiring manager interviews each applicant and subjectively judges his or her fit for the position. Similarly, in college admissions, there is hard information on applicant quality in the form of grades and test scores, plus soft information from an admissions officer's reading of the essays and recommendation letters. A bank deciding which loan applications to approve has access to hard credit scores as well as the soft evaluations of a loan officer.

*Chicago Booth; afrankel@chicagobooth.edu. I thank Manuel Amador, Kyle Bagwell, Ben Brooks, Ging Cee Ng, Lisa Kahn, Danielle Li, Niko Matouschek, Mallesh Pai, Daniel Rappoport, Phil Reny, Jesse Shapiro, Bob Topel, Glen Weyl, and Michael Whinston for their comments.

In each of these cases the agent observing the soft information may have preferences that are only imperfectly aligned with those of the organization. The hiring manager may be idiosyncratically biased in favor of or against certain applicants – she likes the ones who come off as friendly during the interview. Or the manager may have a more systematic bias. She is skilled at evaluating social skills, say, and at the same time overweights the importance of social skills on the job.

If the manager were given full discretion to hire any applicants she wanted, her choices would be distorted by these biases. Instead the firm might require the manager to hire those with the most favorable hard information. This no-discretion policy would inevitably screen out some "diamonds in the rough," high quality candidates who showed their worth only on the softer measures. But, as pointed out by Hoffman et al. (2016), no discretion would improve on full discretion if the manager's biases were sufficiently strong.

Of course, the firm is not limited to the two extreme policies of no discretion or full discretion. The firm can get some input from the manager while still using hard information to constrain her decisions. In this paper I search for the optimal such policy, posed as a contract offered by a principal to an agent.

Spelling out some key features of the model, there is a large number (continuum) of applicants, of whom some share will be accepted. For each applicant there is public or hard information about quality as well as a private or soft signal observed only by an agent. The agent also has some bias in favor of or against each applicant. The principal writes a contract to determine which applicants the agent may accept. The principal's objective is to maximize the average quality of the selected applicants, whereas the agent cares about quality plus bias. As in what is called the delegation literature, the only outcome of the contract is the determination of the decisions in question, i.e., which applicants are accepted. There are no transfer payments.

I begin my analysis by supposing that the firm knows the manager's "type" – the distribution of her information and bias across applicants. Here, the optimal policy takes the form of a specified acceptance rate at each realization of hard information, e.g., at each test result. The manager chooses which applicants to hire subject to this acceptance rate. I provide a general approach for finding the optimal acceptance rate function, essentially by equalizing the quality of the marginal accepted applicant across test results. I then apply the results to a benchmark normal specification: normal distribution of applicant quality, one-dimensional normal signals of quality from both hard and soft sources, and normally distributed idiosyncratic biases. Under the normal specification, the acceptance rate should follow a normal CDF function – an S-curve, in which a higher share of applicants are accepted at higher test results. A manager with either stronger biases or less information faces a steeper acceptance rate function, corresponding to a contract with less discretion: her hiring depends more on the test and less on her personal judgment.

There is also an alternative implementation of these optimal contracts. Realizations of hard information are first mapped into a one-dimensional score. Then the contract specifies that the manager may accept any applicants she wants, subject to fixing the average score of those who are accepted. (Equivalently, the contract specifies a minimum average score, in which case the manager selects applicants so that the floor is binding.) Under the normal specification, the score can be set to be equal to the test result: the manager has the freedom to accept any set of applicants with an appropriate average test result. This contract offers less discretion to the agent when the required average is higher.

Next, I consider the possibility that the firm is uncertain about the manager's type. Different hiring managers may be better or worse at judging applicant quality, and may also have preferences that are more or less aligned with those of the firm. The firm can screen across types by allowing the manager to select from a menu of acceptance rate functions. Under the normal specification, I find conditions under which the firm again has a simple optimal policy. The manager can select applicants according to any normal CDF acceptance rate that is sufficiently steep. Alternatively, the firm specifies a minimum average test result. The floor is binding for more informed or more biased managers, while less informed and less biased ones select applicant pools with average test results above this floor.

The bulk of the work on delegation involves a single one-dimensional decision to be made. See, for example, Holmström (1977, 1984) and Melumad and Shibano (1991) for early work, or Alonso and Matouschek (2008) and Amador and Bagwell (2013) more recently. Papers on delegation or cheap talk – commitment or no commitment – over multiple decisions include Chakraborty and Harbaugh (2007), Frankel (2014), and Frankel (2016). Indeed, if in the current paper there were no hard information available to guide decisions, then the problem would reduce to a straightforward delegation problem over multiple binary decisions; the principal would essentially only be able to give the agent full discretion to accept her favorite applicants or to randomize acceptances. What is novel in the current paper, both in terms of the tradeoffs it generates and in the contracting levers it makes available, is the existence of distinct realizations of hard information across decisions.

Heuristically, one can reinterpret the problem of selecting applicants as a single-decision – though not necessarily one-dimensional – delegation problem. The acceptance rate function takes the role of the action, and the agent's information and bias type (which determines players' preferences over actions) takes the role of the state of the world. Section 4, which constitutes the main technical contribution of the paper, exploits this connection in solving for optimal screening contracts when the principal does not know the agent's type. In particular, I show that, under the normal specification, there is in fact a formal translation of the principal's problem into a one-dimensional delegation problem. The one-dimensional action corresponds to the average test result of the hired applicants, or equivalently the steepness of a normal CDF acceptance rate function. After performing this translation, I can solve for the optimal contract as a floor on this action by applying the one-dimensional delegation results of Amador et al. (2018).

The problem of combining information from hard and soft sources is becoming increasingly relevant as "big data" and the spread of IT supplement traditional subjective evaluations with newly available, or newly quantifiable, hard information. According to the Wall Street Journal (2015), for instance, the number of US employers using pre-employment tests rose from 26% in 2001 to 57% in 2013.[1] Through a mix of theory and data, Autor and Scarborough (2008) and Hoffman et al. (2016) shed light on how pre-employment tests have affected firm hiring, while Einav et al. (2013) looks at the impact of automated credit scoring in the market for consumer loans. In particular, Hoffman et al. (2016) addresses the question of how much discretion to grant to a potentially biased agent. That paper compares full discretion, the policy used by the firm in their data set, to a hypothetical policy of no discretion.

In related theoretical work, Che et al. (2013) looks at the role of hard and soft information in hiring, where hard information is modeled via asymmetric priors on

[1] The article describes these tests as follows:

> Tests in the past gauged only a few broad personality traits. But statistical modeling and better computing power now give employers a choice of customized assessments that, in a single test, can appraise everything from technical and communication skills to personality and whether a candidate is a good match with a workplace's culture—even compatibility with a particular work team.

applicants' quality.[2] Beyond the fact that Che et al. (2013) looks at hiring a single worker instead of many, there is a key difference in the modeling of the agent's bias. Their misalignment is over *how many* candidates to hire, whereas in this paper it is over *which* candidates to hire. In particular, the agent in that paper has the same ranking of candidates as does the principal, but has a bias towards hiring: one of the candidates will be hired, or none will be. I fix the (large) number of candidates to be hired, but take the agent's preferences over candidates to be imperfectly correlated with those of the principal.[3]

I now move into the analysis. Proofs can be found in Appendix G.

2 The model

2.1 Players, payoffs, and information

There is a firm (principal), a manager (agent), and a mass 1 of ex ante identical job applicants. An exogenous fraction $k \in (0, 1)$ of the applicants will be hired (accepted). The firm and manager are the two players in the game; applicants are nonstrategic. The firm will specify rules determining the process by which the manager makes hiring decisions.

Each applicant is associated with a vector of four characteristics: *quality* $Q \in \mathbb{R}$, *hard information* or "test result" $T \in \mathcal{T}$, *soft information* or "private signal" $S \in \mathcal{S}$, and *bias* $B \in \mathbb{R}$. I label generic realizations of Q, T, S, and B by the lowercase q, t, s, and b.

The quality Q indicates the firm's marginal utility of hiring an applicant. The manager's marginal utility is quality plus bias, $Q + B$. That is, the bias for a given applicant is the difference between the manager's marginal utility and the firm's. The principal's and agent's realized payoffs will be their average marginal utilities across the hired applicants.

[2] Some other models of applicant selection include hard but not soft information. Chan and Eyster (2003) study the effects of banning affirmative action in college admissions. In their model, colleges make up-or-down admissions decisions on the basis of test scores and, possibly, minority status. Alonso (2018) studies how much information firms should gather information about job applicants' fit when application decisions are endogenous.

[3] Armstrong and Vickers (2010) and Nocke and Whinston (2013) consider a different sort of mechanism design problem relating to the acceptance or rejection of a single proposed candidate. In their work, the agent's private information is over the set of candidates that may be proposed.

Quality is never directly observed by the players. It can only be inferred from the two pieces of information, T and S. The test result T for an applicant is "hard" in the sense that it is publicly observed by both the firm and the manager and it can be contracted upon. The private signal S is "soft," as is the manager's bias B: they are privately observed by the manager, are noncontractible, and can never be externally verified or audited. The realizations of T, S, and B for all applicants are observed automatically by the appropriate parties. In particular, the agent observes her private signals without undertaking any costly investment of time or effort.[4]

Denote the distribution of applicant quality Q in the population by F_Q, the distribution of the test result T conditional on Q by $F_{T|Q}$, and the distribution of the private signal S conditional on Q and T by $F_{S|Q,T}$. We see that T and S are informative about quality in that their distributions may depend on Q. In examples, I often take the signals T and S to be real-valued. In general, though, their realization spaces $\mathcal{T}$ and $\mathcal{S}$ need not be ordered, and may be highly dimensional.

The final applicant characteristic is the bias, B. Assume that the distribution of B conditional on T and S is independent of Q, and thus that its conditional distribution can be denoted by $F_{B|Q,T,S} = F_{B|T,S}$. In other words – and without loss of generality – the agent's bias contains no information on quality beyond what is captured by her soft information. (Any information in B can be assumed to be included in S as well.)

An *acceptance rule* $\chi(t, s, b)$ describes the probability that an applicant with test result $T = t$, private signal $S = s$, and bias $B = b$ – the three characteristics ever observed by some player – is hired. Formally, $\chi : \mathcal{T} \times \mathcal{S} \times \mathbb{R} \to [0, 1]$ is a measurable function satisfying the budget constraint $\mathbb{E}[\chi(T, S, B)] = k$.[5] The contracting game that determines the equilibrium acceptance rule is introduced in the next section.

Under a given acceptance rule, we can define a random variable Hired $\in \{0, 1\}$ that describes whether a randomly drawn applicant is hired (1) or not (0). Conditional on Q, T, S, and B, Hired takes value 1 with probability $\chi(T, S, B)$ and 0 otherwise.

[4]If the agent were required to put in a costly investment in order to observe her private signal, realizations of hard information might be used as a first screen to decide which applicants to evaluate privately. The Wall Street Journal (2014) reports that at many companies, job applicants who do poorly on pre-employment tests will not have their resumes looked at or will not get interviews.

[5]As discussed below, I consider the possibility that the principal may be uncertain about the joint distribution of Q, T, S, B – specifically, he may have priors over $F_{S|Q,T}$ and $F_{B|T,S}$ without knowing their true realizations. Unless explicitly indicated otherwise, the expectation operator $\mathbb{E}$ always refers to expectations taken under the true distributions. Hence, from an uncertain principal's perspective, the budget constraint of hiring k applicants is "ex post" rather than "ex ante."

The principal's and agent's realized payoffs in the game, V_P and V_A, are defined as their respective average marginal utilities over the hired applicants:[6]

$$V_P \equiv \mathbb{E}[Q|\text{Hired} = 1] = \frac{1}{k}\mathbb{E}[\chi(T, S, B)Q]$$
$$V_A \equiv \mathbb{E}[Q + B|\text{Hired} = 1] = \frac{1}{k}\mathbb{E}[\chi(T, S, B)(Q + B)].$$

Players make decisions in order to maximize expected payoffs.

We see that the model primitives consist of one scalar parameter and four distribution functions: the share k of applicants hired, the quality distribution F_Q, the public and private signal distributions $F_{T|Q}$ and $F_{S|Q,T}$, and the bias distribution $F_{B|T,S}$. I call the first three objects k, F_Q, and $F_{T|Q}$ "principal fundamentals," as they are properties of the applicant pool and how it is judged by the firm. The two distributions $F_{S|Q,T}$ and $F_{B|T,S}$ then describe the extent of the manager's information and biases. I call $F_{S|Q,T}$ and $F_{B|T,S}$ the agent's *type*.

It may be that in the universe of potential hiring managers, some are better than others at evaluating job applicants, and also some care more about hiring the right applicants for the firm rather than the ones they personally like. That is, the potential managers may not all have the same distributions of private information $F_{S|Q,T}$ and bias $F_{B|T,S}$. In the upcoming analysis, I will separately analyze cases where the agent's type is known to the principal (Section 3) and where the principal has a prior over the agent's type but does not know its realization (Section 4). Regardless of whether the agent's type is known to the principal, the agent does know her own type; in particular, the agent knows $F_{S|Q,T}$ and hence knows how to interpret realizations of her private signal S.

While the principal may be uncertain about the agent's type, I assume throughout the paper that there is common knowledge over the principal fundamentals.

Assumption 1. The share of applicants to be hired k, the quality distribution F_Q, and the test score distribution $F_{T|Q}$ are commonly known at the start of the game.

Recall that there is a continuum of applicants, where F_Q and $F_{T|Q}$ are the distributions of Q and $T|Q$. So common knowledge of F_Q establishes that there is no

[6]For any random variable $Y(Q, T, S, B)$ defined as a function of applicant characteristics, the average of Y over hired applicants can be equivalently written as $\mathbb{E}[Y(Q, T, S, B)|\text{Hired} = 1]$ or as $\frac{1}{k}\mathbb{E}[\chi(T, S, B) \cdot Y(Q, T, S, B)]$.

aggregate uncertainty over the distribution of quality in the applicant pool. Common knowledge of $F_{T|Q}$ further implies that there is no aggregate uncertainty over the distribution of test results, nor over the conditional distribution of applicant quality at each test result.

2.2 Contracting

The principal writes a contract that specifies the rules by which the agent selects applicants. The principal has the power to commit to accept any applicant set that the agent selects, and the agent is assumed to participate in whatever contract the principal writes. The only outcome of this contracting relationship is the determination of which applicants are accepted. That is, this model considers a "delegation" framework in which there are no monetary transfers (or other extrinsic incentives) that condition on the agent's behavior or on the realized quality distribution of the accepted applicants.

The contracting mechanism has access to every applicant's public test result. However, it cannot directly condition on the agent's type or the realizations of the agent's private signals and biases for the different applicants. Accordingly, the mechanism asks the agent to make some reports. As discussed below, it is without loss of generality to consider a mechanism in which the agent simply reports the probability with which each applicant is to be accepted.[7] The contract itself is a restriction on the joint distribution of observable test results and the reported acceptance probabilities.

The formal timing of the game is as follows.

(1) The principal gives the agent a contract and the agent's type is realized.

(2) All applicants' test results are publicly observed. At the same time, the agent observes her private signals and biases for each applicant.

(3) The agent reports a probability of acceptance for each applicant, subject to the contractual restrictions.

(4) Applicants are hired according to the stated probabilities.

At step (1) the principal chooses a contract to maximize his subjective expectation of $V_P = \mathbb{E}[Q|\text{Hired} = 1]$, taking into account his beliefs about the agent's type and

[7] I omit mention of the correlation structure over probabilistic acceptances because correlation is payoff-irrelevant to both players. It does not affect the expected marginal utility of hired applicants.

behavior when predicting who will be hired (i.e., what the acceptance rule will be). The agent then makes reports at step (3) to maximize $V_A = \mathbb{E}[Q + B|\text{Hired} = 1]$. Formally, in making this report, the agent chooses a measurable function from realizations of T, S, B into distributions over acceptance probabilities, which induces an acceptance rule χ.

The contract must impose the budget constraint that the agent accepts a share k of the applicants. Beyond that, there may be arbitrary restrictions on the joint distribution of test results and acceptance probabilities. For instance, the contract could include a requirement that all applicants at a given test result $T = t$ are accepted; that anywhere between 10% and 50% of applicants at t are accepted; or that the maximum share of applicants accepted at t depends on the share accepted at another test result t'. The contract can also allow for (or require) stochastic acceptances, in which the agent reports that some applicants are accepted with interior probabilities.

Say that the contract is *deterministic* if it only allows the agent to report acceptance probabilities of either 0 or 1, i.e., rejections or acceptances.[8] Deterministic contracts can be described in a fairly simple manner. First, note that the agent's reports determine an *acceptance rate* at each test result, where I use $\alpha : \mathcal{T} \to [0, 1]$ to denote a generic such acceptance rate function:

$$\alpha(t) = \mathbb{E}[\text{Hired}|T = t].$$

Any two deterministic reports that induce the same acceptance rate function necessarily have the same joint distribution over test results and acceptance probabilities. So any contract either allows both of them or forbids both of them. Hence, deterministic contracts can be fully summarized by the menu of acceptance rate functions from which the agent may choose.[9]

Now let us return to the issue of why it is sufficient to consider contracts of the form above, with steps (1) - (4), rather than working with direct revelation contracts.

[8]Even in a deterministic contract the agent can "mix" and accept some share of applicants at a given realization of (T, S, B) while rejecting the others. This might be necessary if at $T = t$ there is an atom of probability on a particular realization of $(S, B) = (s, b)$, and the contract prohibits the agent from either accepting all of the applicants at the atom or from rejecting all of them. I will rule out the possibility of such atoms in Assumption 2 below.

[9]Stochastic contracts may be more complicated. For instance, two distinct ways to induce the same constant acceptance rate function of $\alpha(T) = 1/2$ are for the agent to deterministically accept one half of applicants at each test result, and to accept each applicant with probability 1/2. A deterministic contract would impose the former rather than the latter.

A direct revelation contract would differ from the above in two ways. First, in between the agent's type realization at step (1) and the arrival of applicants at step (2) – call this step (1.5) – the agent would be asked to report her type. (This step could be omitted if the agent's type were common knowledge.) Second, at messaging step (3) she would report her private signal and bias realization for each applicant instead of an acceptance probability.

The fact that the agent's direct reports at step (3) can be replaced with acceptance probabilities follows from the standard logic of delegation mechanisms: the agent knows how her reports will be translated into acceptances, so she may as well just report the acceptances.[10] The fact that the type reporting at step (1.5) can be omitted, even when the agent's type is not commonly known, is a consequence of Assumption 1. Specifically, Assumption 1 ensures that once the agent knows her type, she faces no aggregate uncertainty about the future. She knows the joint distribution over private signals, biases, and test results of the applicants who have yet to arrive. So if at step (3) the agent would ever be able to identify an ex-post profitable deviation to having previously misreported her type, the agent would have already had all of the information at step (1.5) to know that the deviation would be profitable. Hence, the principal has no reason not to delay all agent reports until step (3).

2.3 Examples of information and biases

To make the model more concrete, I now introduce two examples to illustrate the kinds of information structures and biases that may arise in applications. In the *normal specification*, the agent has an "idiosyncratic" bias for each applicant, independent of all other terms. The *two-factor model* then shows how one might capture a "systematic" bias in which the soft and hard information are informative about different aspects of a job applicant, and the agent values these aspects differently than does the principal. In Appendix D.2, I describe how to capture some other forms of systematic bias that also induce correlation between signals and biases. For instance, the principal may be in favor of affirmative action for job applicants with certain observable attributes while the agent disagrees.

[10] Recall that test results are publicly observed. If the principal could withhold test results from the agent, then the agent would not in fact know how her reports would translate into acceptances. I discuss how the principal might be able to benefit from withholding test results in Section 5.2.

2.3.1 Normal Specification

In the *normal specification*, assume that

$$Q \sim \mathcal{N}(0, \sigma_Q^2), \tag{F_Q}$$

$$T|Q \sim \mathcal{N}(Q, \sigma_T^2), \tag{$F_{T|Q}$}$$

$$S|Q, T \sim \mathcal{N}(Q, \sigma_S^2), \tag{$F_{S|Q,T}$}$$

$$B|T, S \sim \mathcal{N}(0, \sigma_B^2), \tag{$F_{B|T,S}$}$$

where $\mathcal{N}(\mu, \sigma^2)$ indicates a univariate normal distribution with mean μ and variance σ^2. All variances are taken to be positive and less than infinity.

This specification lets us capture the key forces of the model with a small number of parameters – one variance parameter for each distribution. (Indeed, it would be without loss to further normalize one of the variances to 1.) The parameter σ_Q^2 is the variance of quality in the population, with mean normalized to 0. Then σ_T^2 and σ_S^2 describe how informative the public and private signals are about quality: variance going to 0 would be perfectly informative, and variance going to infinity would be uninformative. Finally, σ_B^2 tells us the strength of the agent's biases.[11] The agent's marginal utility for an applicant is $Q + B$, so an agent with higher σ_B^2 is more biased in that her utility comparisons across applicants depend less on variation in Q and more on variation in B. The agent's type consists of the two dimensions (σ_S^2, σ_B^2), the extent of her information and bias.

2.3.2 Two-Factor Model

In the *two-factor model*, let quality Q be decomposed as

$$Q = Q_1 + Q_2.$$

The two quality factors Q_1 and Q_2 follow some joint distribution F_{Q_1,Q_2}. The test result and private signal then follow conditional distributions $F_{T|Q_1,Q_2}$ and $F_{S|Q_1,Q_2,T}$. Assume further that the private signal S does not add any information about the first

[11] The normalization of the mean of B to 0 is without loss of generality. Adding a constant to the agent's marginal utilities would not change her preferences over sets of hired applicants.

quality factor Q_1 beyond what is known from the test result:

$$\mathbb{E}[Q_1|S,T] = \mathbb{E}[Q_1|T].$$

In other words, the agent gets private information only about the second factor. Think of the public test as measuring a job candidate's $t$echnical ability, while the private interview with a hiring manager yields information about the candidate's $s$ocial skills. Or, in college admissions, an applicant's public SAT score and GPA reveal his or her academic skills, whereas the admissions officer subjectively assesses other "holistic" aspects. In the same vein, the agent might be an expert brought in to evaluate applicants only on features related to her specialty: a writing instructor reads and scores the application essays.

Let the agent's objective be given by

$$V_A = \mathbb{E}[Q_1 + \lambda Q_2|\text{Hired} = 1], \text{ for } \lambda > 0.$$

Compared to the principal's objective of $V_P = \mathbb{E}[Q_1 + Q_2|\text{Hired} = 1]$, there is a bias when $\lambda \neq 1$. We might expect that a hiring manager would overemphasize the importance of social skills, or that a writing instructor would overemphasize writing ability. This "advocate" agent, who values the factor that she evaluates more highly than does the principal, corresponds to $\lambda > 1$. The agent may also be a "cynic" with $\lambda < 1$: an interviewer who thinks that social skills don't matter much, or a writing instructor who thinks that writing ability is overrated.[12] The agent's type in the two-factor model corresponds to a bias parameter λ and a signal structure $F_{S|Q_1,Q_2,T}$.

Let us now rewrite the agent's payoff in the notation of Section 2.1, in which the agent's marginal utility is $Q+B$. The agent's marginal utility in the two-factor model was given as $Q_1+\lambda Q_2 = Q+(\lambda-1)Q_2$, and so the agent maximizes $\mathbb{E}[Q+B|\text{Hired} = 1]$ for $B = (\lambda - 1)\mathbb{E}[Q_2|T,S]$.[13] We see that the "systematic" bias manifests itself as a correlation between bias B and signals T, S: an advocate with $\lambda > 1$, for instance, will be biased in favor of applicants for whom the signals reveal positive news on Q_2.

[12] The bias can also be interpreted as a reduced form for disagreement arising from beliefs rather than preferences. An agent with $\lambda > 1$ would be one who thinks that her private signal S is more informative on quality than the principal thinks it is (and the players "agree to disagree").

[13] As required by the formulation of Section 2.1, the (degenerate) distribution of the bias depends only on the realizations of the signals. I have also written primitive distributions with Q_1 separate from Q_2, but we can translate to the appropriate distributions F_Q, $F_{T|Q}$, and $F_{S|Q,T}$.

A richer formulation could of course add an idiosyncratic bias term – an independent "epsilon" – to the agent's utility for each applicant. In Appendix E I write down and analyze an example which combines the systematic biases of the two-factor model with the idiosyncratic biases of the normal specification.

One distinction to highlight between the two-factor model and the normal specification is that under the normal specification, signals were assumed to be conditionally independent given Q. In the two-factor model, the fact that signals are informative about distinct "quality factors" leads to conditional dependence. Indeed, suppose that the distribution of T depends only on Q_1; the distribution of S depends only on Q_2; and that Q_1 and Q_2 are independent. Then T and S would be *unconditionally* independent. But we would expect T and S to be negatively correlated *conditional* on Q. Fixing $Q = Q_1 + Q_2$, an applicant with higher Q_1 would mechanically have lower Q_2.

2.4 Preliminary Analysis

Given the agent's observation of both the public and private signals, her expectation of an applicant's quality is $\mathbb{E}[Q|T, S]$. Let her corresponding expectation of her own marginal utility from hiring an applicant be given by U_A:

$$U_A \equiv \mathbb{E}[Q|T, S] + B. \tag{1}$$

I informally refer to U_A as the agent utility.

The agent's message in a contract assigns applicants to acceptance probabilities based on (T, S, B). At a given test result, the agent prefers to assign higher acceptance probabilities to applicants with higher utility U_A. That is, at any fixed test result $T = t$, the agent will report acceptance probabilities that are weakly increasing – *monotonic* – in U_A.

For all of the analysis that follows, I maintain the following technical assumption stating that the agent almost surely has a strict preference between any two applicants with the same test result. This condition could arise from continuously distributed biases or from a continuously distributed belief on quality arising from the agent's private signals.

Assumption 2. For each $t \in \mathcal{T}$, the distribution of $U_A|T = t$ has no atoms.

Combining monotonicity with Assumption 2, any two applicants with the same realizations of both T and U_A will (almost surely) be accepted with the same probability, even if they differ on S and B.[14] In other words, applicants with the same test score T and same agent utility U_A *cannot be distinguished* by any contract the principal offers. Formalizing the above discussion:

Observation 1. Fix a contract, a test result $t \in \mathcal{T}$, and two pairs of private signal and bias realizations (s^1, b^1) and (s^2, b^2) in $\mathcal{S} \times \mathbb{R}$. For each $i = 1, 2$, let the expected agent utility U_A of an applicant with test result t, private signal s^i, and bias b^i be denoted by $u_A^i = \mathbb{E}[Q|T = t, S = s^i] + b^i$. Then under any equilibrium acceptance rule χ:[15]

1. **Distinguishability.** If $u_A^1 = u_A^2$ then $\chi(t, s^1, b^1) = \chi(t, s^2, b^2)$. That is, at test result t, applicants with the same U_A have the same probability of acceptance as one another.
2. **Monotonicity.** If $u_A^1 > u_A^2$ then $\chi(t, s^1, b^1) \geq \chi(t, s^2, b^2)$. That is, at test result t, applicants with higher U_A have weakly higher acceptance probabilities.

Distinguishability establishes that all applicants with the same test score and same agent utility are treated identically. Some of them may have high perceived quality $\mathbb{E}[Q|T, S]$ and low bias B, while others have low quality and high bias. But no contract can induce the agent to distinguish these applicants. Define $U_P(t, u_A)$ to be the average quality of applicants with test result $T = t$ (observed by the principal) and agent utility $U_A = u_A$ (unobserved by the principal):

$$U_P(t, u_A) \equiv \mathbb{E}[Q|T = t, U_A = u_A]. \tag{2}$$

I informally refer to U_P as the principal utility.

Given distinguishability, we can rewrite acceptance rules as mappings from (T, U_A) – rather than (T, S, B) – into acceptance probabilities. As such, going forward, I take acceptance rules to be functions $\chi : \mathcal{T} \times \mathbb{R} \to [0, 1]$ where $\chi(t, u_A)$ indicates the

[14] Given monotonicity, at any given $T = t$ there can be only countably many values u_A at which applicants with $U_A = u_A$ have different acceptance probabilities. (A correspondence that is monotonic in the strong set order can have only countably many points for which it is not single-valued.) Assumption 2 says that it is zero probability that U_A is in any specified countable set.

[15] Any equilibrium acceptance rule yields identical payoffs for both parties as one which satisfies these properties. But the agent can always "deviate" on a set of applicants of probability 0 without affecting payoffs.

probability of accepting an applicant with $T = t$ and $U_A = u_A$. The principal's payoff V_P – the average quality of hired applicants – under acceptance rule χ can now be written as

$$V_P = \frac{1}{k}\mathbb{E}[\chi(T, U_A) \cdot U_P(T, U_A)]. \tag{3}$$

Monotonicity establishes that any acceptance rule $\chi(t, u_A)$ is weakly increasing in u_A for every t. A deterministic contract will yield acceptance rules $\chi(t, u_A)$ that are step functions in u_A, taking on values in $\{0, 1\}$.

3 Common knowledge of agent type

In this section I consider optimal contracts under *common knowledge of the agent's type*: the distributions $F_{S|Q,T}$ and $F_{B|T,S}$ are known to the principal prior to contracting. Combining common knowledge of the agent's type with common knowledge of F_Q and $F_{T|Q}$ from Assumption 1, the principal knows the induced joint distribution of test result T and agent utility U_A across applicants. The principal knows the function U_P mapping (T, U_A) to average quality. And, given any contract, the principal can predict in advance the acceptance rule χ that will be induced by the agent's choices.

It must hold that the induced acceptance χ rule selects a total of k applicants, and that it is monotonic in agent utility. Writing out these two necessary conditions:

$$\mathbb{E}[\chi(T, U_A)] = k \tag{4}$$

$$\text{For all } t, \chi(t, u_A) \text{ is weakly increasing in } u_A. \tag{5}$$

In fact, any acceptance rule χ satisfying these two conditions can be implemented by some contract. Take some such χ; this χ induces a (commonly known) distribution of acceptance probabilities at each test result. A contract can then specify that at each test result, the agent selects applicants satisfying this distribution of acceptance probabilities. Given such a contract, the agent's optimal behavior of monotonically assigning higher acceptance probabilities to higher agent utilities recovers χ.

So the principal's problem under common knowledge the agent's type can be stated as maximizing the objective (3) over the choice of function $\chi : \mathcal{T} \times \mathbb{R} \to [0, 1]$, subject to the constraints (4) and (5).

3.1 Solving a relaxed problem

One upper bound on the principal's payoff would result from maximizing the objective (3) subject to the budget constraint (4), without imposing the monotonicity constraint (5). Denote this *upper bound acceptance rule* (UBAR) as χ^{UBAR}. If the solution to this relaxed problem satisfies monotonicity (5) then it is a solution to the original problem, i.e., it is implementable as an optimal contract.

UBAR can be described in the following manner. First, find the level of principal expected utility u_P^c such that a share of k agents have $U_P \geq u_P^c$; formally,

$$u_P^c \equiv \sup \left\{ u_P \in \mathbb{R} \mid \text{Prob}[U_P(T, U_A) \geq u_P] \geq k \right\}. \tag{6}$$

UBAR accepts all applicants with U_P above the cutoff and rejects all of those below: $\chi^{\text{UBAR}}(t, u_A) = 1$ if $U_P(t, u_A) > u_P^c$ and $\chi^{\text{UBAR}}(t, u_A) = 0$ if $U_P(t, u_A) < u_P^c$. If there is a mass of applicants with $U_P(T, U_A) = u_P^c$, then there is flexibility over which of these applicants are accepted in order to get k applicants to be hired in total. In that case, over any region of flexibility let χ^{UBAR} take values in $\{0, 1\}$ and let it be monotonic in U_A.[16]

If χ^{UBAR} is monotonic, then it is implementable as an optimal contract. Moreover, χ^{UBAR} is deterministic by construction: it takes values only in $\{0, 1\}$. So if it is monotonic then the contract can be implemented by specifying the appropriate acceptance rates at each test result. Let $\alpha^{\text{UBAR}}(t)$ be the share of applicants accepted at test result $T = t$ under χ^{UBAR}:

$$\alpha^{\text{UBAR}}(t) \equiv \mathbb{E}[\chi^{\text{UBAR}}(T, U_A) | T = t].$$

Sufficient conditions for monotonicity. The following alignment condition guarantees monotonicity of the upper bound acceptance rule.

Definition. Utilities are *aligned up to distinguishability* if for all t, the principal's expected utility $U_P(t, u_A)$ is weakly increasing in u_A over the support of $U_A | T = t$.

Loosely speaking, utilities are aligned up to distinguishability if applicants who are more preferred by the agent are of higher average quality. Think of the agent

[16] That is, on those values (t, u_A) for which $U_P(t, u_A) = u_P^c$, let $\chi^{\text{UBAR}}(t, u_A) = 0$ for u_A below some t-specific cutoff agent utility and let $\chi^{\text{UBAR}}(t, u_A) = 1$ for u_A above the cutoff.

utility $U_A = \mathbb{E}[Q|T = t, S] + B$ as a noisy signal of perceived quality $\mathbb{E}[Q|T = t, S]$; the condition states that higher realizations of U_A uniformly imply a higher average $\mathbb{E}[Q|T = t, S]$.

Under alignment, at every test result the ordering of applicants by agent utility U_A is the same as by principal utility U_P. Therefore UBAR will be monotonic: it accepts applicants with higher U_A over those with lower U_A.[17] Figure 1 illustrates UBAR when alignment up to distinguishability does and does not hold.

I confirm below that alignment up to distinguishability holds for the normal specification (Section 3.3), the two-factor model (Appendix D.1), and for a mixture of the two (Appendix E). However, under particular joint distributions of Q, T, S, and B, it is possible that alignment can be violated, even when the bias B is independent of the soft information S: some high realization of U_A may indicate a very high bias combined with negative information on quality. That said, the following log-concavity condition on the bias distribution guarantees alignment up to distinguishability regardless of the quality and signal distributions.

Lemma 1. *Suppose that for all $t \in \mathcal{T}$, the bias distribution $F_{B|T=t,S}$ is independent of S and is log-concave.*[18] *Then utilities are aligned up to distinguishability.*

The hypothesis of Lemma 1 can easily be checked from the primitives of the problem. For instance, any time the bias is independent of the private signal and is normally distributed (as it is in the normal specification) then utilities must be aligned up to distinguishability.

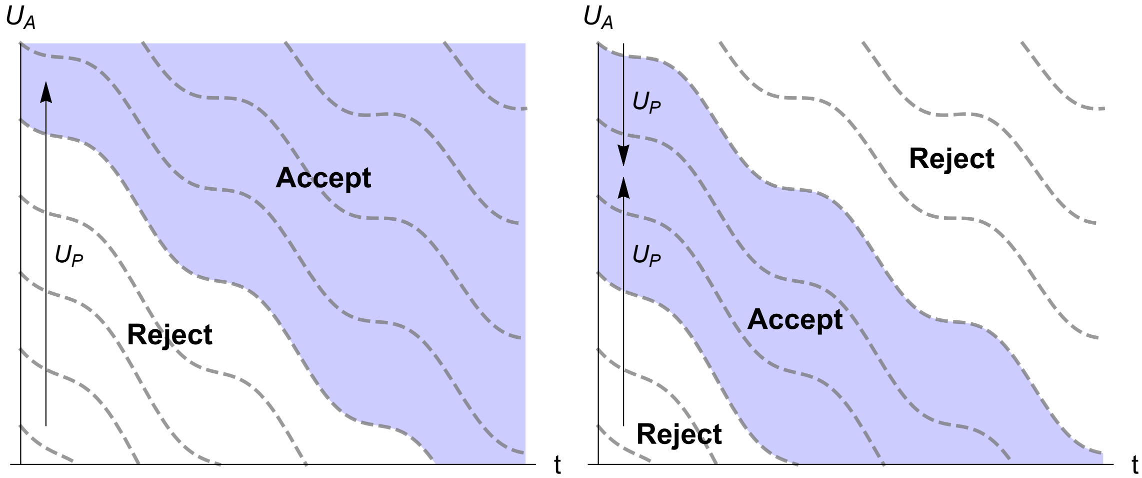
The optimal contract as an acceptance rate function.

Proposition 1. *Under common knowledge of the agent's type, suppose that utilities are aligned up to distinguishability. Then the deterministic contract characterized by requiring acceptance rate at test result t of $\alpha^{\mathrm{UBAR}}(t)$ implements the upper bound acceptance rule. This contract is an optimal contract.*

[17] Alignment up to distinguishability is sufficient but not necessary for UBAR to be monotonic. Holding fixed all other primitives while varying k, however, alignment up to distinguishability is necessary and sufficient for UBAR to be monotonic for all possible $k \in (0, 1)$.

[18] A distribution F on $\mathbb{R}$ is said to be log-concave if it admits a pdf f with convex support, and if $\log f$ is a concave function over the support.

Figure 1: Alignment up to distinguishability and the upper bound acceptance rule



(a) Aligned (b) Not aligned

The dashed curves show possible principal indifference (iso-U_P) curves in (t, U_A)-space. The arrows indicate the direction of higher principal utility. Utilities are aligned up to distinguishability in panel (a) but not panel (b). Under UBAR, applicants with principal utility above some cutoff are accepted; acceptance regions are shaded. UBAR is monotonic, and therefore implementable as the optimal contract, in panel (a) but not panel (b).

The optimal contract as an average score. The contract of Proposition 1 separately describes the acceptance rate at every test result. There is another implementation of UBAR that may in some cases be simpler to express. This alternative implementation first specifies a "score function" mapping every realization of the (arbitrary-dimensional) hard information into a real number. The contract then asks the agent to select k applicants subject to a restriction only on the average score of those selected.[19]

In order to construct the score function, recall that under a monotonic and deterministic acceptance rule, every applicant with U_A above some test-result-specific cutoff is accepted and every applicant with U_A below the cutoff is rejected. We can define the appropriate score function based on these agent utility cutoffs from the upper bound acceptance rule. Specifically, let $u_A^c(t)$ be the cutoff at test result $T = t$:

[19] According to the Wall Street Journal (2014), in many pre-employment tests, "responses to an online personality test are fed into an algorithm that scores each applicant, sometimes on a scale of red, yellow and green. Scoring systems vary by testing provider, and the companies can customize their methods to fit an employer's demands." These scores may be used not only for contractual restrictions (as in the current paper) but also to help the manager make sense of the test result.

take $u_A^c(t) \in \mathbb{R} \cup \{-\infty, \infty\}$ such that for $t \in \mathcal{T}$ and u_A in the support of $U_A|T = t$, it holds that $\chi^{\text{UBAR}}(t, u_A) = 1$ if $u_A > u_A^c(t)$ and $\chi^{\text{UBAR}}(t, u_A) = 0$ if $u_A < u_A^c(t)$.[20] In other words, $u_A^c(t)$ is the function describing the cutoff indifference curve of Figure 1 panel (a), for which $U_P(t, u_A^c(t)) = u_P^c$ at all t.

Now define a *score function* $C : \mathcal{T} \to \mathbb{R} \cup \{-\infty, \infty\}$ to be any negative affine transformation of u_A^c, i.e., $C(t) = a_0 + a_1 u_A^c(t)$ for some $a_1 < 0$.

Proposition 2. *Under common knowledge of the agent's type, suppose that utilities are aligned up to distinguishability. Let C be a score function, and suppose further that $\frac{1}{k}\mathbb{E}[\chi^{\text{UBAR}}(T, U_A)C(T)]$ – the expected score of those applicants hired under UBAR – is finite and equal to $\kappa \in \mathbb{R}$. Then the contract that asks the agent to deterministically select any k applicants satisfying $\mathbb{E}[C(T)|\text{Hired} = 1] = \kappa$ implements the upper bound acceptance rule. This contract is an optimal contract.*

For intuition, consider the following informal Lagrangian argument. Let λ_0 be the multiplier representing the shadow cost on the agent of hiring more applicants and let λ_1 be the shadow cost of increasing the average score of those who are hired. At the optimum, the agent hires an applicant if $U_A \geq \lambda_0 + \lambda_1 C(T)$. For score function $C(T) = a_0 + a_1 u_A^c(T)$, plug in the multipliers $\lambda_0 = -a_0/a_1$ and $\lambda_1 = 1/a_1$: the agent hires an applicant if $U_A \geq u_A^c(T)$, which is the condition defining UBAR.

The role of taking the score function to be a *negative* affine transformation of u_A^c is to ensure that, on the margin, the agent prefers to *decrease* the average score. (With a positive affine transformation, the agent would prefer to increase the average score.) In the Lagrangian argument above, the fact that the agent wants to decrease the score follows from the multiplier $\lambda_1 = 1/a_1$ being negative. More intuitively, perhaps, consider relaxing the constraint $\mathbb{E}[C(T)|\text{Hired} = 1] = \kappa$. The agent would now prefer to reject some applicants at test results with low u_A^c – the agent utility of the marginal applicant – and hire more at test results with high u_A^c. Taking C as a negative transformation of u_A^c means that the newly hired applicants have lower score than the newly rejected ones.

Hence, we can equivalently implement the contract of Proposition 2 as a floor on the average score, replacing the constraint $\mathbb{E}[C(T)|\text{Hired} = 1] = \kappa$ with $\mathbb{E}[C(T)|\text{Hired} = 1] \geq \kappa$. In response, the manager chooses an average score exactly at the minimum.

[20] At $T = t$, if the agent utility can be unboundedly negative [or positive] and the agent is to accept [or reject] all applicants, then $u_A^c(t) = -\infty$ [or $+\infty$].

As discussed in Section 3.4, the floor interpretation may be preferred when we think about applying this form of contract to a setting with aggregate uncertainty.

In Sections 3.3, D.1, and E.2 I apply Propositions 1 and 2 to characterize the optimal contract under common knowledge of the agent's type for the normal specification, the two-factor model, and for a combination of the two.

3.2 The general solution

In Appendix A, I solve for the optimal contract – maximizing (3) subject to (4) and (5) – in the general case where alignment up to distinguishability need not hold, and thus the upper bound acceptance rule may not be implementable.

The solution involves "ironing" test result by test result. For instance, at some test result the agent may be required to treat all applicants from the 70th through 80th percentiles of U_A identically. If the agent is to accept less than 20% of the applicants at this test result, she deterministically accepts her favorites. If she is to accept (20+x)% for $0 < x < 10$, she accepts her favorite 20% deterministically, and gives each of the 10% of applicants in the pooling range an $x/10$ chance of acceptance. If she is to accept 30% of applicants or more, she once again deterministically accepts her favorites. After ironing applicants in an appropriate manner, the problem can be solved using an approach similar to that in Section 3.1.

To see the potential benefit of randomization, consider the special case in which there is no meaningful hard information at all. Under alignment up to distinguishability – higher agent utility U_A implies higher principal utility U_P – the optimal contract would be to give the agent full discretion. Under the extreme case of anti-alignment, though, with higher U_A corresponding to *lower* U_P, the principal would do better by randomly accepting applicants. If U_P were nonmonotonically increasing, decreasing, and increasing in U_A, then the principal might let the agent accept some applicants deterministically and others probabilistically.

One conclusion from Appendix A is that, despite the potential benefit of randomized acceptances, randomization need only be used on at most a single test result. So when test results are continuously distributed, and thus behavior at any single test score is irrelevant, we still get a deterministic optimal contract that can be implemented by specifying an acceptance rate function. This optimal contract may perform worse than UBAR, though.

3.3 The normal specification

Let us now apply the results of Section 3.1 to solve for the optimal contract under the normal specification with common knowledge of the agent's type.

Using standard rules for Bayesian updating with normal priors and normal signals, one can solve for $\mathbb{E}[Q|T,S]$ in the normal specification as

$$\mathbb{E}[Q|T,S] = \frac{\sigma_S^2\sigma_Q^2 T + \sigma_T^2\sigma_Q^2 S}{\sigma_Q^2\sigma_T^2 + \sigma_Q^2\sigma_S^2 + \sigma_T^2\sigma_S^2}. \tag{7}$$

The agent utility is $U_A = \mathbb{E}[Q|T,S] + B$. From the joint distributions of S, T, and B, one can derive $U_P(T, U_A)$, the average quality conditional on the test result and the agent utility, as below; see details of the calculations in Appendix G.1.

$$U_P(T, U_A) = \beta_T \cdot T + \beta_{U_A} \cdot U_A, \text{ for} \tag{8}$$

$$\beta_T = \frac{\sigma_B^2\sigma_Q^2}{(\sigma_Q^2 + \sigma_T^2)(\eta + \sigma_B^2)}, \tag{9}$$

$$\beta_{U_A} = \frac{\eta}{\eta + \sigma_B^2}, \tag{10}$$

$$\text{with } \eta \equiv \frac{\sigma_Q^4\sigma_T^4}{(\sigma_Q^2 + \sigma_T^2)(\sigma_Q^2\sigma_T^2 + \sigma_Q^2\sigma_S^2 + \sigma_T^2\sigma_S^2)}.[21] \tag{11}$$

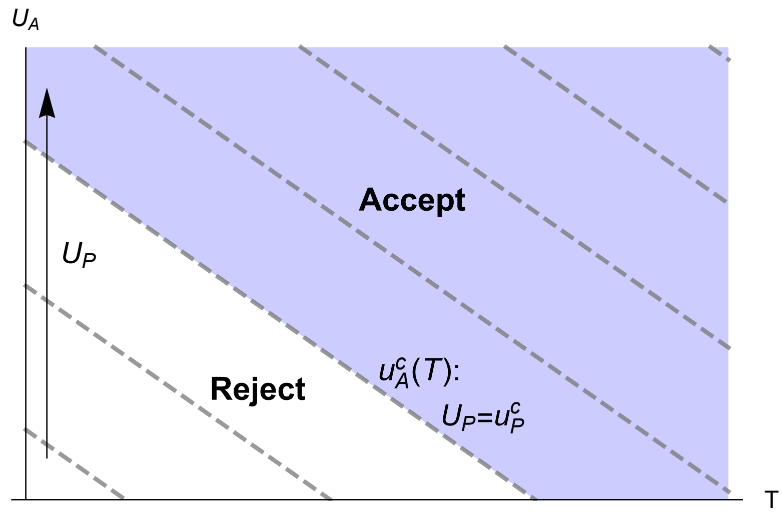
The principal utility U_P is linear in both T and U_A, with respective coefficients $\beta_T > 0$ and $\beta_{U_A} > 0$. The fact that β_{U_A} is positive confirms that utilities are aligned up to distinguishability, as implied by Lemma 1: principal utility is increasing in agent utility. So the upper bound acceptance rule is implementable.

We also see that $\beta_{U_A} < 1$. Under the normalization that the agent and principal both value quality at the same rate, an applicant who is thought to be one utility unit better by the agent is somewhat less than one utility unit better to the principal. The increased agent utility is inferred as partly due to higher quality and partly due to higher bias.

The linearity of U_P in T and U_A means that the principal indifference curves are linear in (T, U_A)-space, with slope $-\frac{\beta_T}{\beta_{U_A}} < 0$. UBAR accepts all applicants "up and to the right" of a cutoff indifference curve. See Figure 2.

[21] Equation (15), below, gives an economic interpretation of η. Conditional on a given test result, η is the variance in the agent's beliefs on applicant quality arising from her private signal.

Figure 2: Upper Bound Acceptance Rule for the Normal Specification



The dashed curves are principal indifference curves, each a line with slope $-\beta_T/\beta_{U_A}$. The arrow indicates that higher indifference curves represent higher principal utilities U_P, i.e., that utilities are aligned up to distinguishability. Under UBAR, applicants are accepted above the cutoff indifference curve at principal utility $U_P = u^c_P$, which induces a mass of k to be accepted in total. This cutoff indifference curve is denoted by $u^c_A(T)$.

Following Proposition 1, one implementation of UBAR as an optimal contract specifies acceptance rates at each test result. Applicants' test results and agent utilities (T, U_A) are distributed joint normally and are positively correlated. So the share of applicants above the cutoff indifference curve – a downward sloping line – is increasing and follows a normal CDF function. Indicating the CDF of a standard normal distribution by Φ, the share of applicants accepted under UBAR at test result T works out to $\alpha^{\text{UBAR}}(T) = \Phi(\gamma^*_T T - \gamma^*_0)$, with $\gamma^*_T > 0$ given by

$$\gamma^*_T \equiv \frac{\sigma^2_Q \sqrt{\eta + \sigma^2_B}}{\eta(\sigma^2_Q + \sigma^2_T)}. \tag{12}$$

The value of γ^*_0 – for which I do not provide an explicit formula – is then set so that a total of k applicants are accepted.

We can equivalently express this contract in the alternative manner of Proposition 2: the agent picks any k applicants subject to $\mathbb{E}[C(T)|\text{Hired} = 1] = \kappa$ (or $\geq \kappa$), for some score function C and average score κ. The score function can be chosen as any negative affine transformation of the cutoff indifference curve of Figure 2. Indifference curves are downward sloping lines and so C can be any increasing linear function. For the normal specification I take the convention of setting $C(T) = T$, i.e., setting the score equal to the test result. (In the context of the normal specification, I hereafter go back and forth between calling T a test result and a test score.) That is, the contract requires the agent to choose any k applicants with a specified average test score. Denote the optimal choice of average test score by κ^*.[22]

The following Proposition formalizes these results.

Proposition 3. *Under the normal specification with common knowledge of the agent's type, the optimal contract can be implemented in either of the following ways. The agent is allowed to hire any set of k applicants, subject to:*

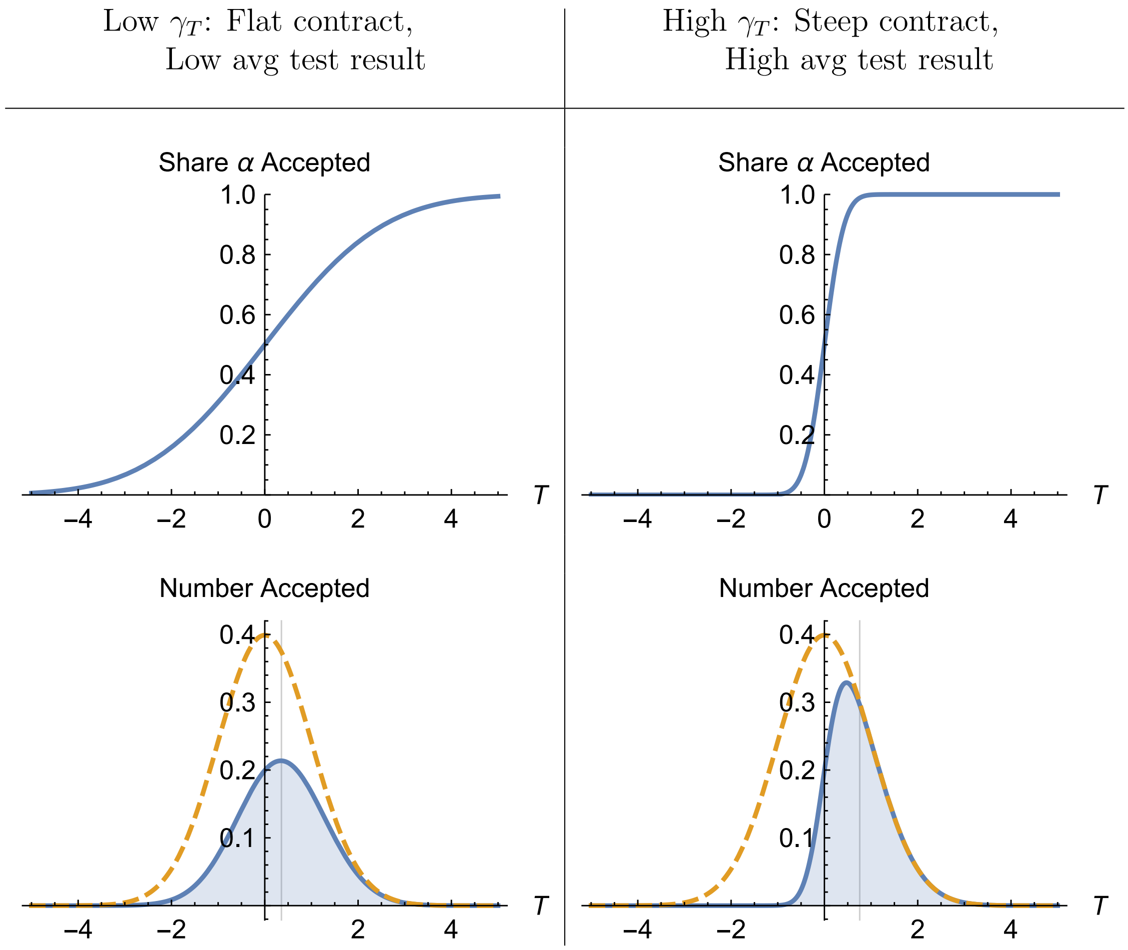
1. *An acceptance rate function of $\alpha(T) = \Phi(\gamma_T^* T - \gamma_0^*)$; or,*
2. *An average test result of accepted applicants, $\mathbb{E}[T|\text{Hired} = 1]$, equal to κ^*.*

Let us focus on the first implementation, the acceptance rate function. The optimal contract induces a normal CDF acceptance rate – an S curve – of the form $\alpha(t) = \Phi(\gamma_T t - \gamma_0)$. More applicants are accepted at higher test results, with the share of applicants accepted approaching 0 as $t \to -\infty$ and approaching 1 as $t \to \infty$. The contracts are characterized by a one-dimensional statistic, the steepness γ_T. A steeper contract with a higher γ_T would correspond to a higher average test score in the second implementation. See Figure 3 for an illustration of such contracts.

Heuristically, a steeper contract means that hiring depends more on the test and less on the agent's input: steeper contracts give the agent "less discretion." Being more precise, in Appendix B I show that the Full Discretion contract, in which the agent selects her favorite applicants, would induce a normal CDF acceptance rate with steepness γ_T^{FD} satisfying $0 < \gamma_T^{\text{FD}} < \gamma_T^*$. The agent prefers a flatter acceptance rate than does the principal ($\gamma_T^{\text{FD}} < \gamma_T^*$) because she cares about idiosyncratic factors in addition to quality. So in the range of contracts that an agent may face ($\alpha(T) = \Phi(\gamma_T T - \gamma_0)$ with $\gamma_T > \gamma_T^{\text{FD}}$), a steeper contract requires the agent to pick fewer applicants with low test scores, whom she prefers on the margin. Taking γ_T to infinity would yield the No Discretion contract in which all applicants below some test score cutoff are rejected and all applicants above are accepted.

[22] A formula for κ^* in terms of the primitive parameters is given by Equation (53) in Appendix G.3, plugging in for $\sigma_{U_P}^2$ from (18): $\kappa^* = \frac{\sigma_Q^2 \sqrt{\sigma_Q^2 + \sigma_T^2} \cdot R(k)}{\sqrt{\sigma_Q^2 + \frac{\eta^2}{\eta + \sigma_B^2}}}$, with $R(k)$ as defined below in (22).

Figure 3: Contracts with $\alpha(T) = \Phi(\gamma_T T - \gamma_0)$ for different γ_T.



The first row illustrates the share of applicants accepted at each test score T under a rule specifying an acceptance rate function of $\Phi(\gamma_T T - \gamma_0)$. Adjusting γ_0 would translate these functions left or right. The solid curves in the second row show the pdf of test results for those accepted, with a grey line at the mean; dashed curves indicate the pdf of test results for the full applicant pool. Steeper contracts with higher γ_T yield higher average test results. In the example the unconditional distribution of test scores is $\mathcal{N}(0, 1)$, the low value of γ_T is .5 and the high value is 3, and $k = .5$, implying $\gamma_0 = 0$ for both low and high γ_T. The flat contract with $\gamma_T = .5$ yields $\mathbb{E}[T|\text{Hired} = 1] = .357$ and the steep contract with $\gamma_T = 3$ yields $\mathbb{E}[T|\text{Hired} = 1] = .757$.

We can now explore comparative statics on the optimal steepness γ_T^* with respect to the five parameters. One qualification to bear in mind is that, as we vary σ_Q^2 or σ_T^2, the unconditional variance of the test scores, $\sigma_Q^2 + \sigma_T^2$, varies as well. So the interpretation of the steepness coefficient as a measure of discretion changes. For instance, if we kept the coefficient γ_T fixed as we increased the variance of test scores, the test would become "more predictive" of hiring – a one standard deviation increase in test scores would have a larger effect on hiring rates. The corrected coefficient, which tells us the impact of a one standard deviation rather than a one unit increase in test scores, is $\gamma_T^* \sqrt{\sigma_Q^2 + \sigma_T^2}$.[23] In the following Proposition, I include comparative statics on this renormalized coefficient when relevant (parts 2 and 5).

Proposition 4. *In the contract of Proposition 3 part 1, the contracting parameter γ_T^* given by* (12) *has the following comparative statics and limits:*

1. γ_T^* *is independent of* k.
2. γ_T^* *and* $\gamma_T^* \sqrt{\sigma_Q^2 + \sigma_T^2}$ *decrease in* σ_T^2, *with* $\lim_{\sigma_T^2 \to 0} \gamma_T^* = \lim_{\sigma_T^2 \to 0} \gamma_T^* \sqrt{\sigma_Q^2 + \sigma_T^2} = \infty$ *and* $\lim_{\sigma_T^2 \to \infty} \gamma_T^* = \lim_{\sigma_T^2 \to \infty} \gamma_T^* \cdot \sqrt{\sigma_Q^2 + \sigma_T^2} = 0$.
3. γ_T^* *increases in* σ_S^2, *with* $\lim_{\sigma_S^2 \to 0} \gamma_T^* \in (0, \infty)$ *and* $\lim_{\sigma_S^2 \to \infty} \gamma_T^* = \infty$.
4. γ_T^* *increases in* σ_B^2, *with* $\lim_{\sigma_B^2 \to 0} \gamma_T^* \in (0, \infty)$ *and* $\lim_{\sigma_B^2 \to \infty} \gamma_T^* = \infty$.
5. γ_T^* *and* $\gamma_T^* \sqrt{\sigma_Q^2 + \sigma_T^2}$ *may increase or decrease in* σ_Q^2, *with* $\lim_{\sigma_Q^2 \to 0} \gamma_T^* = \lim_{\sigma_Q^2 \to 0} \gamma_T^* \sqrt{\sigma_Q^2 + \sigma_T^2} = \infty$, $\lim_{\sigma_Q^2 \to \infty} \gamma_T^* \in (0, \infty)$, *and* $\lim_{\sigma_Q^2 \to \infty} \gamma_T^* \sqrt{\sigma_Q^2 + \sigma_T^2} = \infty$.

Part 1 reiterates that the steepness of the contract does not depend on the number of people to be hired. Hiring fewer or more applicants just translates the acceptance rate function left or right.

Part 2 finds that as the test becomes less informative, the contract gets flatter: it places less weight on the test results, measured either in absolute or relative terms. As the test becomes perfectly uninformative, the contract sets a constant acceptance rate across all scores. As the test becomes perfectly informative, we approach a perfectly steep No Discretion contract in which hiring is entirely based on the test.

Part 3 finds that as the agent becomes better informed, the contract gets flatter, giving the agent more discretion. Part 4 finds that as the agent becomes more biased,

[23] $\Phi(\gamma_T^* T - \gamma_0)$ can be rewritten as $\Phi\left(\gamma_T^* \sqrt{\sigma_Q^2 + \sigma_T^2} \cdot \frac{T}{\sqrt{\sigma_Q^2 + \sigma_T^2}} - \gamma_0\right)$, where $\frac{T}{\sqrt{\sigma_Q^2 + \sigma_T^2}}$ is the z-score of the test result.

the contract gets steeper, giving the agent less discretion. In the limits where the agent is fully uninformed or where her preferences are entirely unrelated to quality, we approach No Discretion.[24]

Hoffman et al. (2016) studies the use of pre-employment tests in hiring and compares the limit contracts of Full versus No Discretion. They argue that Full Discretion is preferred to No Discretion when the agent has low bias and high private information, but not when the agent has high bias and/or low information. Parts 3 and 4 point to similar tradeoffs in the optimal contracts. One should use a flatter contract, yielding more discretion, when bias is low or information is high.

Part 5, included for completeness, establishes that the steepness of the contract can vary nonmonotonically with the variance of quality in the population.

In Appendix B I perform a similar comparative statics analysis for the Full Discretion outcome. The main takeaway is that the principal and agent agree about the impact of agent information, but they disagree about the impact of agent bias. When the agent has better private information (lower σ_S^2), the acceptance rate functions in both the optimal contract and the Full Discretion outcome become flatter. A more informed agent is better at identifying high quality applicants who tested poorly, and the principal wants to let her accept more of them. But when the agent is more biased (higher σ_B^2), the optimal contract becomes steeper while the Full Discretion outcome becomes flatter; the agent wants to accept more low quality applicants who tested poorly, and the principal wants to stop her from accepting them.

3.4 Discussion: Implementation in finite economies

One important simplification of this paper is to assume that there is a "large number" of applicants, modeled as a continuum. This assumption allows me to characterize optimal contracts through two equivalent implementations. First, the firm can specify the acceptance rate at each realization of hard information. Second, the firm can assign all applicants a score based on their observables, and require the manager to select a set of applicants with an appropriate average score.

With a finite number of applicants instead of a continuum, the firm could imple-

[24]Proposition 4 parts 2 and 4 confirm the so-called "uncertainty principle" and "ally principle" of delegation, reviewed in Huber and Shipan (2006): a principal should grant more discretion when he has more uncertainty about what actions to take, and when the agent's preferences are more aligned with his own.

ment approximations of each of these contract forms. These approximated contracts would no longer necessarily be exactly optimal, of course – nor would approximations of the two different implementations remain equivalent.

To highlight the distinction between the approximate implementations, suppose that hard information can take on many values: say that the hard information consists of a continuous-valued test result. Even with a very large finite number of applicants, then, no two applicants would be exactly identical on observables.

A natural finite approximation of the acceptance rate implementation might be to divide realizations of hard information into bins and then specify an acceptance rate at each bin. For instance, with test results ranging from 0 to 100, the firm may divide test results into the four bins of 0-25, 25-50, 50-75, and 75-100.[25] The firm then requires the manager to accept zero applicants in the bottom bin, 10% of applicants in the next higher bin, 25% of applicants in the third bin, and 50% of applicants in the top bin. The firm here faces a tradeoff over the number versus the size of these bins. Having fewer bins that each contained more applicants would recover benefits of aggregation and linking decisions, à la Jackson and Sonnenschein (2007) – the manager could be asked to select the right tail within each bin instead of selecting all applicants or none of them. But having fewer bins would restrict the firm's ability to force the manager to treat observably distinct applicants differently from one another.

In contrast, the average score contract can be approximated for finite economies without any need for binning. Take the score function as specified in the continuum contract, and impose the same form of constraint: the average score of selected applicants must be at or above a floor. (In the continuum economy, without aggregate uncertainty, the average score hits the floor precisely.) The score function effectively puts all applicants into the "same bin" and allows them all to be compared to one another. For the example above, perhaps the firm specifies a minimum average test result of 70. If the manager happens to see unexpectedly strong private signals about some applicants with test results below 50, she has the flexibility to accept more of these low-scoring applicants as long as she also accepts more high-scoring ones (and fewer in the middle).

In Appendix C I explore how one might approximate these two finite implemen-

[25]One might want to condition the cutoffs for these bins on the realized distribution of test results; for example, the approach in Appendix C would have the top bin out of four contain the top quartile of test results. Similarly, for the average score implementation below, one might want to adjust the average score cutoff in response to the realized distribution of test results.

tations in the context of a normal specification example. I suppose that the firm will accept 1/3 of N applicants, with N ranging from 12 to 96; principal payoffs under the two types of finite contracts are reported in Table 1 of Appendix C. I find that, even for these moderate numbers of applicants, both a binned acceptance rate and a minimum average score contract do well – they both recover a large share of the theoretical upper bound on payoffs that is exactly achieved in the continuum model. Consistent with the discussion above, I also find that the minimum average score implementations do perform better than binned acceptance rates.

4 Unknown agent type in the normal specification

Now consider the possibility that the agent's type is unknown to the principal. To analyze the uncertainty in a Bayesian manner, I need to make stronger parametric assumptions. Accordingly, this section focuses only on the normal specification. Here, the agent's type $\theta \equiv (\sigma_S^2, \sigma_B^2)$ describes how informed she is and how strong are her idiosyncratic biases. Let G be the principal's prior belief over the agent's type θ in $\mathbb{R}^2_{++}$. In Appendix E.3, I extend the results of this section to a combined model that puts together the normal specification and two-factor model. There, agents may be heterogeneous on the three type dimensions of information, idiosyncratic bias, and systematic bias.

4.1 Connection to one-dimensional delegation

Before moving to the analysis, it will be helpful to recall the one-dimensional delegation problem, introduced by Holmström (1977, 1984). In that problem, the agent observes a one-dimensional state, which determines principal and agent preferences over a one-dimensional action. The contract takes the form of a "delegation set" specifying the actions that the agent may choose. A common result in this literature is that, under some functional form and distributional assumptions, an interval delegation set is optimal. For instance, an agent who is biased towards high actions may be given a cap, and an agent who is biased towards low actions may be given a floor. Versions of such a result appear in Melumad and Shibano (1991), Martimort and Semenov (2006), Alonso and Matouschek (2008), Goltsman et al. (2009), Kováč and Mylovanov (2009), Amador and Bagwell (2013), Ambrus and Egorov (2015), Amador and Bagwell (2016), and Amador et al. (2018). The latter six of these papers also consider the possibility of some form of "money burning," an auxiliary action that reduces the payoffs of both players – for instance, money burning may be caused by taking a random rather than a deterministic decision action. These papers give various conditions under which money burning is or is not used in conjunction with optimal (interval) delegation sets.

The problem of selecting applicants is posed as a higher-dimensional one. The action effectively corresponds to the entire function mapping test scores to acceptance rates. (In fact, that action space would only describe deterministic contracts; stochastic contracts could be more general.) The players' preferences over this action are determined by the agent's two-dimensional type. When the agent's type was known, and thus the principal faced no aggregate uncertainty, Section 3 found that it was optimal to specify a single acceptance rate function, i.e., a single action. When the agent's type is unknown, however, the principal may offer a menu of acceptance rates to screen across agent types.[26]

I will be able solve for this optimal menu under the normal specification. I do so below by formally transforming the problem of selecting applicants into a one-dimensional delegation problem with money burning.[27] The agent's behavior in any contract is determined by a one-dimensional projection of her two-dimensional type. Moreover, there is a one-dimensional set of "frontier" actions – the normal CDF acceptance rates, parametrized by steepness γ_T. Any other acceptance rate function gives the players the payoffs of some normal CDF acceptance rate, minus money burning that harms both players. After the appropriate transformation, I can apply conditions from the one-dimensional delegation analysis of Amador et al. (2018) to characterize the optimal contract as one that sets a floor on actions and does not burn money.[28] Translating back into acceptance rate functions, the contract gives the agent a menu of normal CDF acceptance rates with a floor on the steepness. That is, the

[26]In a one-dimensional delegation problem, if the agent has some private information prior to contracting, then the principal may benefit from an additional screening step in which the agent is offered a menu over delegation sets; see Krähmer and Kováč (2016) or Tanner (2018).

[27]Guo (2014) similarly solves a contracting problem without transfers by reducing a highly dimensional action space into a one-dimensional frontier plus money burning, then treating it as a one-dimensional delegation problem.

[28]Amador and Bagwell (2013) provides conditions to verify whether a proposed interval is an optimal delegation set. Amador et al. (2018) builds on these results to give sufficient conditions guaranteeing that there exists some interval that is an optimal delegation set.

agent may go steeper, but not flatter – she must follow the test sufficiently closely. The contract can also be implemented by allowing the agent to select any applicants subject to a floor on their average test score.

4.2 Rewriting payoffs

As a preliminary step, let us calculate the distribution of principal and agent utilities at each test result for a fixed agent type θ. The conditional distribution of agent utilities U_A given test score T is derived in the proof of Proposition 1 part 1 as

$$U_A|T \sim \mathcal{N}\left(\mu_{U_A}(T), \sigma^2_{U_A}(\theta)\right), \text{ for} \tag{13}$$

$$\mu_{U_A}(T) = \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} T \tag{14}$$

$$\sigma^2_{U_A}(\theta) = \eta(\theta) + \sigma_B^2. \tag{15}$$

The mean $\mu_{U_A}(T)$ is linear in the test result T but does not depend on the agent's type θ. The variance $\sigma^2_{U_A}(\theta)$ is constant in T but depends on the type θ. Recall that $\eta(\theta)$, defined in (11), depends on the agent's type through σ_S^2 but not σ_B^2.[29]

Continuing to fix θ, equations (8) - (10) give us the principal utility of $U_P(T, U_A) = \beta_T(\theta)T + \beta_{U_A}(\theta)U_A$. Plugging into (13) - (15), we can calculate the distribution of $U_P(T, U_A)$ conditional on T but not U_A.

$$U_P(T, U_A)|T \sim \mathcal{N}\left(\mu_{U_P}(T), \sigma^2_{U_P}(\theta)\right), \text{ for} \tag{16}$$

$$\mu_{U_P}(T) = \mu_{U_A}(T) = \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} T \tag{17}$$

$$\sigma^2_{U_P}(\theta) = \beta^2_{U_A}(\theta) \cdot \sigma^2_{U_A}(\theta) = \frac{\eta(\theta)^2}{\eta(\theta) + \sigma_B^2}. \tag{18}$$

The means μ_{U_P} and μ_{U_A} are the same because the agent's bias is uncorrelated with the test result; the average principal and agent utilities across applicants at a test result are both equal to the average quality. Then each unit of higher utility for the agent translates into $\beta_{U_A}(\theta)$ higher utility for the principal, so the principal's variance is scaled by $\beta^2_{U_A}(\theta)$. Recall that $0 < \beta_{U_A}(\theta) < 1$, implying that $0 < \sigma_{U_P}(\theta) < \sigma_{U_A}(\theta)$.

[29]In this section I write η as a function of θ to emphasize its dependence on the agent's type, and similarly for some other terms such as β_T and β_{U_A}.

Let Z indicate the agent utility *z-score* of an applicant, relative to that applicant's test result:

$$Z \equiv \frac{U_A - \mu_{U_A}(T)}{\sigma_{U_A}(\theta)}. \tag{19}$$

The z-score captures how much the agent likes an applicant, controlling for the public information. Since U_P is increasing in U_A, this term also describes the applicant's relative quality for the principal (up to distinguishability). Of course, while an applicant with a high z-score and low test result is good relative to that test result, this applicant might give lower utility to the agent and principal than an average applicant (z-score of 0) who has a high test result.

I now rewrite the principal and agent utilities in terms of T and Z:

$$U_A = \mu_{U_A}(T) + \sigma_{U_A}(\theta) \cdot Z = \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} \cdot T + \sigma_{U_A}(\theta) \cdot Z$$
$$U_P(T, U_A) = \mu_{U_P}(T) + \sigma_{U_P}(\theta) \cdot Z = \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} \cdot T + \sigma_{U_P}(\theta) \cdot Z.$$

For a given set of k hired applicants, let $\tau \equiv \mathbb{E}[T|\text{Hired} = 1]$ be the average test score and let $\zeta \equiv \mathbb{E}[Z|\text{Hired} = 1]$ be the average agent utility z-score. Taking expectation over the expressions above, the payoffs to the agent and principal from a set of hired applicants are

$$V_A(\tau, \zeta; \theta) = \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} \cdot \tau + \sigma_{U_A}(\theta) \cdot \zeta \tag{20}$$
$$V_P(\tau, \zeta; \theta) = \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} \cdot \tau + \sigma_{U_P}(\theta) \cdot \zeta. \tag{21}$$

The players' payoffs have been reduced to increasing linear functions of two moments: the average test score τ and the average z-score ζ. The agent's preferences over (τ, ζ) depend on the agent's type $\theta = (\sigma_S^2, \sigma_B^2)$ through the induced value $\sigma_{U_A}(\theta)$, and likewise the principal's preferences through $\sigma_{U_P}(\theta)$. Lemma 2 looks at how σ_{U_A} and σ_{U_P} vary with the two components of the type.

Lemma 2.

1. *$\sigma_{U_A}(\sigma_S^2, \sigma_B^2)$ and $\sigma_{U_P}(\sigma_S^2, \sigma_B^2)$ both decrease in σ_S^2.*

2. $\sigma_{U_A}(\sigma_S^2, \sigma_B^2)$ *increases in* σ_B^2 *and* $\sigma_{U_P}(\sigma_S^2, \sigma_B^2)$ *decreases in* σ_B^2.

3. *Across* $\theta = (\sigma_S^2, \sigma_B^2) \in \mathbb{R}_{++}^2$, *the image of* $\sigma_{U_A}(\theta)$ *is* $\mathbb{R}_{++}$. *Given* $\tilde{\sigma}_{U_A} \in \mathbb{R}_{++}$, *the image of* $\sigma_{U_P}(\theta)$ *over* θ *satisfying* $\sigma_{U_A}(\theta) = \tilde{\sigma}_{U_A}$ *is the interval* $\left(0, \min\left\{\tilde{\sigma}_{U_A}, \frac{\sigma_Q^2\sigma_T^2}{\sigma_Q^2+\sigma_T^2}\frac{1}{\tilde{\sigma}_{U_A}}\right\}\right)$.

Part 1 confirms that the effect of making the agent more informed (lower σ_S^2) is to increase both σ_{U_A} and σ_{U_P}, putting more weight on utility z-scores. A more informed agent has higher variance of utilities across applicants at a given test score, and these utility differences become more meaningful to the principal as well. Part 2 shows that as the agent's bias σ_B^2 increases, the agent and principal variances move in opposite directions: σ_{U_A} increases while σ_{U_P} declines. The agent's utilities become more spread out as her biases grow. But because this dispersion is driven by idiosyncratic factors, the principal infers a smaller change to his own utilities from a one standard deviation change in agent utilities.

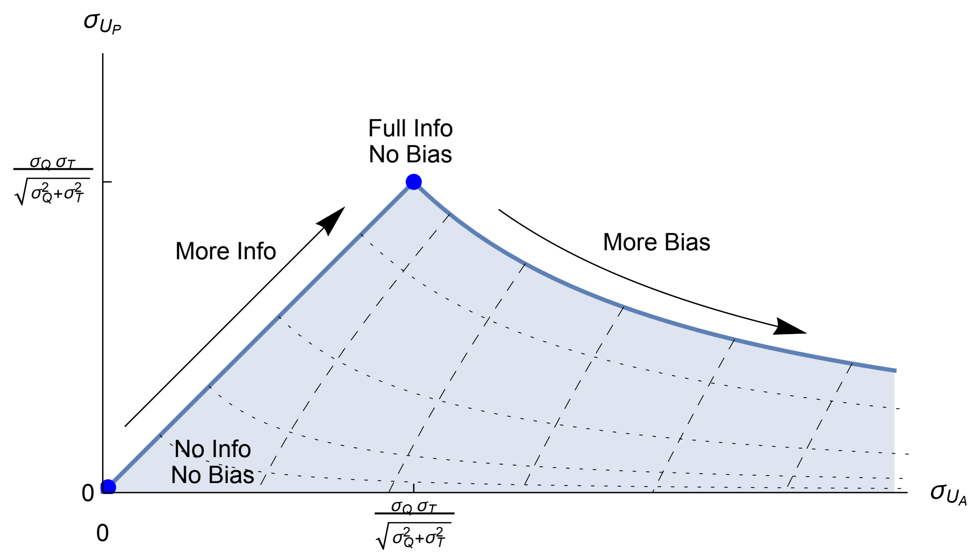
Part 3 describes the set of possible pairs of σ_{U_A} and σ_{U_P} across all values of θ; see Figure 4. To interpret the upper bound for σ_{U_P} at a given σ_{U_A}, first note that if the agent has no information ($\sigma_S^2 \to \infty$) and no bias ($\sigma_B^2 \to 0$), then $\sigma_{U_P} = \sigma_{U_A} = 0$. Improving information moves us up the $y = x$ line in $(\sigma_{U_A}, \sigma_{U_P})$-space. The maximum possible value of σ_{U_P} is achieved when the agent has no bias ($\sigma_B^2 \to 0$) and perfect information ($\sigma_S^2 \to 0$), with $\sigma_{U_P} = \sigma_{U_A} = \frac{\sigma_Q\sigma_T}{\sqrt{\sigma_Q^2+\sigma_T^2}}$. The value of σ_{U_A} can then be increased without bound by increasing the bias, but increasing the bias lowers σ_{U_P} at a rate of $\frac{1}{\sigma_{U_A}}$.

4.3 Rewriting the contracting space

Players' preferences over average test scores τ and average z-scores ζ depend on the agent's type θ through Equations (20) - (21). Given a contract and given her type, the agent will choose the message that maximizes (20). In other words, any contract reduces to a set of possible (τ, ζ) from which the agent may choose. What pairs of average test scores τ and average z-scores ζ are possible?

As a first step, for $x \in (0, 1)$, let $R(x)$ denote the expected value of the top x quantiles of a standard normal distribution. That is, $R(x)$ is the mean of a standard normal that is truncated below at a point r such that $x = 1 - \Phi(r)$. Letting ϕ be the

Figure 4: The region of possible $(\sigma_{U_A}(\theta), \sigma_{U_P}(\theta))$.



The shaded region shows the possible values of $\sigma_{U_A}(\theta)$ and $\sigma_{U_P}(\theta)$ across $\theta = (\sigma_S^2, \sigma_B^2) \in \mathbb{R}^2_{++}$. Increasing information (reducing σ_S^2) moves the values up and right in the region; see the dashed curves. Increasing the bias σ_B^2 moves the values down and right; see dotted curves.

pdf of a standard normal and Φ^{-1} the inverse cdf, standard results imply that

$$R(x) = \frac{\phi(\Phi^{-1}(1-x))}{x}. \tag{22}$$

The function $R(x)$ decreases from infinity to 0 as x goes from 0 to 1.

Lemma 3. *Let $\overline{W} \subseteq \mathbb{R}^2$ be defined as*

$$\overline{W} \equiv \left\{ (\tau, \zeta) \mid \frac{\tau^2}{R_T^2} + \frac{\zeta^2}{R_Z^2} \leq 1 \right\}, \text{ with} \tag{23}$$

$$R_T \equiv \sqrt{\sigma_Q^2 + \sigma_T^2} \cdot R(k) \tag{24}$$

$$R_Z \equiv R(k). \tag{25}$$

There exists a set of k applicants yielding average test scores and z-scores (τ, ζ) if and only if $(\tau, \zeta) \in \overline{W}$.

That is, the set $\overline{W}$ of possible (τ, ζ) is an ellipse centered at $(0, 0)$ with principal axes R_T and R_Z; see Figure 5. For intuition, recall that the empirical distribution of test scores is normal with mean 0 and variance $\sigma_Q^2 + \sigma_T^2$, while the distribution of z-scores at every test result is normal with mean 0 and variance 1. The highest possible average test score for any set of k applicants comes from selecting the k applicants with the highest test results (a step function acceptance rate), yielding $\tau = R_T$ and $\zeta = 0$. The highest possible average z-score comes from selecting the k applicants with the highest z-scores (a constant acceptance rate function), yielding $\tau = 0$ and $\zeta = R_Z$. We get the opposite points if we select applicants with the lowest test results or the lowest z-scores. The boundary connecting these four extreme points is an ellipse, in some sense following the elliptical shape of the joint normal distribution.

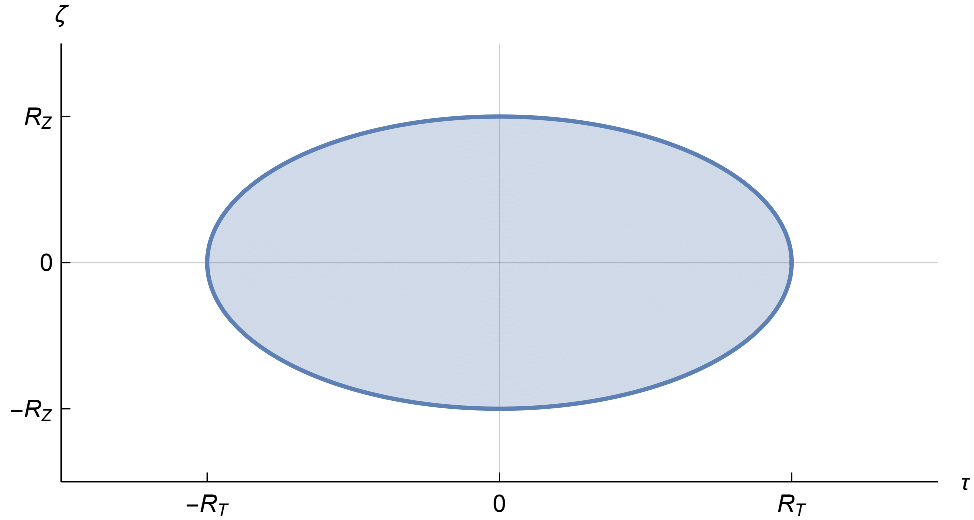
As previously discussed (Proposition 3), if a contract asks an agent to select any set of applicants subject only to a given average test score, that contract induces a normal CDF acceptance rate. Such a contract can now be interpreted as one which restricts τ but not ζ. Given that the agent's payoff is increasing in ζ, we see that the agent must be choosing ζ on the upper frontier of $\overline{W}$. In other words, the (τ, ζ) values on the upper boundary of the ellipse in Figure 5 are those induced by deterministic contracts with normal CDF acceptance rates of the form $\alpha(T) = \Phi(\gamma_T T - \gamma_0)$. Applicant pools with $\tau > 0$ correspond to $\gamma_T > 0$; the value $(\tau, \zeta) = (0, R_Z)$ is achieved by a constant acceptance rate, $\gamma_T = 0$; and $\tau < 0$ is achieved by $\gamma_T < 0$.

Any contract can be characterized as some subset $W \subseteq \overline{W}$ of feasible (τ, ζ) pairs. The contract can specify any (measurable) subset of possible τ in $[-R_T, R_T]$, since test scores are observable. Then, for each allowed τ in W, there is some range of possible ζ.[30] Monotonicity – Observation 1 part 2 – implies that the highest possible ζ at each allowed τ must be weakly positive. (Otherwise the agent could "permute" her report, listing her less preferred applicants as more preferred, to flip the sign of ζ.) And for any chosen τ the agent would always pick this highest ζ. For the purposes of contracting, then, we can restrict attention to the subset of $\overline{W}$ with $\zeta \geq 0$; graphically, the upper half of the ellipse in Figure 5.

The principal's contracting problem can now be stated as a choice of W, a subset of (the top half of) $\overline{W}$. Given W, the agent observes $\sigma_{U_A}(\theta)$ and chooses $(\tau, \zeta) \in W$ to maximize V_A from (20). The principal chooses the set W to maximize $\mathbb{E}_{\theta \sim G}[V_P]$ from (21), taking into account predictions of the agent's behavior at each type.

[30]The average test score of any set of selected applicants is directly contractible. The possible z-scores at a given average test score can be inferred from the rules of the contract.

Figure 5: The space of feasible (τ, ζ) values, $\overline{W}$.



4.4 Projecting the type space to one dimension

Equations (20) and (21) show that the principal and agent preferences over (τ, ζ) depend on θ only through the standard deviation terms $\sigma_{U_A}(\theta)$ and $\sigma_{U_P}(\theta)$. Indifference curves are downward sloping and linear in (τ, ζ)-space. A higher value of the respective standard deviation leads to a higher weight on ζ relative to τ, implying flatter indifference curves. Because $\sigma_{U_P}(\theta) < \sigma_{U_A}(\theta)$, the agent has flatter indifference curves than does the principal. The ideal point for each player is on the upper-right frontier of the ellipse $\overline{W}$, defined in Lemma 3 and illustrated in Figure 5. Due to her flatter indifference curves, the agent's ideal point has a higher average z-score ζ and a lower average test score τ than the principal's.

The agent's utility over (τ, ζ) depends on θ only through the one-dimensional statistic $\sigma_{U_A}(\theta)$; we can equivalently write $V_A(\tau, \zeta; \theta)$ as $V_A(\tau, \zeta; \sigma_{U_A}(\theta))$. Hence, the principal can never separate any two agent types θ with the same $\sigma_{U_A}(\theta)$. They choose the same (τ, ζ) given any contract.[31] So in solving for the optimal contract, it is without loss of generality to average the principal's payoffs across types θ with

[31] I impose the standard contracting assumption that, if the agent is indifferent between two (τ, ζ) outcomes, she will break her indifference in the principal's favor. Since $\sigma_{U_A}(\theta) > \sigma_{U_P}(\theta)$ for all θ, we see from the payoff expressions (20) and (21) that this tie-breaking rule always chooses the outcome with higher τ and lower ζ. Hence, agent types with the same $\sigma_{U_A}(\theta)$ act identically even at possible indifferences. (Under the contract I derive in Proposition 5, there will actually be no such indifferences.)

the same $\sigma_{U_A}(\theta)$. Formally, I now replace the principal's objective function $V_P(\tau, \zeta; \theta)$ from (21), the payoff given an agent of type θ, with $\hat{V}_P(\tau, \zeta; \sigma_{U_A}(\theta))$, the principal's subjective expectation of $V_P(\tau, \zeta; \theta)$ conditional on $\sigma_{U_A}(\theta)$:

$$V_A(\tau, \zeta; \sigma_{U_A}(\theta)) = \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} \cdot \tau + \sigma_{U_A}(\theta) \cdot \zeta \tag{26}$$

$$\begin{aligned}\hat{V}_P(\tau, \zeta; \sigma_{U_A}(\theta)) &\equiv \mathbb{E}_{\theta \sim G}[V_P(\tau, \zeta; \theta) | \sigma_{U_A}(\theta)] \\ &= \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} \cdot \tau + \mathbb{E}_{\theta \sim G}[\sigma_{U_P}(\theta) | \sigma_{U_A}(\theta)] \cdot \zeta \\ &= \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} \cdot \tau + \hat{\sigma}_{U_P}(\sigma_{U_A}(\theta)) \cdot \zeta, \end{aligned} \tag{27}$$

for $\hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A}) \equiv \mathbb{E}_{\theta \sim G}[\sigma_{U_P}(\theta) | \sigma_{U_A}(\theta) = \tilde{\sigma}_{U_A}]$.

The function $\hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A})$ describes the principal's belief on the expectation of $\sigma_{U_P}(\theta)$ across all agent types with $\sigma_{U_A}(\theta) = \tilde{\sigma}_{U_A}$. To understand this function, it may help to recall that for any fixed $\tilde{\sigma}_{U_A}$, the range of possible $\sigma_{U_P}(\theta)$ across types such that $\sigma_{U_A}(\theta) = \tilde{\sigma}_{U_A}$ is described by Lemma 2 part 3. The average across all such types lies in the same range. Graphically, then, $\hat{\sigma}_{U_P}$ is some function mapping realizations of $\sigma_{U_A}(\theta)$ into the shaded region of Figure 4.

We have now effectively projected the type space from the two-dimensional θ to the one-dimensional $\sigma_{U_A}(\theta)$. The two-dimensional distribution G over the type θ determines both the distribution of the projected type $\sigma_{U_A}(\theta)$ and the function $\hat{\sigma}_{U_P}(\cdot)$ summarizing the principal's preferences at a given realization of $\sigma_{U_A}(\theta)$. Let H denote the cdf of $\sigma_{U_A}(\theta)$ induced by $\theta \sim G$.

4.5 Projecting the action space to one dimension plus money burning

Focus on the top half of the ellipse $\overline{W}$ of Figure 5, the values of (τ, ζ) that may be induced by an agent's choices in a contract. Any pool of applicants with (τ, ζ) off of the upper-right frontier of the (half-)ellipse – an acceptance rate that is not of the form $\Phi(\gamma_T T - \gamma_0)$, for $\gamma_T \geq 0$ – is dominated. There is another pool of k applicants with strictly higher τ at the same ζ that improves the payoff of both players. It is as if there is a one-dimensional action space along this upper-right frontier, plus the possibility of joint "money burning" that hurts both players. Let us make that formal.

For $\zeta \in [0, R_Z]$, define $\bar{\tau}(\zeta)$ as the maximum possible τ given ζ, from Lemma 3:

$$\bar{\tau}(\zeta) \equiv R_T \cdot \sqrt{1 - \frac{\zeta^2}{R_Z^2}} = \sqrt{(\sigma_Q^2 + \sigma_T^2) \cdot (R(k)^2 - \zeta^2)}. \tag{28}$$

The minimum possible τ given ζ is $-\bar{\tau}(\zeta)$.

Now take any (τ, ζ) in the upper half of the ellipse. From (26) and (27), the agent and principal payoffs can be written as

$$V_A(\tau, \zeta; \sigma_{U_A}(\theta)) = \left[\frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} \cdot \bar{\tau}(\zeta) + \sigma_{U_A}(\theta) \cdot \zeta\right] - \delta \tag{29}$$

$$\hat{V}_P(\tau, \zeta; \sigma_{U_A}(\theta)) = \left[\frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} \cdot \bar{\tau}(\zeta) + \hat{\sigma}_{U_P}(\sigma_{U_A}(\theta)) \cdot \zeta\right] - \delta \tag{30}$$

$$\text{for } \delta = \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} \cdot (\bar{\tau}(\zeta) - \tau).$$

The bracketed terms give the payoff from an applicant pool with the same ζ, but with τ projected to $\bar{\tau}(\zeta)$ on the right edge of the ellipse. We then subtract the "money burning cost" of $\delta \geq 0$, the loss from taking τ below rather than equal to $\bar{\tau}(\zeta)$.

Importantly, the money burning cost δ is the same for both players, and does not depend on the agent's type θ. This will mean that it fits the framework of one-dimensional delegation with money burning developed in Amador and Bagwell (2013). Instead of thinking about an applicant pool as having payoff-relevant moments τ and ζ, we can equivalently think about it as having payoff-relevant moments ζ and δ. An outcome of a contract corresponds to an average z-score $\zeta \in [0, R_Z]$ along with a level of money burning $\delta \geq 0$.[32]

4.6 Optimal contracts

We can now formally translate the current model into a one-dimensional delegation model. Treat ζ as a one-dimensional "action" to be taken, and allow for the possibility

[32] Given ζ, the money burning cost δ cannot exceed $2\bar{\tau}(\zeta)\frac{\sigma_Q^2}{\sigma_Q^2+\sigma_T^2}$. I will focus on contracts in which money burning is not used even when any $\delta \geq 0$ is feasible, so the upper limit of δ will not bind.

of required money burning $\delta(\zeta) \geq 0$ when action ζ is taken. Payoffs over actions are determined by a one-dimensional "state," $\sigma_{U_A}(\theta)$. The agent is biased towards higher actions than the principal: her ideal ζ is larger for every realization of $\sigma_{U_A}(\theta)$, because $\hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A}) < \tilde{\sigma}_{U_A}$ for all $\tilde{\sigma}_{U_A}$. This is the one-dimensional delegation setting in which – under appropriate regularity conditions – *action ceilings* are often found to be optimal. Specifically, Amador et al. (2018) provides regularity conditions on utility functional forms and distributions to guarantee that a ceiling on actions without money burning is optimal. Their results imply conditions on H and $\hat{\sigma}_{U_P}(\cdot)$ – implicitly, conditions on G – guaranteeing that an optimal contract can be expressed as a choice over any ζ less than or equal to a ceiling. Money burning is identically 0, meaning that given ζ the agent chooses $\tau = \bar{\tau}(\zeta)$.

More meaningfully, a *ceiling* on the unobservable ζ is exactly equivalent to a *floor* on the observable average test score τ of accepted applicants. In either case, the agent picks (τ, ζ) from an interval on the upper-right frontier of the ellipse $\overline{W}$. See Figure 6.

We can also interpret this contract as specifying a menu of acceptance rate functions. As we have seen, when given the freedom to choose any applicants subject to a restriction on average test scores, the agent's picks generate a normal-CDF acceptance rate of the form $\Phi(\gamma_T T - \gamma_0)$. A floor on τ corresponds to a floor on the coefficient γ_T. Proposition 5 gives the formal result.[33]

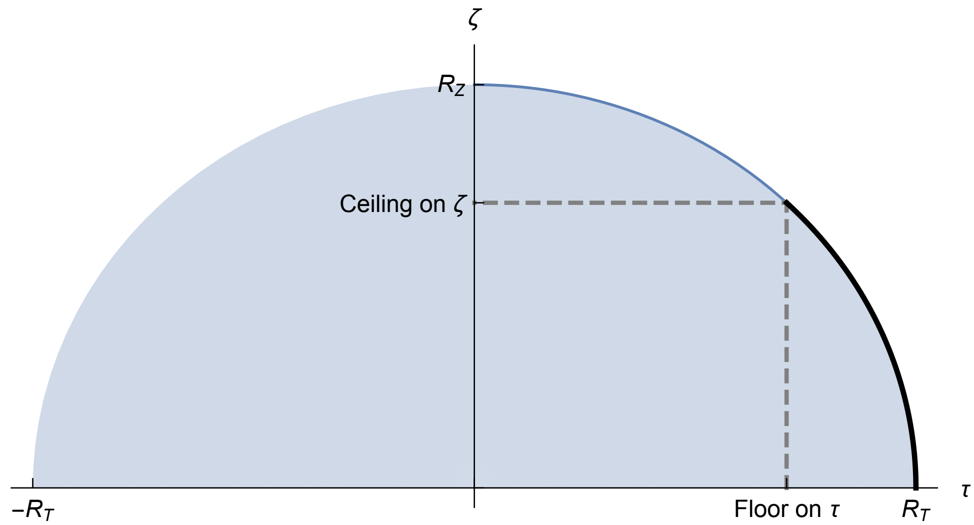
Proposition 5. *Let the distribution H have continuous pdf h over its support, with the support a bounded interval in $\mathbb{R}_+$, and let $\hat{\sigma}_{U_P}(\cdot)$ be continuous over the support. If $H(\tilde{\sigma}_{U_A}) + (\tilde{\sigma}_{U_A} - \hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A}))h(\tilde{\sigma}_{U_A})$ is nondecreasing in $\tilde{\sigma}_{U_A}$, then the optimal contract is deterministic and can be characterized in either of the following two ways:*

1. *The agent is given a floor on the steepness of the acceptance rate function. She may select any k applicants she wants as long as the induced acceptance rate $\alpha(T)$ is of the form $\Phi(\gamma_T T - \gamma_0)$, with γ_T at or above some specified level $\Gamma > 0$.*
2. *The agent is given a floor on the average test score. She may select any k applicants she wants, subject to the average test score of hired applicants τ being at or above some specified level $\kappa > 0$.*

As in Amador and Bagwell (2013) and other papers on one-dimensional delegation,

[33] Proposition 9 in Appendix E.3 derives these same contract forms as optimal in a model combining the idiosyncratic biases of the normal specification with the systematic biases of the two-factor model.

Figure 6: The contract as a floor on the average test score.



Here I highlight on the upper-half of the ellipse $\overline{W}$ from Figure 5. An agent of any type values both τ and ζ positively. So if the agent is given a ceiling on ζ, she chooses some (τ, ζ) point on the thick black curve on the upper-right frontier, and has full flexibility among these points. The agent acts identically if given a corresponding floor on τ.

the floor (Γ or κ) is set to a level that is correct, on average, for the agents who are bound by the floor.[34] It is possible that the floor always binds – that all agent types choose the same average test score of κ. Or the principal might "screen" agent types by setting the floor to be low enough that agents sometimes exceed it.[35]

Returning to the intuition behind the possible benefit of screening, there are two

[34]By itself, the first-order condition on how the floor is set only implies that there is no benefit from making the minimum acceptance rate steeper or shallower in the class of normal CDFs; it does not imply that there is no benefit from pointwise increases or decreases in the acceptance rate at individual test results that take us outside the normal CDF class. However, this latter claim also holds. To see why, consider applying an arbitrary newly proposed acceptance rate function (which still accepts k applicants) to the set of agents who are currently bound at the floor. This new acceptance rate function induces some average test score and average z-score (τ, ζ) in the ellipse $\overline{W}$ (see Lemma 3) – the same pair (τ, ζ) for each of the agents. Any (τ, ζ) is weakly dominated for the principal by some point on the frontier, induced by a normal CDF acceptance rate; and the current normal CDF acceptance rate has been established to be preferred to any other.

[35]Call the principal's ex ante preferred test score the one he would set if he were restricted to giving the agent a (no-screening) contract that fixed the average test score. By Lemma 2 of Amador et al. (2018), the floor in the contract of Proposition 5 will be always-binding if and only if the agent's highest possible ideal average test score over the type support is below the principal's ex ante preferred test score. In this case the principal would set the floor at his ex ante preferred level and the agent would always want a lower score.

One sufficient condition for a non-binding floor is that the support of H extends to $\tilde{\sigma}_{U_A} = 0$, in which case any floor would be nonbinding for those agents with $\sigma_{U_A}(\theta)$ small enough.

reasons why an agent would prefer lower average test scores: because she has better information, or because she is more biased. As discussed above in Section 3.3 and Lemma 2, the principal and agent are aligned with respect to information σ_S^2 and misaligned with respect to bias σ_B^2. If an agent wants lower average test scores because she is more informed, the principal also wants lower test scores; if an agent wants lower test scores because she is more biased, the principal wants higher test scores. Hence, any benefit of screening – of allowing the agent some flexibility over the average test score – would seem to come from the principal's uncertainty about the agent's information, not about her bias. Let us formalize this intuition by considering uncertainty about only one of bias or information at a time.

Commonly known information, uncertain bias.

Proposition 6. *Suppose that the agent's information level σ_S^2 is commonly known. Then the optimal contract can be characterized in either of the following two ways:*

1. *The agent is allowed to choose any k applicants as long as the induced acceptance rate $\alpha(T)$ is $\Phi(\gamma_T T - \gamma_0)$, with γ_T equal to some specified level $\Gamma > 0$.*
2. *The agent may select any k applicants as long as the average test score of hired applicants τ is equal to some specified level $\kappa > 0$.*

Proposition 6 states that when the agent's bias is uncertain but her information is commonly known, the principal doesn't screen across agent types and simply fixes the acceptance rate function, or the average test score, in advance. Additional distributional assumptions such as those in Proposition 5 are not needed. (Once again, the agent interprets this equality constraint as a floor: given flexibility, she would choose weakly flatter acceptance rate functions and lower average test scores.)

In fact, Proposition 6 follows from the slightly stronger result of Proposition 10 in Appendix G. Proposition 10 shows that there will likewise be no screening if the principal's minimum ideal τ over all possible types is above the agent's maximum ideal τ – that is, if the maximum $\hat{\sigma}_{U_P}(\sigma_{U_A}(\theta))$ in the support is below the minimum $\sigma_{U_A}(\theta)$. For instance, the conclusions of Proposition 6 apply any time $\hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A})$ is always decreasing in $\tilde{\sigma}_{U_A}$. Commonly known information is sufficient to imply that $\hat{\sigma}_{U_P}$ is decreasing (as seen in Lemma 2 part 2 and in Figure 4) but is not necessary.

Commonly known bias, uncertain information. In the reverse case with commonly known bias but uncertain information, there is a potential benefit from flexibility. I still require the distributional assumptions of Proposition 5 in order to derive the optimal mechanism as a (possibly binding) floor on steepness or the average test score. But I can give a slightly simpler sufficient condition for these assumptions to hold.

Lemma 4. *Let bias σ_B^2 be commonly known, and let the distribution H have continuous pdf h over its support, with the support a bounded interval in $\mathbb{R}_+$. If $h(\tilde{\sigma}_{U_A})$ is increasing in $\tilde{\sigma}_{U_A}$ over the support, then the hypotheses of Proposition 5 are satisfied: $\hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A})$ is continuous and $H(\tilde{\sigma}_{U_A}) + (\tilde{\sigma}_{U_A} - \hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A}))h(\tilde{\sigma}_{U_A})$ is nondecreasing.*

5 Discussion and extensions

In the Harvard Business Review, McAfee (2013) reports that algorithms have been trained to outperform human experts in making medical diagnoses, in predicting the recidivism of parolees or the outcomes of sports matches, and in many other domains. Algorithms often even improve on experts who first observe the algorithm's suggestions – human decisionmakers introduce biases and add noise. But, as McAfee writes, information from human experts can still be valuable: "things get a lot better when we flip this sequence around and have the expert provide input to the model, instead of vice versa. When experts' subjective opinions are quantified and added to an algorithm, its quality usually goes up."

The current paper can be thought of as studying how subjective opinions should be incorporated into an algorithm when the agent may be biased. I take the machine learning or statistics problem of optimal prediction from a variety of information sources as a black box. Instead, I focus on a strategic issue. If an agent is biased, then any mechanism that allows for her soft information to influence outcomes must be allowing her biases to do so as well. The information that one recovers will depend on the mechanism that is to be used. The contracts I study make the best possible use of the agent's soft information, subject to incentive-compatibility.

I conclude by discussing some issues that have been raised by the above analysis.

5.1 Statistical discrimination and commitment

Becker (1957) and the subsequent literature suggests a test for bias – "taste-based" rather than "statistical" discrimination – when an agent makes a number of binary decisions. Suppose one wants to test for, say, racial discrimination in hiring. After an applicant is hired, we observe his or her race as well as a measure of ex post quality. The hiring manager is demonstrated to be biased if the quality of the marginal applicants – the ones she was just indifferent about hiring – varies across races.

In this paper, I begin with the assumption that an agent is in fact biased and I search for an optimal contract restricting her behavior. Unsurprisingly, though, there is a connection to the problem of testing whether an agent is biased with respect to observables. When the bias and information structure are known, the "upper bound acceptance rule" of Propositions 1 and 2 equalizes marginal qualities across realizations of hard information. In other words, it "de-biases" the agent by inducing her to select applicants in a manner that passes the bias test.

Interestingly, agents are *not* necessarily fully de-biased by the optimal contract when their types are unknown. The screening contract of Proposition 5 proposes a floor on the steepness of the acceptance rate as a function of the test score. First consider the agents who find this floor binding: a mix of those with good information and/or a strong bias. Those with a stronger bias will pick worse marginal applicants at low test scores, and those with better information will pick better marginal applicants at low test scores. Averaged across all agents at the floor, the quality of the marginal applicant is indeed equalized across scores – the agents collectively pass the bias test.[36] Now consider those agents who choose steeper acceptance rates than required: those with poor information and/or a weak bias. They hire their first-best pool of applicants. The marginal applicants they hire at high test scores are of higher quality, on average, than those at low test scores. These agents fail the bias test.

The bias test has a nice connection to the role of the principal's commitment power. If an agent's choices pass the bias test, the principal does not want to adjust acceptance rates ex post. When the bias test is failed, though, the principal is tempted to intervene. He wants the agent to accept more applicants at test results with high marginal quality, and fewer at test results with low marginal quality. In particular,

[36] Recall from footnote 34 that the acceptance rate at the floor is the principal-optimal acceptance rate function for the distribution of agents who are bound by the floor. This fact implies that, averaging across these agents, marginal quality is equalized across test scores.

if an agent facing the contract of Proposition 5 chooses an acceptance rate steeper than what is required, the principal wants to force this agent to go back and choose an even steeper acceptance rate. Of course, the agent would alter her initial choice of applicants if she did not trust the principal to honor the contract; the principal is ex ante better off by committing not to change the rules after the fact.

This commitment logic is analogous to that explored in one-dimensional delegation problems. Suppose an agent is always biased towards an action below what the principal wants, and so the principal sets a floor on the agent's actions. The floor is optimally set so that, across all of the agents who choose an action at the floor, the action is correct on average. But when an agent chooses an action above the floor, the principal would want to intervene and choose an even higher action.

5.2 Hidden test results and multiple agents

If the principal can hide the test results from the agent, he can potentially improve on the contracts that I consider. As a simple illustration, consider the normal specification with known agent type. There, when test results can be hidden, the principal can actually achieve his first-best outcome. One such contract would be as follows. The principal hides the test results T and asks the agent to report her privately observed signal S for every applicant. The principal then calculates the variance of the reported values of S as well as the covariance of S with T. If the agent were in fact to report each S truthfully, the variance of S reports would be $\sigma_S^2 + \sigma_Q^2$ and the covariance between S and T would be σ_Q^2. Any misreport that maintained the variance of S would reduce the covariance. So, if the variance and covariance match the predictions, the principal infers that S has been truthfully reported and the contract implements the principal's first best. If not, the contract chooses applicants uniformly at random. Because the agent prefers some discretion to none, the threat of random selection incentivizes her to report truthfully.

Now return to public test results, but suppose that multiple agents evaluate each applicant. From the perspective of one agent, another agent's soft reports are exactly like hidden test results – they are observed by the principal, but can be hidden from the agent in question. Once again, the ability to hide these reports gives the principal levers to extract additional information from the agent(s).

5.3 Inference from performance data

This paper has been studying the principal's problem of choosing a contract given some specified beliefs about an agent's type. Now suppose that the principal is looking to set the contract for an agent of unknown type by using data from past hires. In particular, the principal has access to ex post performance data – quality realizations – of previously hired applicants. How might this performance data be used to determine contracts going forward?

To remain consistent with the previous analysis, assume that the newly introduced performance data will only used to design new contracts: there is still no way to directly reward agents for hiring applicants who end up performing better.[37]

One simple exercise is as follows. Suppose that (i) we are in the environment described by the normal specification; (ii) an agent had previously been given full discretion to hire her favorite k applicants; and (iii) this agent made these past decisions "myopically" – she selected applicants without realizing that the principal would use her behavior to change the contract in the future. The principal is now setting a new contract after observing the agent's previous acceptance rate, as well as the distribution of realized quality of the accepted applicants, at each test result.

In fact, the principal can infer the agent's type, and thus implement the optimal contract of Proposition 3 going forward, by looking at just two moments of the data. The principal need only calculate the *average test score* of previously accepted applicants, and their *average quality.* In Appendix F, I give an explicit formula for the optimal contract as a function of these two moments. An agent whose previous hires were of higher average quality tends to be less biased and/or more informed, and should be given a flatter contract. Fixing the average quality, an agent who chose a higher average test score should be given a steeper contract.

There are two obvious objections to the above exercise. First, the agent might not act myopically – she might alter her hiring behavior at early periods if she knows that her behavior will affect the contract she is offered in later periods. Second, performance data for one agent's hires might only be available after a long delay. A more reasonable exercise may be to suppose that the principal does not use any

[37] Also maintain the assumption that "principal fundamentals" – the distribution of applicant quality in the population and the informativeness of the test – are known. It is an interesting question in its own right to consider how the principal would best learn about these fundamentals from the data. See the discussion "Inferring σ_Q^2 and σ_T^2" in Appendix F.

individual agent's past performance to update her own contract. Instead, he gathers performance data from a pool of agents and uses the aggregated results to determine contracts.

Sticking with the normal specification, if we take the model literally then the principal can give a sample of agents full discretion; observe the average test score and average performance of each agent; then use this information to infer the joint distribution of bias and information (σ_S^2, σ_B^2) in the population. The principal then imposes the optimal contract for that joint distribution. Alternately, one can take a first-order approach. Proposition 5 highlights a one-dimensional parametric class of contracts, those which give a floor on the average test score or on the steepness of the acceptance rate. The principal can try different floors, look at average performance at each floor, and adjust over time until he finds the floor yielding the highest quality hired applicants.

References

Alonso, R. (2018): "Recruiting and selecting for fit," Working paper.

Alonso, R. and N. Matouschek (2008): "Optimal Delegation," *Review of Economic Studies*, 75, 259–293.

Amador, M. and K. Bagwell (2013): "The theory of optimal delegation with an application to tariff caps," *Econometrica*, 81, 1541–1599.

——— (2016): "Money burning in the theory of delegation," Tech. rep., Mimeo, Stanford University.

Amador, M., K. Bagwell, and A. Frankel (2018): "A Note on Interval Delegation," *Economic Theory Bulletin*, 6, 239–249.

Ambrus, A. and G. Egorov (2015): "Delegation and Nonmonetary Incentives," Working paper.

Armstrong, M. and J. Vickers (2010): "A model of delegated project choice," *Econometrica*, 78, 213–244.

Autor, D. H. and D. Scarborough (2008): "Does job testing harm minority workers? Evidence from retail establishments," *The Quarterly Journal of Economics*, 219–277.

Becker, G. S. (1957): *The economics of discrimination*, University of Chicago press.

Chakraborty, A. and R. Harbaugh (2007): "Comparative cheap talk," *Journal of Economic Theory*, 132, 70–94.

Chan, J. and E. Eyster (2003): "Does banning affirmative action lower college student quality?" *The American Economic Review*, 93, 858–872.

Che, Y., W. Dessein, and N. Kartik (2013): "Pandering to Persuade," *The American Economic Review*.

Einav, L., M. Jenkins, and J. Levin (2013): "The impact of credit scoring on consumer lending," *The RAND Journal of Economics*, 44, 249–274.

Frankel, A. (2014): "Aligned Delegation," *American Economic Review*, 104, 66–83.

——— (2016): "Delegating multiple decisions," *AEJ: Microeconomics*, 8, 16 – 53.

Goltsman, M., J. Hörner, G. Pavlov, and F. Squintani (2009): "Mediation, arbitration and negotiation," *Journal of Economic Theory*, 144, 1397–1420.

Guo, Y. (2014): "Dynamic delegation of experimentation," Tech. rep., mimeo, Yale University.

Hoffman, M., L. B. Kahn, and D. Li (2016): "Discretion in Hiring," Working paper.

Holmström, B. (1977): "On incentives and control in organizations," Ph.D. thesis, Stanford University.

——— (1984): *On the theory of delegation*, Elsevier Science Ltd, In: Bayesian models in economic theory.

Huber, J. D. and C. R. Shipan (2006): "Politics, Delegation, and Bureaucracy," *The Oxford Handbook of Political Economy*, 7, 256.

Jackson, M. O. and H. F. Sonnenschein (2007): "Overcoming Incentive Constraints by Linking Decisions," *Econometrica*, 75, 241–257.

Kováč, E. and T. Mylovanov (2009): "Stochastic mechanisms in settings without monetary transfers: the regular case," *Journal of Economic Theory*, 114, 1373–1395.

Krähmer, D. and E. Kováč (2016): "Optimal Sequential Delegation," *Journal of Economic Theory*, 163, 849–888.

Marshall, A. W. and I. Olkin (2007): *Life distributions*, vol. 13, Springer.

McAfee, A. (2013): "Big Data's Biggest Challenge? Convincing People NOT to Trust Their Judgment," *Harvard Business Review*.

Melumad, N. D. and T. Shibano (1991): "Communication in settings with no transfers," *RAND Journal of Economics*, 22, 173–198.

Myerson, R. B. (1981): "Optimal auction design," *Mathematics of operations research*, 6, 58–73.

Nocke, V. and M. D. Whinston (2013): "Merger policy with merger choice," *The American Economic Review*, 103, 1006–1033.

Tanner, N. (2018): "Optimal Delegation Under Unknown Bias: The Role of Concavity," Working paper.

Wall Street Journal (2014): "Are Workplace Personality Tests Fair?" `http://www.wsj.com/articles/are-workplace-personality-tests-fair-1412044257`.

——— (2015): "Today's Personality Tests Raise the Bar for Job Seekers," `http://www.wsj.com/articles/a-personality-test-could-stand-in-the-way-of-your-next-job-1429065001`.

Appendix

A Common knowledge without alignment

Under alignment up to distinguishability, the monotonicity constraint is not binding in searching for the optimal contract under common knowledge of agent type. This section shows how to apply now-standard "ironing" logic (see, e.g., Myerson (1981)) to solve the optimal contracting problem when alignment up to distinguishability does not hold, and thus when the monotonicity constraint may be binding.

First, let us rewrite the function describing the principal's utility for an applicant. Previously I defined the expected quality in (T, U_A)-space through $U_P(t, u_A) = \mathbb{E}[Q|T = t, U_A = u_A]$. Now define a similar function, l, which tells us the expected quality at a given test result and a *quantile* (rather than a realization) of U_A. Specifically, at each test result t, there is a continuous conditional distribution of U_A which can be rewritten in terms of its quantiles (i.e., by going from a CDF to an inverse CDF): U_A increases in quantile at each t, with quantile 0 at the infimum of the support of $U_A|T = t$ and quantile 1 at the supremum. For $t \in \mathcal{T}$ and $x \in [0, 1]$, let $l(t, x)$ be equal to $U_P(t, u_A)$ for u_A at the x^{th} quantile of the distribution of $U_A|T = t$. Higher x gives higher u_A; alignment up to distinguishability is equivalent to the statement that $l(t, x)$ is weakly increasing in x for every t.

When alignment up to distinguishability fails, there exist test results t for which $l(t, \cdot)$ is not weakly increasing. At these test results, define an *ironed* version of the function l as follows. First integrate l over quantiles to get $L(t,x) \equiv \int_0^x l(t,x')dx'$. Now "iron" L, separately at each test result t, by defining $\overline{L}(t,\cdot)$ to be the convex hull of $L(t,\cdot)$, i.e., the highest convex function that is weakly below $L(t,\cdot)$. Finally, let the ironed l be defined as $\bar{l}(t,x) \equiv \frac{\partial \overline{L}}{\partial x}(t,x)$. The function $\bar{l}$ is defined for almost every $x \in [0,1]$ by convexity of $\overline{L}(t,x)$ in x, and furthermore $\bar{l}(t,x)$ is weakly increasing in x at every t. At any t for which $l(t,\cdot)$ is weakly increasing, it holds that $\bar{l}(t,x) = l(t,x)$ for all x.[38]

To restate, the principal's utility for an applicant with test result t and agent utility quantile x is $l(t,x)$. The ironed principal utility is $\bar{l}(t,x)$. Loosely speaking, we now proceed as if we were solving for UBAR as in Section 3.1, after replacing true principal utilities for each applicant with ironed – and therefore aligned – utilities.

More formally, let us now write an acceptance rule as $\chi : \mathcal{T} \times [0,1] \to [0,1]$, mapping test result and quantile (t,x) into an acceptance probability. As before, implementable acceptance rules must be monotonic – the acceptance rate weakly increases in x – and lead to a total mass of k acceptances. A new ironing constraint also states that any applicants with the same test result and the same ironed principal utility must be given the same acceptance rate:

$$\text{For any } x, x', t \text{ with } \bar{l}(t,x) = \bar{l}(t,x'), \text{ it must hold that } \chi(t,x) = \chi(t,x').$$

Continuing the ironing procedure, the optimizing acceptance rule χ is constructed as follows: accept k applicants so as to maximize the average ironed principal utility of those accepted, $\frac{1}{k}\mathbb{E}[\chi(T,X) \cdot \bar{l}(T,X)]$, for X uniformly drawn on $[0,1]$. The value of this problem is unaffected by the ironing constraint, so that constraint can be costlessly imposed. The constructed acceptance rule will satisfy monotonicity because $\bar{l}(t,x)$ is weakly increasing in x even if $l(t,x)$ is not.

This acceptance rule amounts to first finding the cutoff ironed principal utility level l^c that will lead to accepting k applicants. The acceptance rule then accepts applicants (t,x) with $\bar{l}(t,x) > l^c$ and rejects those with $\bar{l}(t,x) < l^c$. One can choose arbitrary acceptance probabilities in [0,1] when $\bar{l}(t,x) = l^c$ as long as the total share of applicants accepted is k, and as long as we satisfy the ironing constraint at each t.

One way of satisfying this ironing constraint is to choose a single acceptance

[38]If $l(t,x)$ is increasing in x, then $L(t,x)$ is convex in x with $\frac{\partial}{\partial x}L(t,x) = l(t,x)$. Convexity of $L(t,x)$ in x implies that $L(t,x) = \overline{L}(t,x)$ for every x, and therefore that $\bar{l}(t,x) = \frac{\partial}{\partial x}\overline{L}(t,x) = \frac{\partial}{\partial x}L(t,x) = l(t,x)$.

probability in $[0, 1]$ for all applicants (t, x) with $\bar{l}(t, x) = l^c$, where the probability is set so that a total of k applicants are accepted. This does indeed give an optimal (implementable) acceptance rule. It involves randomization at any test result t for which there is an interval of x over which $\bar{l}(t, x) = l^c$.

Alternatively, we can satisfy the ironing constraint by choosing acceptance probabilities for those applicants with $\bar{l}(t, x) = l^c$ that are constant in x (as above) but may vary in t. For instance, it is always possible to order the possibly multidimensional test results in $\mathcal{T}$ in such a manner that the acceptance probability for applicants with $\bar{l}(t, x) = l^c$ is set at 1 for test results t below a threshold t^*; 0 for test results above t^*; and some intermediate level in [0,1] for the single threshold test result t^*.

In this alternative way of satisfying the ironing constraint, there is at most a single test result for which an interior acceptance rate is ever used. That is to say, it is always possible to find an optimal contract which is either deterministic, or in which there are stochastic acceptances at just a single test result. When test results are continuously distributed, of course, behavior at any single test result can be disregarded. So with continuously distributed test scores there exists a deterministic optimal contract.

B Full Discretion under the normal specification

An agent who has full discretion to select k applicants will choose those with U_A above some fixed level – in Figure 2, above a horizontal line. We can solve explicitly for this Full Discretion acceptance rate under the normal specification. Working through the algebra of Section 3.3, but replacing the UBAR acceptance cutoff line $u_A^c(T)$ with a constant in T, under Full Discretion the agent chooses an acceptance rate of $\Phi(\gamma_T^{\text{FD}} T - \gamma_0)$ for

$$\gamma_T^{\text{FD}} = \frac{\sigma_Q^2}{(\sigma_Q^2 + \sigma_T^2)\sqrt{\sigma_B^2 + \eta}}, \tag{31}$$

with η as defined in (11). Putting together (12) and (31), $\gamma_T^* = \gamma_T^{FD} + \frac{\sigma_Q^2 \sigma_B^2}{\eta(\sigma_Q^2 + \sigma_T^2)\sqrt{\sigma_B^2 + \eta}}$, and hence $0 < \gamma_T^{\text{FD}} < \gamma_T^*$. The agent with Full Discretion accepts a greater share of applicants at higher test scores ($0 < \gamma_T^{\text{FD}}$) because she places some weight on quality. But, as discussed in Section 3.3, the Full Discretion outcome is flatter than the principal's optimal contract under knowledge of the agent's type ($\gamma_T^{\text{FD}} < \gamma_T^*$).

We can replicate the comparative statics of Proposition 4 for the Full Discretion outcome rather than the optimal contract, where $\gamma_T^{\mathrm{FD}}\sqrt{\sigma_Q^2+\sigma_T^2}$ is the coefficient on the z-score of the test result.

Proposition 7. *Under the normal specification, the Full Discretion steepness parameter γ_T^{FD} from* (31) *has the following comparative statics and limits:*

1. γ_T^{FD} *is independent of k.*
2. γ_T^{FD} *and* $\gamma_T^{\mathrm{FD}}\sqrt{\sigma_Q^2+\sigma_T^2}$ *decrease in* σ_T^2*, with* $\lim_{\sigma_T^2\to 0}\gamma_T^{\mathrm{FD}}\in(0,\infty)$, $\lim_{\sigma_T^2\to 0}\gamma_T^{\mathrm{FD}}\sqrt{\sigma_Q^2+\sigma_T^2}\in(0,\infty)$*, and* $\lim_{\sigma_T^2\to\infty}\gamma_T^{\mathrm{FD}}=\lim_{\sigma_T^2\to\infty}\gamma_T^{\mathrm{FD}}\sqrt{\sigma_Q^2+\sigma_T^2}=0$.
3. γ_T^{FD} *increases in* σ_S^2*, with* $0<\lim_{\sigma_S^2\to 0}\gamma_T^{\mathrm{FD}}<\lim_{\sigma_S^2\to 0}\gamma_T^{*}$ *and* $\lim_{\sigma_S^2\to\infty}\gamma_T^{\mathrm{FD}}\in(0,\infty)$.
4. γ_T^{FD} *decreases in* σ_B^2*, with* $\lim_{\sigma_B^2\to 0}\gamma_T^{\mathrm{FD}}=\lim_{\sigma_B^2\to 0}\gamma_T^{*}$ *and* $\lim_{\sigma_B^2\to\infty}\gamma_T^{\mathrm{FD}}=0$.
5. γ_T^{FD} *and* $\gamma_T^{\mathrm{FD}}\sqrt{\sigma_Q^2+\sigma_T^2}$ *increase in* σ_Q^2*, with* $\lim_{\sigma_Q^2\to 0}\gamma_T^{\mathrm{FD}}=\lim_{\sigma_Q^2\to 0}\gamma_T^{\mathrm{FD}}\sqrt{\sigma_Q^2+\sigma_T^2}=0$, $0<\lim_{\sigma_Q^2\to\infty}\gamma_T^{\mathrm{FD}}<\lim_{\sigma_Q^2\to\infty}\gamma_T^{*}$*, and* $\lim_{\sigma_Q^2\to\infty}\gamma_T^{\mathrm{FD}}\sqrt{\sigma_Q^2+\sigma_T^2}=\infty$.

There are a few main observations to make. The first is that, relative to Proposition 4, the sign is the same on the derivative with respect to the agent's information σ_S^2 (part 3) but is reversed for the derivative with respect to her bias σ_B^2 (part 4). As discussed in the main text, the principal and agent agree that a more informed agent should have a flatter acceptance rate function. But when the agent is more biased, she prefers flatter acceptance rates, while the principal prefers steeper ones.

Second, part 4 confirms that as the agent's bias disappears, the agent's preferred outcome goes to that of the principal's optimal contract: $\gamma_T^{\mathrm{FD}}\to\gamma_T^{*}$. Without bias, the incentives of the two parties are perfectly aligned.

Third, we now get a clean comparative static on σ_Q^2, whereas its sign under the optimal contract was ambiguous. The Full Discretion acceptance rate gets steeper with respect to test scores (in both absolute and relative terms) when the variance of population quality increases. When this variance goes to zero, the agent's preferences are entirely driven by bias, and so the Full Discretion outcome becomes flat even as the principal-optimal contract becomes infinitely steep.

C Approximated mechanisms in a finite economy

In this section, I explore how one might implement finite approximations of the optimal contract when the agent's type is commonly known. The body of the paper develops two characterizations of the continuum optimal contract, through the acceptance rate function of Proposition 1 and the average score of Proposition 2, which are summarized for the normal specification in Proposition 3 parts 1 and 2. Here, I separately explore approximations of these two contract forms through binned acceptance rates and a minimum average score. This exercise is intended to illustrate how one might put these contract forms into practice, while also giving insight into how the two contract forms compare.

C.1 A finite example

Example primitives. In the continuum model of the paper, there is a mass 1 of applicants, of which k will be accepted. The aggregate distributions of applicant characteristics Q, T, S, and B are given by F_Q, $F_{T|Q}$, $F_{S|T,Q}$, and $F_{B|T,S}$. For this section, I instead suppose that that there is a finite number N of applicants with characteristics drawn iid from these distributions, from which kN will be accepted.

Specifically, let the distributions follow the normal specification, with parameters set to $\sigma_Q^2 = 1$, $\sigma_T^2 = 4$, $\sigma_S^2 = 1$, and $\sigma_B^2 = 1$. The agent will accept a share $k = 1/3$ of the applicants.[39] I will consider finite economies with $N = 12$, 24, 48, or 96 applicants, out of which 4, 8, 16, or 32 will be accepted.

Overview of the mechanisms. In the context of this example, I will go through what I view as the most natural finite approximations of the two contract forms of Proposition 3, with technical details in Appendix C.2. Of course, the exact implementations I consider are certainly not the only possible ways of approximating the continuum contracts for a finite economy. It is doubtless the case that some further tuning could improve payoffs.

To approximate the acceptance rate function of Proposition 3 part 1, I use a *binned acceptance rate* implementation: applicants are divided into bins based on

[39] The numbers were chosen in part to guarantee that in the continuum economy, the optimal contract does considerably better than contracts which give the agent either No Discretion or Full Discretion. See numerical details in Appendix C.2.1.

their test scores, and then the agent chooses a specified number of applicants from each bin. I consider bins that put together a uniform number of applicants M for every possible bin size M that is a factor of N. For instance, at $N = 24$, I consider $M = 1, 2, 3, 4, 8, 12, 24$. (The numbers of applicants N have been chosen as multiples of 12 to allow for many possible bin sizes.) The top M scores are binned together, then the next M, and so forth. The manager then selects some predetermined number of applicants from each bin. At the extremes, $M = 1$ corresponds to No Discretion, in which applicants are selected based only on their test scores; and $M = N$ corresponds to Full Discretion, in which the manager can select any kN applicants. After calculating the principal's expected payoffs for every possible M at some fixed N, we can say that the value M yielding the highest payoff is the preferred bin size.

To approximate the average score contract of Proposition 3 part 2, I use a *minimum average score* implementation: the manager can select any kN applicants whose average test score is sufficiently high. I consider two possible floors for the average score. The first is the "naive floor" that is set in advance at the level that is optimal for the continuum contract. Among other concerns with this naive floor, it may be the case that realized test scores were lower than expected and no set of kN applicants have average scores at or above this level. I correct the naive implementation in an ad hoc manner by supposing that if the floor is not achievable, then the manager must select the applicants with the top kN test scores. The second floor I consider is a "responsive floor" that adjusts the floor up or down when the applicant pool has high or low realized test scores; see Appendix C.2.5 below for the adjustment formula. For instance, under the responsive floor, if all test scores go up by some increment, then the floor itself shifts up by the same amount. Moreover, the applicants with the top kN test scores are guaranteed to have an average score above the floor. I consider the responsive floor to be the preferred floor in every case, but I include analysis of the naive floor for comparison.

Numerical results and takeaways. I numerically simulated the principal's expected payoffs for each of the contract implementations and values of N discussed above. Table 1 summarizes the results. There are three main takeaways that I draw from this table.

First, we see that for both the binned acceptance rate and the minimum average score contracts, payoffs of the preferred implementation improve as we increase the

Table 1: Principal payoffs in finite economies for different contract implementations.

	$N = 12$ Accept 4	$N = 24$ Accept 8	$N = 48$ Accept 16	$N = 96$ Accept 32	Continuum Accept 1/3
Benchmarks					
(*Principal Payoff*)					
No Discretion	.4556	.4712	.4794	.4836	.4878
Full Discretion	.4537	.4693	.4775	.4817	.4859
UBAR upper bound	.5516	.5706	.5805	.5856	.5907
Binned Acc Rate					
(*No Disc to UBAR %*)					
Bin size $M = 1$	0%	0%	0%	0%	
2	40	36	34	33	
3	42	48	52	51	
4	**59**	56	62	60	
6	53	68	72	73	
8		**71**	76	77	
12	-2	57	**80**	83	
16			77	84	
24		-2	63	**85**	
32				81	
48			-2	68	
96				-2	
Min Avg Score					
(*No Disc to UBAR %*)					
Naive floor	32	43	62	77	
Responsive floor	**72**	**82**	**93**	**96**	

Payoffs for the binned acceptance rate and minimum average score contracts are reported as a percentage of the way from the No Discretion to the UBAR payoff for the corresponding value of N, rounded to the nearest percent. For each N, I have bolded the best payoff within each contract type. Binned acceptance rate payoffs for bins of size M with $1 < M < N$ and all minimum average score payoffs are calculated from simulations with 100,000 draws each; the standard error of each such payoff is between .5 and 1.6 percentage points.

number of applicants N. Moreover, as we would expect, with larger N the payoffs seem to be approaching those from the "large numbers" continuum model.

Putting numbers to those points, for each of these finite economies I first derive the principal payoffs from the No Discretion and Full Discretion contracts, which can both easily be implemented (by mechanically selecting the applicants with the top test scores, or by letting the agent choose her favorite applicants). It turns out that under the given parameters, the No Discretion contract does slightly better. Next, I derive the payoffs from the outcomes of the upper bound acceptance rule, which is a theoretical upper bound on the performance of an optimal contract that is exactly achieved in the continuum limit (see Appendix C.2.2). Finally, I simulate the performance of the various binned acceptance rate and minimum average score implementations. These simulations find that when accepting 4 out of 12 applicants, a binned acceptance rate contract already achieves 59% of the benefit of moving from No Discretion to the Upper Bound; a minimum average score contract does even better, achieving 72% of the benefit. Accepting 32 out of 96 applicants, a binned acceptance rate contract achieves 85% while a minimum average score contract achieves 96%. It is not shown in the table, but a minimum average score contract achieves about 99% of this theoretical upper bound when accepting 100 out of 300 applicants.[40]

In other words, the analysis from the continuum model translates well to a finite model of reasonable size. Without solving for exactly optimal contracts in the finite model, I can confirm that these straightforward translations of continuum contracts into finite ones deliver a high share of any possible payoff gains from optimal discretion.

Second, we see that the responsive average score implementation gives higher payoffs than any of the binned acceptance rate implementations. This numerical result does not prove that there would not have been a different finite approximation of "binned acceptance rates" that did even better, of course. But the observation is consistent with the informal argument of Section 3.4 that, in a finite economy, we might expect to prefer some version of a minimum average score contract. Binned acceptance rates impose a number of constraints on the agent while a minimum average score contract links all of the constraints into a single inequality, which can mitigate sampling variation.

Finally, fixing the number of applicants N, we see evidence of the tradeoff over bin size versus number of bins in the binned acceptance rate implementation. As we move from many small bins to fewer large bins by increasing M, the principal payoff tends to increase and then decrease.

[40] Simulating 500,000 draws of $N = 300$ applicants, I found that the principal's payoff from the responsive minimum average score was 98.9% of the way from the No Discretion benchmark to the UBAR upper bound, estimated with a standard error of .1%.

C.2 Implementation Details

C.2.1 Continuum benchmark

For this running example, the (continuum) optimal contract can be expressed in two equivalent ways. First, as an acceptance rate function, the manager accepts a share $\Phi(\gamma_T^* t - \gamma_0^*)$ of applicants with test score t, where from (12) we have $\gamma_T^* = \sqrt{\frac{549}{1280}} \simeq .6549$ and we can numerically calculate that $\gamma_0^* \simeq .7638$ given $k = 1/3$. Second, as an average test score restriction, the manager accepts her favorite k applicants subject to their average test score being at or above $\kappa^* \simeq 2.0143$, which is the average test score of accepted applicants when using the above acceptance rate function. (The floor will be binding.)

Applicants have a mean quality of 0 with a standard deviation of 1. Hence, if applicants were accepted completely randomly, the principal payoff (the expected quality of hired applicants) would be 0. On the other hand, if quality were perfectly observable and the firm could accept the 1/3 of applicants with the highest quality, the principal payoff would work out to 1.0908.

The optimal contract, which implements the upper bound acceptance rule, leads to a principal payoff of .5907. By comparison, under the No Discretion contract in which the applicants with the highest 1/3 of test scores were hired automatically, the principal's payoff would be .4878. Under the Full Discretion contract in which the manager selected her favorite 1/3 of applicants, the principal's payoff would be .4859.

C.2.2 Finite economy benchmarks

Under the No Discretion outcome, the applicants with the top kN out of N test scores are accepted. Under the Full Discretion outcome, the applicants with the top kN out of N realizations of $U_A = \mathbb{E}[Q|T, S] + B$ are accepted. Under the Upper Bound Acceptance Rule (UBAR) outcome, the applicants with the top kN out of N realizations of $U_P(T, U_A) = \mathbb{E}[Q|T, U_A]$ are accepted. Note that UBAR is not necessarily implementable by any incentive compatible contract in the finite economy, but it provides an upper bound for what any incentive compatible contract can achieve.

As we vary N, the No Discretion, Full Discretion, and UBAR benchmark payoffs all scale linearly with each other as some constants times the expectation of the top kN draws out of N from a standard normal distribution. This expectation term increases in N, asymptoting to the expectation of an appropriately truncated normal.[41]

C.2.3 Further notation for the finite economy

Denote the realized test score of applicant $i \in \{1, ..., N\}$ by t^i, where without loss of generality we label scores so that $t^1 \leq t^2 \leq \cdots \leq t^N$.

I now define a term $\alpha^{i,N}$, which is a finite approximation of the continuum contract's acceptance rate of the applicant with the i^{th} lowest test score out of N. Recalling that the distribution of test scores in the population is $\mathcal{N}(0, \sigma_Q^2 + \sigma_T^2)$, the "$i^{th}$ lowest test score out of N" essentially corresponds to test scores ranging from the $\frac{i-1}{N}$ to $\frac{i}{N}$ quantiles of this distribution. This range of quantiles corresponds to test scores in the interval $[x^{i-1,N}, x^{i,N}]$ for

$$x^{i,N} \equiv \sqrt{(\sigma_Q^2 + \sigma_T^2)}\Phi^{-1}\left(\frac{i}{N}\right).$$

In the continuum economy, the optimal acceptance share at test score t is $\Phi(\gamma_T^* t - \gamma_0^*)$. Integrating the acceptance share over this range of quantiles, the finite approximation of the acceptance rate of the i^{th} lowest scoring applicant out of N, $\alpha^{i,N}$, is given by

$$\alpha^{i,N} \equiv \frac{N}{\sqrt{(\sigma_Q^2 + \sigma_T^2)}} \int_{x^{i-1,N}}^{x^{i,N}} \Phi(\gamma_T^* t - \gamma_0^*)\phi\left(\frac{t}{\sqrt{(\sigma_Q^2 + \sigma_T^2)}}\right) dt.$$

C.2.4 Approximate Implementation #1: Binned acceptance rates

Here we approximate the continuum contract through one that imposes a specified acceptance rate on some binned sets of applicants. Fixing the number of applicants N, we will have one parameter to optimize over, the bin size M.

For any bin size M that is a factor of the number of applicants N, now create N/M bins of size M each. Given the notation that applicants are labeled in order of lowest to highest test scores, the first bin $\beta_1^{M,N} = \{1, ..., M\}$ consists of the M lowest-scoring applicants, the j^{th} bin are applicants $\beta_j^{M,N} = \{(j-1)M + 1, ..., jM\}$, and so on up to the highest-scoring applicant bin $\beta_{N/M}^{M,N} = \{N - M + 1, ..., N\}$.

[41] It holds for any distribution that the expectation of the top kN draws of out N is increasing in N. For instance, with $k = 1/2$, the top 1 draw out of 2 from a $U[0, 1]$ distribution has an expectation of $2/3$, while for large N the expectation of the average of the top $N/2$ draws out of N approaches $3/4$.

Fixing M and N, we will determine the number of applicants to accept in bins $\beta_1^{M,N}$, ... , $\beta_{N/M}^{M,N}$ as follows. At the j^{th} such bin $\beta_j^{M,N}$, recall that the finite approximation of the acceptance rate of applicants in this bin is given by the real number $\sum_{i\in\beta_j^{M,N}} \alpha^{i,N}$. We will simply round these values to integers to find the number of applicants to accept at bin $\beta_j^{M,N}$. We do the least rounding possible consistent with making sure that the when we add up across all bins j, we accept a total of kN applicants.

For instance, suppose we will accept 4 out of $N = 12$ applicants using six bins that are each of size $M = 2$. Adding up the $\alpha^{i,N}$ values, we get – prior to rounding – that we should accept .0084 from the first bin, .0853 from the second bin, .2905 from the third bin, .6556 from the fourth bin, 1.1816 from the fifth bin, and 1.7785 from the top bin. Rounding these values, we accept 0 of the applicants from the bottom three bins (the bottom six test scores); 1 of the applicants with the next two higher test scores; 1 of the applicants with the next two higher test scores; and, finally, both of the applicants with the top two test scores. Similarly, when $N = 12$ and $M = 3$, adding up the $\alpha^{i,N}$ values implies that we should accept .0344 from the bottom bin, then .3498 from the next bin, then 1.1721 from the next one, and finally 2.4437 of the top bin. Rounding as little as possible to get to a total of four accepted applicants, we accept 0 of the applicants with the bottom three test scores, 0 of the next three, 1 of the next three, and finally all 3 of the top three scoring applicants.

From Table 1, the best bin size at $N = 12$ is $M = 4$. For $N = 12$ and $M = 4$, the number of applicants accepted from each of the three bins – from lowest scoring to highest – is 0, 1, 3. The best bin size at $N = 24$ is $M = 8$, in which case the number of applicants accepted from each of the three bins is 0, 2, 6. The best bin size at $N = 48$ is $M = 12$, in which case the number of applicants accepted from each of the four bins is 0, 1, 5, 10. Finally, the best bin size at $N = 96$ is $M = 24$, in which case the number of applicants accepted from the four bins is 0, 3, 9, and 20.

C.2.5 Approximate Implementation #2: Minimum average scores

Here we approximate the continuum contract through one that lets the agent accept any applicants she wants subject to a constraint on the minimum average test score of those that she hires. I will introduce two possible test score floors. The naive floor is set in advance, while the responsive floor depends on the realized distribution of test scores.

Let the *naive floor* be equal to the predetermined value of $\kappa^* = 2.0143$. Unfortunately, sometimes the top kN applicants have test scores that actually average less than κ^*; in simulations, this happens about 38% of the time with $N = 12$ and 9% of the time with $N = 96$. When this is the case, we simply use the ad hoc correction that the agent must default to the No Discretion rule of accepting the applicants with the top kN test scores.

To motivate the construction of a responsive floor, recall that the naive floor is determined by solving for the average test score of accepted applicants in the continuum limit. This can be thought of as taking a weighted mean over the theoretical distribution of population test scores, with weights given by the acceptance shares. We will create a new responsive floor that is based on a similar weighted average of the realized rather than theoretical distribution of test scores. We will use the weights that we have already solved for above, the $\alpha^{i,N}$ values that give us the finite approximation of the theoretical acceptance rate for the applicant with the i^{th} lowest test score out of N.

In particular, let the *responsive floor* given N applicants with ordered realized test scores $\{t^1, ..., t^N\}$ be equal to $\sum_i \frac{\alpha^{i,N}}{kN} t^i$. Since the weights $\frac{\alpha^{i,N}}{kN}$ are all in $[0, 1]$ and add up to 1, we know that the responsive floor is always less than the average of the top kN test scores. Hence, it is always possible to find at least one combination of kN applicants whose average test scores are above the responsive floor.

D Additional analysis of systematic biases

D.1 Two-factor model under common knowledge of agent type

Consider the two-factor model of Section 2.3.2 under common knowledge of the agent's type (consisting of F_S and λ). Recalling that $\mathbb{E}[Q_1|T, S] = \mathbb{E}[Q_1|T]$, it holds that $U_A = \mathbb{E}[Q_1|T] + \lambda \mathbb{E}[Q_2|T, S]$. Rearranging,

$$\mathbb{E}[Q_2|T, S] = \frac{U_A - \mathbb{E}[Q_1|T]}{\lambda}.$$

The principal's utility $U_P(T, U_A)$ is therefore

$$U_P(T, U_A) = \mathbb{E}[Q_1 + Q_2|T, U_A] = \mathbb{E}[Q_1|T] + \frac{U_A - \mathbb{E}[Q_1|T]}{\lambda} = \frac{(\lambda - 1)}{\lambda}\mathbb{E}[Q_1|T] + \frac{U_A}{\lambda}.$$

Given the assumption that $\lambda > 0$, the coefficient $\frac{1}{\lambda}$ on U_A is positive. Hence, utilities are aligned up to distinguishability.

The sign of the coefficient $\frac{\lambda-1}{\lambda}$ on $\mathbb{E}[Q_1|T]$ in the expression for U_P depends on the agent's bias term λ. Supposing that test results are real numbers normalized so that higher t yields higher $\mathbb{E}[Q_1|T = t]$, the sign determines whether indifference curves in (T, U_A)-space are sloped upwards or downwards. (Under a further normalization to $\mathbb{E}[Q_1|T = t] = t$, we would get linear indifference curves.)

For the advocate with $\lambda > 1$, there is a positive coefficient on $\mathbb{E}[Q_1|T]$, so the principal has downward-sloping indifference curves – just as with the normal specification. Think of an agent with $\lambda = 2$ who is indifferent between an applicant with a low test result $\underline{t}$ indicating $\mathbb{E}[Q_1|T = \underline{t}] = 0$, and an applicant with a high test score $\bar{t}$ indicating $\mathbb{E}[Q_1|T = \bar{t}] = 1$. For the agent to be indifferent, it must be that the agent observes that $\mathbb{E}[Q_2|T, S]$ is one-half a unit lower for the applicant with the high test result. So the principal prefers the applicant with the high test result: one unit higher Q_1 and one-half unit lower Q_2.

For the cynic with $\lambda < 1$, however, we get upward-sloping indifference curves. If an agent with $\lambda = \frac{1}{2}$ is indifferent between low and high test result applicants, then the principal prefers the applicant with the lower test result: one unit lower Q_1 and two units higher Q_2.

One implementation of the upper bound acceptance rule is to specify the acceptance rate $\alpha(T)$ at $\alpha^{\text{UBAR}}(T)$. This acceptance rate is determined only by the joint distribution of $\mathbb{E}[Q_1|T]$ and $\mathbb{E}[Q_2|T, S]$; the principal selects the k applicants with the highest $\mathbb{E}[Q_1|T] + \mathbb{E}[Q_2|T, S]$ as if T and S were both observable. This contract implements the principal's first-best outcomes, thanks to the assumed absence of idiosyncratic biases. Moreover, while the acceptance rate function α^{UBAR} depends on the agent's information structure, it does not depend on the bias term λ. Restating this point, the same acceptance rate function would be optimal for a principal with *any* beliefs on λ, even if λ is not commonly known.

We can also implement UBAR in the alternative manner in which we fix the average of a score function. For the advocate with $\lambda > 1$, for which indifference curves were downward sloping in $\mathbb{E}[Q_1|T]$, we can choose the score function as $C(T) = \mathbb{E}[Q_1|T]$. For the cynic with $\lambda < 1$ and upward sloping indifference curves, we can choose the score function as $C(T) = -\mathbb{E}[Q_1|T]$. The contract then specifies that $\mathbb{E}[C(T)]|\text{Hired} = 1] = \kappa$, for some κ. The different signs of $C(T)$ based on the magnitude of λ do not actually affect the form of the optimal contract when stated as an equality constraint – we fix the average value of $\mathbb{E}[Q_1|T]$ either way. Rather, the different signs indicate that advocates prefer to push $\mathbb{E}[Q_1|T]$ to be lower than what the principal wants, whereas cynics prefer $\mathbb{E}[Q_1|T]$ to be higher.

D.2 Utility weight on a public signal

One source of systematic bias which is not covered by the two-factor model is that the principal or agent may care directly about the realization of an applicant's hard information. Think about two specific applications.

First, there may be a third party ranking organization (e.g., US News) that rates colleges based on the public hard information of the applicants who matriculate. The college cares about its rankings in addition to the "true quality" of its students. So the school is willing to admit a slightly worse applicant who looks better on paper – a worse essay paired with a better SAT score. The admissions officer doesn't care about rankings, though, and just wants to maximize true student quality.

Second, one or both of the principal and agent may be "prejudiced" or may support "affirmative action" based on an observable characteristic such as race, included as one component of the vector T. This induces a bias – misaligned objectives – if the racial preferences are not perfectly shared by both parties.

To model this, let an applicant's "true quality" be denoted by Q_1. We have distribution $Q_1 \sim F_{Q_1}$ of true quality, with corresponding signal distributions $T \sim F_{T|Q_1}$ and $S \sim F_{S|Q_1,T}$. The expected value of true quality given all information is $\mathbb{E}[Q_1|S,T]$.

Then there is an addition to the principal utility, $Q_{2P}(T)$, and an addition to agent utility, $Q_{2A}(T)$, where $Q_{2P}(\cdot)$ and $Q_{2A}(\cdot)$ are arbitrary functions of the realization of hard information. Marginal utilities for the two players are as follows:

$$
\begin{aligned}
&P: Q_1 + Q_{2P}(T) = Q \\
&A: Q_1 + Q_{2A}(T) = Q + B, \text{ for } B = Q_{2A}(T) - Q_{2P}(T).
\end{aligned}
$$

We see that this form of systematic bias shows up as a relationship between the bias realization B and the hard information.

Utilities are aligned up to distinguishability: at any test result, applicants more preferred by the agent are more preferred by the principal. Formally, given T and $U_A = \mathbb{E}[Q_1|S,T] + Q_{2A}(T)$, we can rearrange to get $\mathbb{E}[Q_1|S,T] = U_A - Q_{2A}(T)$. The induced principal utility $U_P(T, U_A)$ is

$$U_P(T, U_A) = \mathbb{E}[Q_1|S,T] + Q_{2P}(T) = U_A - Q_{2A}(T) + Q_{2P}(T)$$

which is increasing in U_A. Hence, we can apply the results of Section 3 to solve for the optimal acceptance rate function. Just as with the analysis of the two-factor model in Section D.1, this acceptance rate does not depend on the agent's bias function $Q_{2A}(\cdot)$. Likewise, due to the assumed lack of idiosyncratic bias, the optimal contract implements the first-best payoff for the principal.

E Combining idiosyncratic and systematic biases

This section puts together the idiosyncratic biases of the normal specification with the systematic biases of the two-factor model into a *combined model*. I show that the qualitative results for the normal specification in Sections 3.3 and 4 extend to the combined model.

E.1 Setup of the combined model

As in the two-factor model, quality Q in the combined model can be decomposed into two quality factors Q_1 and Q_2, and the test result reveals everything that can be inferred about Q_1: $\mathbb{E}[Q_1|T,S] = \mathbb{E}[Q_1|T]$. The private signal S then gives additional information about quality factor Q_2. Now, adding normally distributed idiosyncratic biases to the payoffs of the two-factor model, assume that principal cares about $Q = Q_1 + Q_2$ and the agent cares about $Q_1 + \lambda Q_2 + \epsilon_B$, for $\epsilon_B \sim \mathcal{N}(0, \sigma_B^2)$ independent of T, S, and with λ and σ_B^2 are in $\mathbb{R}_{++}$. All together, then, conditional on signals T

and S, the principal and agent marginal utilities of hiring an applicant are given by

$$\begin{aligned} &\text{Principal:} && \mathbb{E}[Q|T,S] = \mathbb{E}[Q_1 + Q_2|T,S] = \mathbb{E}[Q_1|T] + \mathbb{E}[Q_2|T,S] \\ &\text{Agent:} && U_A \equiv \mathbb{E}[Q_1 + \lambda Q_2|T,S] + \epsilon_B = \mathbb{E}[Q_1|T] + \lambda\mathbb{E}[Q_2|T,S] + \epsilon_B. \end{aligned}$$

I will now add additional distributional assumptions on the two signals. In particular, rather than specifying the conditional distributions $F_{T|Q}$ and $F_{S|Q,T}$, I will write out the expectations of Q_1 and Q_2 given the signals in linear reduced forms. Let the signal realization spaces $\mathcal{T}$ and $\mathcal{S}$ both be equal to $\mathbb{R}$ and assume that

$$\begin{aligned} \mathbb{E}[Q_1|T] &= T \\ \mathbb{E}[Q_2|T,S] &= rT + S \end{aligned}$$

for some $r \in \mathbb{R}$. Finally, assume that the distribution of S conditional on T (but unconditional on the quality factors) is given by $S|T \sim \mathcal{N}(0, l \cdot \sigma_2^2)$ for $l \in (0,1)$ and $\sigma_2 \in \mathbb{R}_{++}$. Note that, while I do not commit to the details of the updating model that would get us these posteriors, it would be straightforward to "microfound" these reduced form assumptions through appropriate joint-normal priors on the two quality factors and normally distributed signals.[42]

We have introduced five relevant parameters: r, σ_2^2, l, λ, and σ_B^2. Two of these, λ and σ_B^2, are familiar as the systematic bias term of the two-factor model and the idiosyncratic bias term of the normal specification. The interpretation of the other three parameters is as follows. First, conditional on the observation of $T = t$, the distribution of Q_2 has mean rt and variance of σ_2^2. A value $r > 0$ indicates a positive correlation of the two quality factors, and $r < 0$ a negative correlation. The parameter l corresponds to the level of the agent's information on Q_2: a more informative private signal means higher l. An agent who perfectly observed the realization of Q_2 would

[42]Here is one collection of primitives that would give rise to these reduced form distributional assumptions. Take Q_1 and Q_2 to be joint normally distributed, and have T perfectly reveal Q_1 – it has a degenerate distribution at $T = Q_1$. (I have not specified the distribution of T outside of this footnote, but under this assumption the empirical distribution would be normal.) The mean of Q_2, unconditional on other signals, will be linear in T with slope depending on the variances and covariance of Q_1 and Q_2. The agent then receives a private signal equal to Q_2 plus some normally distributed noise (where a higher variance of noise corresponds to less information, and so lower l); normalize the signal S to be the resulting deviation of the posterior belief from the mean. The agent's posterior expectation on Q_2 is normally distributed about the mean with a variance somewhere between 0 (no information) and σ_2^2 (full information).

have $l \to 1$, and one who received no private information would have $l \to 0$.

Putting these assumptions together, we can rewrite the principal and agent marginal utilities as

$$\begin{aligned} &\text{Principal:} && \mathbb{E}[Q_1|T] + \mathbb{E}[Q_2|T,S] = (1+r)T + S \\ &\text{Agent:} && \mathbb{E}[Q_1|T] + \lambda\mathbb{E}[Q_2|T,S] + \epsilon_B = (1+\lambda r)T + \lambda S + \epsilon_B. \end{aligned}$$

In the notation $Q = Q_1 + Q_2$, we have that the agent maximizes the expectation of $Q + B$ for $B = (\lambda - 1)(rT + S) + \epsilon_B$.

The agent's type θ consists of three parameters: $l \in (0,1)$ for information (replacing, but analogous to, σ_S^2 in the normal specification), $\sigma_B^2 \in (0,\infty)$ for idiosyncratic bias, and $\lambda \in (0,\infty)$ for systematic bias.[43] In line with Assumption 1, I take all other parameters to be commonly known.

Conditional on $T = t$ (but unconditional on S or ϵ_B), it holds that $\mathbb{E}[Q|T = t, S]$ and U_A are normally distributed. The marginal distribution of U_A is given by

$$U_A|T = t \sim \mathcal{N}\left(\mu_{U_A}(t), \sigma_{U_A}^2\right), \text{ for} \tag{32}$$

$$\mu_{U_A}(t) = t(1+\lambda r) \tag{33}$$

$$\sigma_{U_A}^2 = \lambda^2 l \sigma_2^2 + \sigma_B^2. \tag{34}$$

The marginal distribution of $\mathbb{E}[Q|T = t, S]$ is normal with mean $t(1+r)$ and variance $l\sigma_2^2$. The covariance of $\mathbb{E}[Q|T = t, S]$ with $U_A|T = t$ is $\lambda l \sigma_2^2$. Hence, we can calculate

[43] The combined model here embeds the normal specification after a notational adjustment. Call S and T the signals in the normal specification, and S' and T' the signals in the combined model. The normal specification maps into the combined model if we take $\lambda = 1$, along with

$$S' = \left(S - T\frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2}\right) \cdot \frac{\sigma_T^2\sigma_Q^2}{\sigma_S^2\sigma_T^2 + \sigma_Q^2\sigma_S^2 + \sigma_Q^2\sigma_T^2} \qquad T' = T \cdot \frac{\sigma_S^2\sigma_Q^2}{\sigma_S^2\sigma_T^2 + \sigma_Q^2\sigma_S^2 + \sigma_Q^2\sigma_T^2}$$

$$\sigma_2^2 = \frac{\sigma_Q^2\sigma_T^2}{\sigma_Q^2 + \sigma_T^2} \qquad l = \frac{\sigma_Q^2\sigma_T^2}{\sigma_S^2\sigma_T^2 + \sigma_Q^2\sigma_S^2 + \sigma_Q^2\sigma_T^2} \qquad r = \frac{\sigma_Q^2\sigma_T^2}{\sigma_S^2(\sigma_Q^2 + \sigma_T^2)}.$$

$U_P(t, u_A) = \mathbb{E}[Q|T = t, U_A = u_A] = \mathbb{E}[\mathbb{E}[Q|T = t, S]|T = t, U_A = u_A]$ as

$$U_P(t, u_A) = \beta_T t + \beta_{U_A} u_A, \text{ for} \tag{35}$$

$$\beta_T = 1 - \frac{\lambda l \sigma_2^2 - r\sigma_B^2}{\lambda^2 l \sigma_2^2 + \sigma_B^2} \tag{36}$$

$$\beta_{U_A} = \frac{\lambda l \sigma_2^2}{\lambda^2 l \sigma_2^2 + \sigma_B^2}. \tag{37}$$

The coefficient β_{U_A} on agent utility is positive, implying that utilities are aligned up to distinguishability. Larger idiosyncratic biases σ_B^2 reduce this coefficient, making beliefs on quality less responsive to agent utilities, but do not affect the sign. Let us look next at the coefficient β_T on test scores. Without idiosyncratic shocks – that is, plugging in $\sigma_B^2 = 0$ – β_T reduces to $\frac{\lambda-1}{\lambda}$ as in the two-factor model (Appendix D.1). Adding idiosyncratic shocks through σ_B^2 pulls the coefficient β_T towards $1+r$. Putting the effects on β_{U_A} and β_T together, we see that increasing the idiosyncratic preference shocks (larger σ_B^2) takes the principal's belief, at any given test score $T = t$ and utility realization $U_A = u_A$, in the direction of $(1 + r)t$ – the estimate of quality given $T = t$, and unconditional on U_A. In the case where there is weakly positive correlation of the two factors ($r \geq 0$), larger idiosyncratic shocks monotonically increase the coefficient β_T. In particular, with $r \geq 0$, the sign of β_T is always positive for advocates ($\lambda > 1$). The sign of β_T can be negative for cynical agents ($\lambda < 1$) but it switches to positive for sufficiently large idiosyncratic biases σ_B^2.

From (35), we see that the distribution of $U_P(t, U_A)$ – that is, expected quality conditional on $T = t$ across realizations of U_A – has mean and variance of

$$\mu_{U_P}(t) = t(1 + r)$$

$$\sigma_{U_P}^2 = \beta_{U_A}^2 \sigma_{U_A}^2 = \frac{(\lambda l \sigma_2^2)^2}{\lambda^2 l \sigma_2^2 + \sigma_B^2}.$$

In the normal specification, the principal and agent had equal mean utilities conditional on test score T. But now the coefficients on T differ if the quality factors are correlated ($r \neq 0$) and the agent has a systematic bias ($\lambda \neq 1$). The coefficient on T for the principal is $1 + r$, compared to the agent's $1 + \lambda r$. When there is positive correlation between the two quality factors ($r > 0$), the mean as a function of test score will have a steeper slope for an advocate agent than for the principal, and a flatter slope for a cynic.

E.2 Contracting with known agent type

Under common knowledge of agent type, we can solve for the optimal policy exactly as in the normal specification and two-factor model. The upper bound acceptance rule sets a cutoff utility u_P^c and accepts all applicants with $U_P \geq u_P^c$. This is implemented by a normal CDF acceptance rate, $\alpha(T) = \Phi(\gamma_T T - \gamma_0)$, at some appropriate steepness $\gamma_T = \gamma_T^{\text{comb}}$. We can solve for this optimal coefficient from the equations for the distribution of $U_P(t, U_A)$ as

$$\gamma_T^{\text{comb}} = \frac{1+r}{\sigma_{U_P}} = \frac{(1+r)\sqrt{\lambda^2 l \sigma_2^2 + \sigma_B^2}}{\lambda l \sigma_2^2}. \tag{38}$$

The acceptance rate is increasing in the test score (positive γ_T^{comb}) even if β_T is negative, as long as $r \geq -1$. This holds because higher quality applicants tend to have higher test scores, so the principal wants to accept more of them. If there is a sufficiently strong negative correlation between quality factors that $r < -1$, then higher quality applicants tend to have lower test scores (they are lower on the first quality factor), and γ_T^{comb} is negative.[44]

Thanks to the linearity of $U_P(T, U_A)$ in both T and U_A, the policy could also be implemented by fixing the average test score of hired workers at some level κ^{comb} (for which I do not provide a formula). Gathering together these observations:

Proposition 8. *Under the combined model with common knowledge of the agent's type, the optimal contract can be implemented in either of the following ways. The agent is allowed to hire any set of k applicants, subject to:*

1. *An acceptance rate function of $\alpha(t) = \Phi(\gamma_T^{\text{comb}} T - \gamma_0)$; or,*
2. *An average test score of accepted applicants equal to some value κ^{comb}.*

[44]While we might expect T to be normally distributed, as it would be under the normal prior/normal signal microfoundation of footnote 42, this result did not impose any assumptions on the distribution of T. The distribution of T affects which γ_0 will set the aggregate share of acceptances to k, but does not affect the coefficient γ_T on test scores in the acceptance rate function.

E.3 Contracting with unknown agent type

With uncertainty over the agent's type $\theta = (l, \lambda, \sigma_B^2)$, we can replicate much of the analysis of Section 4 in solving for the optimal policy. Say that θ follows distribution function G. Going forward, I write μ_{U_A}, σ_{U_A}, and σ_{U_P} as functions of θ. For this analysis assume that the unconditional distribution of test scores T is normally distributed, with mean normalized to 0 and variance of Var_T, as motivated in footnote 42.

As in the main text, define Z as an agent utility z-score for a given applicant, and τ and ζ as the average test score and average z-score for a pool of accepted applicants:

$$Z \equiv \frac{U_A - \mu_{U_A}(T;\theta)}{\sigma_{U_A}(\theta)}$$
$$\tau \equiv \mathbb{E}[T|\text{Hired} = 1]$$
$$\zeta \equiv \mathbb{E}[Z|\text{Hired} = 1].$$

The outcome space in terms of (τ, ζ) is exactly as in Lemma 3, with $R_Z = R(k) = \frac{1}{k}\phi\left(\Phi^{-1}(1-k)\right)$ and $R_T = \sqrt{\sigma_Q^2 + \sigma_T^2}R(k)$. As before, let $\bar{\tau}(\zeta) \equiv R_T \cdot \sqrt{1 - \frac{\zeta^2}{R_Z^2}}$ be the maximum possible τ for a given $\zeta \in [-R_Z, R_Z]$.

When the agent is of type θ, hiring an applicant with test score T and utility z-score Z gives marginal utilities to the agent and principal of

$$U_A = \mu_{U_A}(T;\theta) + \sigma_{U_A}(\theta)Z = (1 + \lambda r)T + \sigma_{U_A}(\theta)Z$$
$$U_P(T, U_A) = \mu_{U_P}(T) + \sigma_{U_P}(\theta)Z = (1 + r)T + \sigma_{U_P}(\theta)Z$$

In terms of τ and ζ, agent and principal payoffs for hiring a pool of applicants are

$$\text{A: } V_A = \sigma_{U_A}(\theta)\zeta + (1 + \lambda r)\tau$$
$$\text{P: } V_P = \sigma_{U_P}(\theta)\zeta + (1 + r)\tau.$$

We see that the agent's behavior depends only on the ratio of $\sigma_{U_A}(\theta)$ to $(1 + \lambda r)$; her problem is equivalent to maximizing $\frac{\sigma_{U_A}(\theta)}{1+\lambda r}\zeta + \tau$, or to maximizing $\sigma_{U_A}(\theta)\frac{1+r}{1+\lambda r}\zeta + (1 + r)\tau$. Define ρ as this coefficient on ζ:

$$\rho(\theta) \equiv \sigma_{U_A}(\theta)\frac{1 + r}{1 + \lambda r} = \sqrt{\lambda^2 l \sigma_2^2 + \sigma_B^2}\frac{1 + r}{1 + \lambda r}.$$

The coefficient $\rho(\theta)$ is a one-dimensional sufficient statistic for the agent's preferences. For any $\tilde{\rho}$, all agent types θ with $\rho(\theta) = \tilde{\rho}$ act identically. Let the distribution of $\rho(\theta)$ induced by $\theta \sim G$ be given by the cdf H.

Because the principal can never distinguish agents with the same $\rho(\theta)$, it is convenient to define the principal's average value of σ_{U_P} across all agent types with $\rho(\theta) = \tilde{\rho}$ as $\hat{\sigma}_{U_P}(\tilde{\rho})$:

$$\hat{\sigma}_{U_P}(\tilde{\rho}) \equiv \mathbb{E}_{\theta \sim G}\left[\sigma_{U_P}(\theta) \mid \rho(\theta) = \tilde{\rho}\right].$$

Now rewrite the principal and agent maximization problems as

$$\text{Agent: } \max \Big(\rho(\theta) \cdot \zeta + (1+r)\bar{\tau}(\zeta)\Big) - \delta \tag{39}$$

$$\text{Principal: } \max \mathbb{E}_{\rho(\theta) \sim H}\left[\Big(\hat{\sigma}_{U_P}(\rho(\theta)) \cdot \zeta + (1+r)\bar{\tau}(\zeta)\Big) - \delta\right] \tag{40}$$

$$\text{for } \delta \equiv (1+r)(\bar{\tau}(\zeta) - \tau). \tag{41}$$

Once again δ represents "money burning" due to taking τ below its maximum possible value. The contract induces a menu of (ζ, δ) from which the agent may select, given her observation of $\rho(\theta)$.

We can now give the analog of Proposition 5.

Proposition 9. *In the combined model, let the distribution H have continuous pdf h over its support, with the support a bounded interval in $\mathbb{R}_+$, and let $\hat{\sigma}_{U_P}(\cdot)$ be continuous over the support. If $H(\tilde{\rho}) + (\tilde{\rho} - \hat{\sigma}_{U_P}(\tilde{\rho}))h(\tilde{\rho})$ is nondecreasing in $\tilde{\rho}$, then the optimal contract can be characterized in either of the following two ways:*

1. *The agent is given a floor on the steepness of the acceptance rate function. She may select any k applicants she wants as long as the induced acceptance rate $\alpha(T)$ is of the form $\Phi(\gamma_T T - \gamma_0)$, with γ_T at or above some specified level $\Gamma > 0$.*

2. *The agent is given a floor on the average test score. She may select any k applicants she wants, subject to the average test score of hired applicants τ being at or above some specified level $\kappa > 0$.*

This result embeds Proposition 5 – up to some changes of notation – when there is no systematic bias, i.e., $\lambda = 1$. In that case the projection of the agent's type $\rho(\theta) = \sigma_{U_A}(\theta)\frac{1+r}{1+\lambda r}$ is exactly just $\sigma_{U_A}(\theta)$. But we also now have a generalization of the conditions under which the simple contract forms from the body of the paper remain optimal even when agents have a commonly known systematic bias $\lambda \neq 1$, or when there is a distribution of λ across agents.

F Inference from performance data

Consider the normal specification, and take σ_Q^2 and σ_T^2 to be commonly known while the agent's type, (σ_S^2, σ_B^2), is not known. (Below, I address how one might also infer σ_Q^2 and σ_T^2 if those were unknown.) I proceed here in a prior-free manner and thus do not specify the principal's prior beliefs over the agent's type.

Let there be two periods over which the agent's type is persistent. In the first period the principal gives the agent a Full Discretion contract in which she chooses k applicants. For each applicant that is hired, the principal observes the public test result T and also the quality Q – the realized performance. Then in the second period the principal uses the first-period data to choose a contract that will select another k applicants.

Assume that the agent selects applicants myopically, i.e., her behavior in the first period maximizes her first period payoff. That is, she has no dynamic consideration for how her behavior affects the contract she will be offered in the future.

In the second period, the principal has access to the acceptance rate as a function of test results, plus the entire distribution of realized qualities for the accepted applicants at each score. I will find that this data is sufficient for the principal to perfectly infer the agent's type, and therefore to set the optimal contract in the second period given the knowledge of her type. Indeed, the principal only needs to look at two moments of the data. Let τ_1 be the average test score of the applicants accepted in the first period, and let ξ_1 be the average realized quality. The principal can calculate the optimal second-period contract from τ_1 and ξ_1. As in Proposition 3 part 2, the contract can be summed up as a requirement that the average test score of accepted applicants in the second period, τ_2, must equal some level κ^*.

Lemma 5. *Given τ_1 and ξ_1, the principal's period-2 optimal contract allows the agent to accept any k applicants with average test score τ_2 equal to κ^*, with*

$$\kappa^* = \frac{R(k)\sigma_Q^2\sqrt{\sigma_Q^2+\sigma_T^2}}{\sqrt{\sigma_Q^2 + \frac{(\xi_1(\sigma_Q^2+\sigma_T^2)-\sigma_Q^2\tau_1)^2}{(\sigma_Q^2+\sigma_T^2)((\sigma_Q^2+\sigma_T^2)R(k)^2-\tau_1^2)}}} \tag{42}$$

and $R(\cdot)$ given by (22). Moreover, over the domain of possible ξ_1 and τ_1, it holds that κ^* decreases in ξ_1 and increases in τ_1.

That is, for any fixed average test score in the Full Discretion first period, better ex post performance of the hired applicants ξ_1 leads to a lower required average test score (a flatter contract, one closer to the agent's preferred outcome) in the second period. On the other hand, an agent who picks a higher average test score τ_1 (steeper contract) in the first period is required to pick a higher average test score (steeper contract) in the second period.

Inferring σ_Q^2 and σ_T^2.

What if the principal fundamentals σ_Q^2 and σ_T^2 will be the same from period 1 to period 2, but the values are not known in advance? In fact, these two parameters can also be inferred from the period-1 Full Discretion data. Their imputed values can then be plugged into the formulas above.

To see this, first define Var_T as the empirical variance of the test score distribution across all applicants. This empirical variance is directly observable in period 1. Under the predictions of the model, Var_T will be equal to $\sigma_Q^2 + \sigma_T^2$.

Next, let $\bar{q}_1(t)$ indicate the average realized period-1 quality of accepted applicants at test score t. Suppose that, under full discretion, a share $\alpha(t)$ of applicants are accepted at this test score. Then the model predicts that

$$\bar{q}_1(t) = \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} t + \sigma_{U_P}(\theta) R(\alpha(t)) = \frac{\sigma_Q^2}{\mathrm{Var}_T} t + \sigma_{U_P}(\theta) R(\alpha(t)).$$

The value of $\sigma_{U_P}(\theta)$ can be inferred from performance data, with the formula given in (54) in the proof of Lemma 5. Plugging in that formula, and replacing all occurrences of $\sigma_Q^2 + \sigma_T^2$ with Var_T, we get:

$$\bar{q}_1(t) = \frac{\sigma_Q^2}{\mathrm{Var}_T} t + \left(\frac{\xi_1 \mathrm{Var}_T - \sigma_Q^2 \tau_1}{\sqrt{\mathrm{Var}_T (\mathrm{Var}_T R(k)^2 - \tau_1^2)}} \right) R(\alpha(t)).$$

Solving this equation for σ_Q^2 gives a separate estimate of σ_Q^2 at each test score t:

$$\sigma_Q^2 = \frac{\bar{q}_1(t) \sqrt{\mathrm{Var}_T (R(k)^2 \mathrm{Var}_T - \tau_1^2)} - R(\alpha(t)) \mathrm{Var}_T \xi_1}{t \sqrt{(R(k)^2 - \tau_1^2)/\mathrm{Var}_T} - R(\alpha(t)) \tau_1}. \tag{43}$$

Of course, under the theoretical model the estimate should be identical at every t. With actual performance data, one would presumably want to take an average or a weighted average of these estimates across all of the test scores. At any rate, given an estimate of σ_Q^2 from (43), we have $\sigma_T^2 = \text{Var}_T - \sigma_Q^2$.

G Proofs

G.1 Proofs for Section 2 and 3

Proof of Lemma 1. Consider two independent random variables X and Y, for which Y has a log-concave distribution. I seek to show that $\mathbb{E}[X|X+Y=z]$ is weakly increasing in z. Conditioning on $T=t$ and interpreting X as $\mathbb{E}[Q|S,T]$, Y as B, and z as u_A will then yield the desired conclusion that $U_P(t,u_A)$ is increasing in u_A. Specifically, these substitutions give us $X+Y = \mathbb{E}[Q|S,T]+B = U_A$, and $\mathbb{E}[X|X+Y=z, T=t] = \mathbb{E}[\mathbb{E}[Q|S,T]|U_A = u_A, T=t] = \mathbb{E}[Q|U_A=u_A, T=t] = U_P(t,u_A)$, where the second-to-last equality holds by the law of iterated expectations.

To show that $\mathbb{E}[X|X+Y=z]$ is weakly increasing in z (for any prior over X), it suffices to show that the distribution of $X+Y|X=x$ satisfies the monotone likelihood ratio property in x, and thus that higher realizations of $X+Y$ indicate higher posteriors on X. By independence of X and Y, it holds that $X+Y|X=x$ follows the distribution of $Y+x$. In other words, it suffices to show that $Y+x$ has monotone likelihood ratio in x. Indeed, log-concavity of Y implies that $Y+x$ has monotone likelihood ratio in x; see, e.g., Marshall and Olkin (2007, Example 2.A.15).

To give some intuition for that final step, indicate the pdf of Y by f_Y and the pdf of $Y+x$ by f_{Y+x}, where $f_{Y+x}(z) = f_Y(z-x)$. The random variable $Y+x$ has monotone likelihood ratio in x if (ignoring zeroes in the denominator) it holds that for all $\overline{z} > \underline{z}$ and $\overline{x} > \underline{x}$,

$$\frac{f_{Y+\overline{x}}(\overline{z})}{f_{Y+\underline{x}}(\overline{z})} \geq \frac{f_{Y+\overline{x}}(\underline{z})}{f_{Y+\underline{x}}(\underline{z})}, \text{ i.e., } \frac{f_Y(\overline{z}-\overline{x})}{f_Y(\overline{z}-\underline{x})} \geq \frac{f_Y(\underline{z}-\overline{x})}{f_Y(\underline{z}-\underline{x})}.$$

And log-concavity of Y is equivalent to $f_Y(\overline{z}-\overline{x})f_Y(\underline{z}-\underline{x}) \geq f(\overline{z}-\underline{x})f(\underline{z}-\overline{x})$ for all $\overline{z} > \underline{z}$, $\overline{x} > \underline{x}$ (Marshall and Olkin, 2007, Proposition 21.B.8), yielding the expression above. □

Proof of Proposition 1. The bulk of the argument is in the text. The only missing step is to prove that alignment up to distinguishability implies monotonicity of χ^{UBAR}.

To see that this is so, take some test result t and some agent utilities $\underline{u}_A < \overline{u}_A$. Alignment implies that $U_P(t, \underline{u}_A) \le U_P(t, \overline{u}_A)$. I seek to confirm that $\chi^{\text{UBAR}}(t, \underline{u}_A) \le \chi^{\text{UBAR}}(t, \overline{u}_A)$. If $U_P(t, \underline{u}_A) < u_P^c$ then the desired inequality holds because $\chi^{\text{UBAR}}(t, \underline{u}_A) = 0$; similarly if $U_P(t, \overline{u}_A) > u_P^c$, the desired inequality holds because $\chi^{\text{UBAR}}(t, \overline{u}_A) = 1$. The remaining possibility is that $U_P(t, \underline{u}_A) = U_P(t, \overline{u}_A) = u_P^c$. In that case, acceptance probabilities are ordered due to the selection of χ^{UBAR} as monotonic over the flexible region. □

Proof of Proposition 2. Take some cutoff function $u_A^c(t)$ and a corresponding score function $C(t) = a_0 + a_1 u_A^c(t)$ for $a_1 < 0$. Rearranging, $u_A^c(t) = \frac{C(t)}{a_1} - \frac{a_0}{a_1}$. It is assumed that the expectation of $C(T)$ exists and is finite, which implies that $C(\cdot)$ and $u_A^c(\cdot)$ are almost everywhere finite-valued.

The agent chooses an acceptance rule χ, a map from test results and agent utilities to acceptance probabilities. Her problem is to choose χ to maximize her objective $\frac{1}{k}\mathbb{E}[\chi(T, U_A) \cdot U_A]$, subject to the two constraints of accepting k applicants and of setting the expectation of $C(T)$ conditional on hiring to κ:

$$\mathbb{E}[\chi(T, U_A)] = k \tag{44}$$

$$\frac{1}{k}\mathbb{E}[\chi(T, U_A) \cdot C(T)] = \kappa. \tag{45}$$

I claim that this problem is solved by $\chi = \chi^{\text{UBAR}}$.

By construction, χ^{UBAR} satisfies the constraints (44) and (45). Suppose for the sake of a contradiction that $\hat{\chi}$ also satisfies the constraints, but yields a strictly higher value of the objective:

$$\begin{aligned} \frac{1}{k}\mathbb{E}[\hat{\chi}(T, U_A)U_A] &> \frac{1}{k}\mathbb{E}[\chi^{\text{UBAR}}(T, U_A)U_A] \\ \Rightarrow \ \frac{1}{k}\mathbb{E}[\hat{\chi}(T, U_A)U_A] - \frac{\kappa}{a_1} + \frac{a_0}{a_1} &> \frac{1}{k}\mathbb{E}[\chi^{\text{UBAR}}(T, U_A)U_A] - \frac{\kappa}{a_1} + \frac{a_0}{a_1}. \end{aligned}$$

Now apply (44) and (45) to both $\hat{\chi}$ and χ^{UBAR} to bring the extra terms inside the expectations:

$$\frac{1}{k}\mathbb{E}\left[\hat{\chi}(T,U_A)\cdot\left(U_A\frac{C(T)}{a_1}+\frac{a_0}{a_1}\right)\right] > \frac{1}{k}\mathbb{E}\left[\chi^{\text{UBAR}}(T,U_A)\cdot\left(U_A-\frac{C(T)}{a_1}+\frac{a_0}{a_1}\right)\right]$$
$$\Rightarrow \frac{1}{k}\mathbb{E}\left[\hat{\chi}(T,U_A)\cdot(U_A-u_A^c(T))\right] > \frac{1}{k}\mathbb{E}\left[\chi^{\text{UBAR}}(T,U_A)\cdot(U_A-u_A^c(T))\right].$$

But this last line yields a contradiction, because χ^{UBAR} maximizes $\chi(T,U_A)(U_A - u_A^c(t))$ pointwise over realizations of (T,U_A), subject to the constraint that χ takes values in $[0,1]$. It sets χ to 1 if $U_A > u_A^c(T)$, and to 0 if $U_A < u_A^c(T)$.[45] □

Before I address the proofs of the formally stated results of Section 3.3, let me work out the derivations of formulas (8) - (11) in the text of that section. These will follow from standard updating rules of normal distributions. First, take a multivariate normal random vector X that can be decomposed as $X = (X_1, X_2)$ with mean (μ_1, μ_2) and covariance matrix $\begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix}$. The conditional distribution of X_1 given $X_2 = x_2$ is given by

$$X_1|X_2 = x_2 \sim \mathcal{N}\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right). \tag{46}$$

Lemma 6. *The variables (Q,T,U_A) are joint normally distributed, with means of 0 and covariance matrix of*

$$\begin{bmatrix} \sigma_Q^2 & \sigma_Q^2 & \frac{\sigma_Q^4(\sigma_T^2+\sigma_S^2)}{\sigma_Q^2\sigma_T^2+\sigma_Q^2\sigma_S^2+\sigma_T^2\sigma_S^2} \\ \sigma_Q^2 & \sigma_Q^2+\sigma_T^2 & \sigma_Q^2 \\ \frac{\sigma_Q^4(\sigma_T^2+\sigma_S^2)}{\sigma_Q^2\sigma_T^2+\sigma_Q^2\sigma_S^2+\sigma_T^2\sigma_S^2} & \sigma_Q^2 & \frac{\sigma_Q^4(\sigma_T^2+\sigma_S^2)}{\sigma_Q^2\sigma_T^2+\sigma_Q^2\sigma_S^2+\sigma_T^2\sigma_S^2} + \sigma_B^2 \end{bmatrix}.$$

Given Lemma 6, we can apply (46) to calculate $U_P(T,U_A)$, defined as the expectation of Q conditional on T and U_A, by taking Q as X_1 and (T,U_A) as X_2. Working out the algebra yields Equations (8) - (11).

Proof of Proposition 3.

[45] While I do not formally work with Lagrangians to avoid technical complications, this proof follows a standard Lagrangian-style argument for showing that a constrained maximization problem can be replaced with an unconstrained one having the same objective minus a multiplier times each constraint. The implied Lagrange multipliers would be $\lambda_0 = -\frac{a_0}{a_1}$ on constraint (44) and $\lambda_1 = \frac{1}{a_1}$ on constraint (45).

1. Utilities are aligned up to distinguishability, and so we can apply Proposition 1 to find one implementation of the optimal contract. To show the desired result, then, it suffices to show that for any fixed principal utility cutoff u_P^c, as we vary t the share of applicants with $U_P(t, U_A) \geq u_P^c$ takes the form $\Phi(\gamma_T^* t - \gamma_0)$ for γ_T^* as in (12) and for some γ_0.

   The first step is to calculate the conditional distribution of U_A given T. Applying Lemma 6 and (46), taking T as X_1 and U_A as X_2, we find that

$$U_A | T \sim \mathcal{N}\left(\mu_{U_A}(T), \sigma_{U_A}^2\right), \text{ for} \tag{13}$$

$$\mu_{U_A}(t) = \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} t \tag{14}$$

$$\sigma_{U_A}^2 = \eta + \sigma_B^2. \tag{15}$$

   (These equations appear in the body of the paper as well, in Section 4.2.)

   Restating (13), for any t,

$$\frac{U_A - \mu_{U_A}(t)}{\sigma_{U_A}} \big| T = t \sim \mathcal{N}(0, 1). \tag{47}$$

   For any t and any u_P^c, we can now calculate the acceptance rate under UBAR. An applicant with $T = t$ is accepted under UBAR if

$$\begin{aligned} & \beta_T t + \beta_{U_A} U_A \geq u_P^c \\ \iff & \frac{U_A - \mu_{U_A}(t)}{\sigma_{U_A}} \geq \frac{\frac{u_P^c - \beta_T t}{\beta_{U_A}} - \mu_{U_A}(t)}{\sigma_{U_A}}. \end{aligned}$$

   Conditional on $T = t$, the LHS of the last line is distributed according to a standard normal. Plugging in $\mu_{U_A}(t)$ from (14) on the RHS and collecting terms, the acceptance condition can be rewritten as

$$\frac{U_A - \mu_{U_A}(t)}{\sigma_{U_A}} \geq \gamma_0 - \gamma_T^* t,$$

for $\gamma_0 = \frac{u_P^c}{\beta_{U_A}\sigma_{U_A}}$ and

$$\gamma_T^* = \frac{\beta_T}{\beta_{U_A}\sigma_{U_A}} + \frac{\sigma_Q^2}{(\sigma_Q^2+\sigma_T^2)\sigma_{U_A}}.$$

So the share of applicants with $U_P(t, U_A) \geq u_P^c$ at test score $T = t$ is $1 - \Phi(\gamma_0 - \gamma_T^* t) = \Phi(\gamma_T^* t - \gamma_0)$.

As stated in the text, I will not explicitly calculate the optimal value of γ_0 as a function of primitives, as u_P^c is itself a function of k. But plugging (9), (10), and (15) into the above expression for γ_T^* and simplifying yields the expression (12) for γ_T^*.

2. From Proposition 2, it suffices to derive the formula for a cutoff indifference curve, $u_A^c(t)$, and set $C(t)$ as any negative affine transformation. Solving for $u_A^c(t)$ as the solution to $U_P(t, u_A^c(t)) = u_P^c$ for a given u_P^c:

$$\begin{aligned} & U_P(t, u_A^c(t)) = u_P^c \\ \overset{\text{Eq (8)}}{\Longrightarrow} \quad & \beta_T t + \beta_{U_A} u_A^c(t) = u_P^c \\ \Longrightarrow \quad & u_A^c(t) = \frac{u_P^c - \beta_T t}{\beta_{U_A}}. \end{aligned}$$

$C(t) = t$ is a negative affine transformation of $u_A^c(t)$, a linear function of t with a negative slope. □

Proof of Proposition 4. Restating (11) and (12),

$$\begin{cases} \gamma_T^* = \frac{\sigma_Q^2\sqrt{\eta+\sigma_B^2}}{\eta(\sigma_Q^2+\sigma_T^2)} \\ \gamma_T^*\sqrt{\sigma_Q^2+\sigma_T^2} = \frac{\sigma_Q^2\sqrt{\eta+\sigma_B^2}}{\eta\sqrt{\sigma_Q^2+\sigma_T^2}} \end{cases},$$

$$\text{for } \eta = \frac{\sigma_Q^4\sigma_T^4}{(\sigma_Q^2+\sigma_T^2)(\sigma_Q^2\sigma_T^2+\sigma_Q^2\sigma_S^2+\sigma_T^2\sigma_S^2)}.$$

1. The parameter k does not appear in the formula for γ_T^*.
2. Taking $\gamma_T^* = \frac{\sigma_Q^2\sqrt{\eta+\sigma_B^2}}{\eta(\sigma_Q^2+\sigma_T^2)}$ as a function of σ_Q^2, σ_B^2, σ_T^2, and η, we can write $\frac{d\gamma_T^*}{d\sigma_T^2}$ as $\frac{\partial\gamma_T^*}{\partial\sigma_T^2} + \frac{\partial\gamma_T^*}{\partial\eta}\frac{d\eta}{d\sigma_T^2}$. It is easy to confirm by routine differentiation that $\frac{\partial\gamma_T^*}{\partial\sigma_T^2} < 0$, $\frac{\partial\gamma_T^*}{\partial\eta} < 0$, and $\frac{d\eta}{d\sigma_T^2} > 0$. Therefore $\frac{d\gamma_T^*}{d\sigma_T^2} < 0$.

Taking $\gamma_T^* \sqrt{\sigma_Q^2 + \sigma_T^2} = \frac{\sigma_Q^2 \sqrt{\eta + \sigma_B^2}}{\eta \sqrt{\sigma_Q^2 + \sigma_T^2}}$ as a function of σ_Q^2, σ_B^2, σ_T^2, and η, we can write $\frac{d\left(\gamma_T^* \sqrt{\sigma_Q^2+\sigma_T^2}\right)}{d\sigma_T^2}$ as $\frac{\partial\left(\gamma_T^* \sqrt{\sigma_Q^2+\sigma_T^2}\right)}{\partial\sigma_T^2} + \frac{\partial\left(\gamma_T^* \sqrt{\sigma_Q^2+\sigma_T^2}\right)}{\partial\eta} \frac{d\eta}{d\sigma_T^2}$. Once again, $\frac{\partial\left(\gamma_T^* \sqrt{\sigma_Q^2+\sigma_T^2}\right)}{\partial\sigma_T^2} < 0$, $\frac{\partial\left(\gamma_T^* \sqrt{\sigma_Q^2+\sigma_T^2}\right)}{\partial\eta} < 0$, and $\frac{\partial\eta}{\partial\sigma_T^2} > 0$. Therefore $\frac{d\left(\gamma_T^* \sqrt{\sigma_Q^2+\sigma_T^2}\right)}{d\sigma_T^2} < 0$.

Taking limits,

$$\lim_{\sigma_T^2 \to 0} \gamma_T^* = \lim_{\sigma_T^2 \to 0} \gamma_T^* \sqrt{\sigma_Q^2 + \sigma_T^2} = \infty, \text{ because } \lim_{\sigma_T^2 \to 0} \eta = 0$$

$$\lim_{\sigma_T^2 \to \infty} \gamma_T^* = \lim_{\sigma_T^2 \to \infty} \gamma_T^* \sqrt{\sigma_Q^2 + \sigma_T^2} = 0, \text{ because } \lim_{\sigma_T^2 \to \infty} \eta = \frac{\sigma_Q^4}{\sigma_Q^2 + \sigma_S^2}.$$

3. The parameter σ_S^2 appears in γ_T^* only through η. Routine differentiation shows that $\frac{\partial\gamma_T^*}{\partial\eta} < 0$ and $\frac{d\eta}{d\sigma_S^2} < 0$, and so by the chain rule $\frac{d\gamma_T^*}{d\sigma_S^2} > 0$.

   Taking limits,

$$\lim_{\sigma_S^2 \to 0} \gamma_T^* = \frac{1}{\sigma_T^2} \sqrt{\frac{\sigma_T^2 \sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} + \sigma_B^2}, \text{ because } \lim_{\sigma_S^2 \to 0} \eta = \frac{\sigma_T^2 \sigma_Q^2}{\sigma_Q^2 + \sigma_T^2}$$

$$\lim_{\sigma_S^2 \to \infty} \gamma_T^* = \infty, \text{ because } \lim_{\sigma_S^2 \to \infty} \eta = 0.$$

4. The value η remains constant as we vary σ_B^2. Taking the derivative of γ_T^* with respect to σ_B^2 gives

$$\frac{\sigma_Q^2}{2\eta(\sigma_Q^2 + \sigma_T^2)\sqrt{\eta + \sigma_B^2}} > 0.$$

   Taking limits,

$$\lim_{\sigma_B^2 \to 0} \gamma_T^* = \frac{\sigma_Q^2}{(\sigma_Q^2 + \sigma_T^2)\sqrt{\eta}}$$

$$\lim_{\sigma_B^2 \to \infty} \gamma_T^* = \infty.$$

5. Numerical examples (not shown) verify that, depending on parameters, both γ_T^* and $\gamma_T^* \sqrt{\sigma_Q^2 + \sigma_T^2}$ can either be locally increasing or decreasing in σ_Q^2.

   It is easy to verify that $\lim_{\sigma_Q^2 \to 0} \frac{\sigma_T^2 \sigma_Q^2}{\sigma_B} \gamma_T^* \to 1$. Therefore $\lim_{\sigma_Q^2 \to 0} \gamma_T^* = \lim_{\sigma_Q^2 \to 0} \gamma_T^* \sqrt{\sigma_Q^2 + \sigma_T^2} =$

∞. Taking $\sigma_Q^2 \to \infty$, we get $\lim_{\sigma_Q^2 \to \infty} \eta = \frac{\sigma_T^4}{\sigma_T^2 + \sigma_S^2}$ and so

$$\lim_{\sigma_Q^2 \to \infty} \gamma_T^* = \frac{(\sigma_T^2 + \sigma_S^2)\sqrt{\frac{\sigma_T^4}{\sigma_T^2+\sigma_S^2} + \sigma_B^2}}{\sigma_T^4}$$

$$\lim_{\sigma_Q^2 \to \infty} \gamma_T^* \sqrt{\sigma_Q^2 + \sigma_T^2} = \infty. \qquad \square$$

G.2 Proofs for Section 4

Proof of Lemma 2. Rewriting (15), (18), and (11),

$$\sigma_{U_A}(\theta) = \sqrt{\eta + \sigma_B^2}$$
$$\sigma_{U_P}(\theta) = \frac{\eta}{\sqrt{\eta + \sigma_B^2}}$$
$$\text{for } \eta = \frac{\sigma_Q^4 \sigma_T^4}{(\sigma_Q^2 + \sigma_T^2)(\sigma_T^2\sigma_Q^2 + \sigma_S^2\sigma_T^2 + \sigma_S^2\sigma_Q^2)}.$$

1. Observe that η decreases in σ_S^2. Fixing σ_B^2, both σ_{U_A} and σ_{U_P} increase in η.
2. The term η is constant in σ_B^2. Fixing σ_S^2, σ_{U_A} increases in σ_B^2 while σ_{U_P} decreases in σ_B^2.
3. Fixing $\sigma_{U_A}(\theta) = \tilde{\sigma}_{U_A} > 0$, the range of possible σ_{U_P} is an open interval in $\mathbb{R}_+$. One can achieve the minimum of this interval by taking $\sigma_B \to \tilde{\sigma}_{U_A}$ and $\sigma_S \to \infty$, implying $\sigma_{U_P} \to 0$. It remains to show that the supremum of $\sigma_{U_P}(\theta)$ given $\sigma_{U_A}(\theta) = \tilde{\sigma}_{U_A}$ is $\min\left\{\tilde{\sigma}_{U_A}, \frac{\sigma_Q^2\sigma_T^2}{\sigma_Q^2+\sigma_T^2}\frac{1}{\tilde{\sigma}_{U_A}}\right\}$, or equivalently that this supremum is $\tilde{\sigma}_{U_A}$ for $\tilde{\sigma}_{U_A} \le \frac{\sigma_Q\sigma_T}{\sqrt{\sigma_Q^2+\sigma_T^2}}$ and is $\frac{\sigma_Q^2\sigma_T^2}{\sigma_Q^2+\sigma_T^2}\frac{1}{\tilde{\sigma}_{U_A}}$ for $\tilde{\sigma}_{U_A} > \frac{\sigma_Q\sigma_T}{\sqrt{\sigma_Q^2+\sigma_T^2}}$.

   The term η is independent of σ_B^2, and decreases from $\frac{\sigma_Q^2\sigma_T^2}{\sigma_Q^2+\sigma_T^2}$ to zero as σ_S^2 increases from zero to infinity. From Parts 1 and 2, we maximize $\sigma_{U_P}(\theta)$ given $\sigma_{U_A}(\theta) = \tilde{\sigma}_{U_A}$ over $\theta = (\sigma_S^2, \sigma_B^2)$ by finding the mixture of the lowest σ_S^2 (most information) and the lowest σ_B^2 (least bias) consistent with $\sigma_{U_A}(\theta) = \tilde{\sigma}_{U_A}$.

   For $\tilde{\sigma}_{U_A} \le \frac{\sigma_Q\sigma_T}{\sqrt{\sigma_Q^2+\sigma_T^2}}$, we achieve a supremum for σ_{U_P} of $\tilde{\sigma}_{U_A}$ by taking $\sigma_B^2 \to 0$ and setting σ_S^2 so that $\eta = \tilde{\sigma}_{U_A}^2$, implying $\sigma_{U_A}(\theta) = \tilde{\sigma}_{U_A}$.

   For $\tilde{\sigma}_{U_A} > \frac{\sigma_Q\sigma_T}{\sqrt{\sigma_Q^2+\sigma_T^2}}$, take $\sigma_S^2 \to 0$, which implies $\eta \to \frac{\sigma_Q^2\sigma_T^2}{\sigma_Q^2+\sigma_T^2}$; and set $\sigma_B^2 = \tilde{\sigma}_{U_A}^2 - \eta$ to get $\sigma_{U_A}(\theta) = \tilde{\sigma}_{U_A}$. Plugging $\eta \to \frac{\sigma_Q^2\sigma_T^2}{\sigma_Q^2+\sigma_T^2}$ into the identity $\sigma_{U_P}(\theta) = \eta/\sigma_{U_A}(\theta) =$

$\eta/\tilde{\sigma}_{U_A}$ then gives $\sigma_{U_P} \to \frac{\sigma_Q^2 \sigma_T^2}{\sigma_Q^2+\sigma_T^2} \frac{1}{\tilde{\sigma}_{U_A}}$. $\square$

Proof of Lemma 3.

Step 1. Fixing k, let us start by deriving the "upper-right frontier" of (τ, ζ), the pairs that maximize $p\tau + (1-p)\zeta$ for some $p \in [0,1]$. One maximizes $p\tau + (1-p)\zeta$ by selecting the k applicants with the highest values of $pT + (1-p)Z$. (For $p \in (0,1)$, these are exactly the applicants above a downward sloping line in (T, U_A)-space – accepting such applicants induces a normal CDF acceptance rate.) In the population, T and Z are independently normally distributed with means of 0, and respective variances of $\sigma_Q^2 + \sigma_T^2$ and 1. Therefore $pT + (1-p)Z$ has mean 0 and variance σ_{comb}^2, for $\sigma_{\text{comb}} \equiv \sqrt{p^2(\sigma_Q^2 + \sigma_T^2) + (1-p)^2}$. The applicants with the k highest values of $pT + (1-p)Z$ are those with $\frac{pT+(1-p)Z}{\sigma_{\text{comb}}} \geq r^*$, for r^* satisfying $\Phi(r^*) = 1 - k$. I seek to calculate the expected value of T and of Z conditional on $\frac{pT+(1-p)Z}{\sigma_{\text{comb}}} \geq r^*$.

We have the following joint normal distribution among the three random variables T, Z, and $\frac{pT+(1-p)Z}{\sigma_{\text{comb}}}$:

$$\begin{bmatrix} T \\ Z \\ \frac{pT+(1-p)Z}{\sigma_{\text{comb}}} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_Q^2 + \sigma_T^2 & 0 & \frac{p}{\sigma_{\text{comb}}}(\sigma_Q^2 + \sigma_T^2) \\ 0 & 1 & \frac{1-p}{\sigma_{\text{comb}}} \\ \frac{p}{\sigma_{\text{comb}}}(\sigma_Q^2 + \sigma_T^2) & \frac{1-p}{\sigma_{\text{comb}}} & 1 \end{bmatrix} \right).$$

As in expression (46) of Appendix G.1, we can calculate conditional means of T and Z conditional on any realization $\frac{pT+(1-p)Z}{\sigma_{\text{comb}}} = r$:

$$\mathbb{E}[T | \frac{pT + (1-p)Z}{\sigma_{\text{comb}}} = r] = \frac{p(\sigma_Q^2 + \sigma_T^2)}{\sigma_{\text{comb}}} r$$
$$\mathbb{E}[Z | \frac{pT + (1-p)Z}{\sigma_{\text{comb}}} = r] = \frac{1-p}{\sigma_{\text{comb}}} r.$$

This holds for every realization r. Therefore, for every r,

$$\mathbb{E}[T \mid \frac{pT + (1-p)Z}{\sigma_{\text{comb}}} \geq r] = \frac{p(\sigma_Q^2 + \sigma_T^2)}{\sigma_{\text{comb}}} \mathbb{E}[\frac{pT + (1-p)Z}{\sigma_{\text{comb}}} \mid \frac{pT + (1-p)Z}{\sigma_{\text{comb}}} \geq r]$$
$$\mathbb{E}[Z \mid \frac{pT + (1-p)Z}{\sigma_{\text{comb}}} \geq r] = \frac{1-p}{\sigma_{\text{comb}}} \mathbb{E}[\frac{pT + (1-p)Z}{\sigma_{\text{comb}}} \mid \frac{pT + (1-p)Z}{\sigma_{\text{comb}}} \geq r].$$

And given that $\frac{pT+(1-p)Z}{\sigma_{\text{comb}}}$ follows a standard normal, the truncated mean $\mathbb{E}[\frac{pT+(1-p)Z}{\sigma_{\text{comb}}} \mid \frac{pT+(1-p)Z}{\sigma_{\text{comb}}} \geq r]$ is equal to $\frac{\phi(r)}{1-\Phi(r)}$. Evaluating the above expressions at $r = r^*$:

$$\tau = \mathbb{E}[T \mid \frac{pT + (1-p)Z}{\sigma_{\text{comb}}} \geq r^*] = \frac{p(\sigma_Q^2 + \sigma_T^2)}{\sigma_{\text{comb}}} \frac{\phi(r^*)}{1 - \Phi(r^*)} = \frac{p(\sigma_Q^2 + \sigma_T^2)}{\sqrt{p^2(\sigma_Q^2 + \sigma_T^2) + (1-p)^2}} R(k)$$

$$\zeta = \mathbb{E}[Z \mid \frac{pT + (1-p)Z}{\sigma_{\text{comb}}} \geq r^*] = \frac{1-p}{\sigma_{\text{comb}}} \frac{\phi(r^*)}{1 - \Phi(r^*)} = \frac{1-p}{\sqrt{p^2(\sigma_Q^2 + \sigma_T^2) + (1-p)^2}} R(k).$$

As p goes from 0 to 1, τ goes from 0 to $R_T = \sqrt{\sigma_Q^2 + \sigma_T^2} R(k)$ and ζ goes from $R_Z = R(k)$ to 0. For any $p \in [0, 1]$,

$$\begin{aligned} \frac{\tau^2}{R_T^2} + \frac{\zeta^2}{R_Z^2} &= \frac{1}{\sigma_Q^2 + \sigma_T^2} \frac{p^2(\sigma_Q^2 + \sigma_T^2)^2}{p^2(\sigma_Q^2 + \sigma_T^2) + (1-p)^2} + \frac{(1-p)^2}{p^2(\sigma_Q^2 + \sigma_T^2) + (1-p)^2} \\ &= \frac{p^2(\sigma_Q^2 + \sigma_T^2) + (1-p)^2}{p^2(\sigma_Q^2 + \sigma_T^2) + (1-p)^2} = 1. \end{aligned}$$

So we see that varying $p \in [0, 1]$ traces out the upper-right boundary of the ellipse $\overline{W}$.

Step 2. We can proceed similarly to show that we trace out the entire boundary of the ellipse as we maximize the four combinations of $\pm p\tau \pm (1-p)\zeta$ for $p \in [0, 1]$. In other words, every (τ, ζ) that is a boundary point of $\overline{W}$ is achieved by some set of k applicants. Moreover, no set of k applicants achieves a pair (τ, ζ) that is outside the boundaries of this ellipse, the interior of which is convex – otherwise this point would yield a higher value of an appropriately signed $\pm p\tau \pm (1-p)\zeta$ than any value on the boundary.

Step 3. Finally, the set of achievable (τ, ζ) across applicant pools is convex: choosing a convex combination of applicants from the two pools yields the same convex combination of the average test score and average z-score (τ, ζ). So all points in the interior of $\overline{W}$ are achievable as well. □

Proof of Proposition 5. I will show this result as an application of the one-dimensional delegation results in Amador et al. (2018), an extension of Amador and Bagwell (2013). Specifically, Lemma 2 of Amador et al. (2018) provides sufficiency conditions for the optimality of an action ceiling. As a delegation problem in the framework of those two papers, let the "state" $\sigma_{U_A}(\theta)$ be distributed according to H, let $\zeta \in [0, R_Z]$

be the contractible "action," and let the level of joint "money burning" be $\delta \in \mathbb{R}_+$.[46] Those papers take a contract to be an arbitrary set of allowed actions ζ and an arbitrary function from allowed actions to nonnegative money burning; in my problem money burning is restricted, bounded at $\delta \leq 2\bar{\tau}(\zeta)$ under action ζ. However, under the conditions of Lemma 2 in Amador et al. (2018), money burning will be identically zero in the optimal delegation contract; any upper bound on money burning will therefore not be binding.[47]

Following the notation of those other papers, the agent's payoff over the state and action, prior to money burning, can be written as

$$\sigma_{U_A}(\theta) \cdot \zeta + b(\zeta), \text{ for } b(\zeta) \equiv \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} R_T \cdot \sqrt{1 - \frac{\zeta^2}{R_Z^2}}. \tag{48}$$

The principal's payoff prior to money burning can be written as $w(\sigma_{U_A}(\theta), \zeta)$, with

$$w(\tilde{\sigma}_{U_A}, \zeta) \equiv \hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A}) \cdot \zeta + b(\zeta). \tag{49}$$

Money burning of δ reduces both payoffs by that same amount. These payoffs are not just of the general form considered in Amador and Bagwell (2013), but of the form in Equation (6) of Amador et al. (2018): $w(\sigma_{U_A}(\theta), \zeta) = A[b(\zeta) + B(\sigma_{U_A}(\theta)) + C(\sigma_{U_A}(\theta))\zeta]$ for $A = 1$, $B(\sigma_{U_A}(\theta)) = 0$, and $C(\sigma_{U_A}(\theta)) = \hat{\sigma}_{U_P}(\sigma_{U_A}(\theta))$.

Denote the agent's interim optimal action at state $\sigma_{U_A}(\theta)$ – her "flexible" action – as $\zeta_f(\sigma_{U_A}(\theta))$. Taking the first order condition of (48),

$$\zeta_f(\sigma_{U_A}(\theta)) = \frac{\sigma_{U_A}(\theta) R_Z}{\sqrt{\sigma_{U_A}(\theta)^2 + \frac{\sigma_Q^4}{(\sigma_Q^2 + \sigma_T^2)^2} \frac{R_T^2}{R_Z^2}}}. \tag{50}$$

[46] The notation of Amador and Bagwell (2013) and **?** has state γ distributed according to cdf F, with pdf f; action π; and money burning t. They use $\pi_f(\gamma)$ for the function describing the agent's ideal action, what I will call $\zeta_f(\sigma_{U_A}(\theta))$; and the latter paper uses π^* for the principal's ex ante optimal action, what I will call ζ^*. I follow these papers in using $b(\cdot)$ as the component of the agent's payoff function that depends on the action, despite my previous use of b as the bias realization for a given applicant; there should be no confusion between the two distinct terms.

[47] A delegation contract takes a delegation set of allowed actions and a money burning function as direct objects of choice, whereas these are induced objects – expectations over a selected applicant pool – in the problem of the current paper. Hence, even if a delegation contract satisfies the necessary condition of $\delta \leq 2\bar{\tau}(\zeta)$ at each ζ, I still need to show how to find a contract in my setting that implements this delegation outcome.

Denote the principal's ex ante optimal action by ζ^*, the arg max over ζ of the expectation of (49):

$$\zeta^* = \arg\max_{\zeta \in [0, R_Z]} \mathbb{E}_{\sigma_{U_A}(\theta) \sim H}[\hat{\sigma}_{U_P}(\sigma_{U_A}(\theta))]\zeta + b(\zeta). \tag{51}$$

Because $\hat{\sigma}_{U_P}(\sigma_{U_A}(\theta)) \in (0, \sigma_{U_A}(\theta))$, and because the Proposition assumes that $\sigma_{U_A}(\theta)$ has bounded support, it holds that $\mathbb{E}_{\sigma_{U_A}(\theta) \sim H}[\hat{\sigma}_{U_P}(\sigma_{U_A}(\theta))]$ is finite and strictly positive. Moreover, the derivative of $b(\zeta)$ is 0 as $\zeta \to 0$ and minus infinity as $\zeta \to R_Z$. Therefore ζ^* is interior, contained in $(0, R_Z)$; we have now verified Assumption 2 of Amador et al. (2018).

We can now verify the regularity conditions of Assumption 1 of Amador et al. (2018). Going through the list, (i) w is continuous; (ii) $w(\tilde{\sigma}_{U_A}, \cdot)$ is concave and twice differentiable for every $\tilde{\sigma}_{U_A}$; (iii) $b(\cdot)$ is strictly concave and twice differentiable; (iv) $\zeta_f(\cdot)$ is twice-differentiable and strictly increasing; and (v) the function w_ζ, the derivative of w with respect to ζ, is continuous.

Next, let us evaluate w_ζ at the agent's ideal point from (50). Putting together $w_\zeta(\tilde{\sigma}_{U_A}, \zeta) = \hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A}) + b'(\zeta)$ with the fact that the agent's ideal point $\zeta_f(\sigma_{U_A}(\theta))$ is derived from the first order condition $b'(\zeta_f(\tilde{\sigma}_{U_A})) = -\tilde{\sigma}_{U_A}$, it holds that

$$w_\zeta(\tilde{\sigma}_{U_A}, \zeta_f(\tilde{\sigma}_{U_A})) = \hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A}) - \tilde{\sigma}_{U_A}. \tag{52}$$

From (52) combined with $\hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A}) \in (0, \tilde{\sigma}_{U_A})$, we see that $w_\zeta(\tilde{\sigma}_{U_A}, \zeta_f(\tilde{\sigma}_{U_A}))$ is strictly negative at each $\tilde{\sigma}_{U_A} > 0$, and is equal to zero in the limit as $\tilde{\sigma}_{U_A} \to 0$ (if this limit is in the support of H); in economic terms, the agent's bias is always towards higher ζ. Therefore $w_\zeta(\tilde{\sigma}_{U_A}, \zeta_f(\tilde{\sigma}_{U_A}))$ satisfies the sign restrictions of Lemma 2 of Amador et al. (2018) for $\tilde{\sigma}_{U_A}$ at the lower and upper bounds of the support.

To apply that Lemma 2, it remains only to check condition (Gc1) of Amador et al. (2018). The parameter κ appearing in (Gc1) is equal to 1 because – as mentioned above – the payoffs are of the form in Equation (6) of that paper, with $A = 1$. Plugging in $\kappa = 1$ and $w_\zeta(\tilde{\sigma}_{U_A}, \zeta_f(\tilde{\sigma}_{U_A})) = \hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A}) - \tilde{\sigma}_{U_A}$, condition (Gc1) states that $H(\tilde{\sigma}_{U_A}) + (\tilde{\sigma}_{U_A} - \hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A}))h(\tilde{\sigma}_{U_A})$ is nondecreasing over $\tilde{\sigma}_{U_A}$ in the support of H. This condition is exactly what is assumed in the statement of the proposition.

We have now confirmed all of the hypotheses of Lemma 2 of that paper. We can therefore conclude that in the delegation problem with money burning allowed, the optimal delegation set is of the form of a ceiling on ζ – possibly everywhere binding, implying a one-point delegation set of $\zeta = \zeta^*$ – with money burning δ identically equal to 0.

Finally, as discussed in the body of the paper and illustrated in Figure 6, a ceiling on ζ and no money burning corresponds to an applicant selection contract that takes the form of a floor on the average test score τ. Furthermore, as in Section 3.3, each average test score that the agent may choose is equivalent to a normal CDF acceptance rate function, with higher test scores mapping to steeper normal CDFs. □

Proof of Proposition 6. Follows immediately from Proposition 10, below. □

Proposition 10. *Suppose that* $\sup_{\tilde{\sigma}_{U_A} \in \text{Supp}H} \hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A}) \leq \inf_{\tilde{\sigma}_{U_A} \in \text{Supp}H} \tilde{\sigma}_{U_A}$. *Then the optimal contract can be characterized in either of the following two ways:*

1. *The agent is allowed to choose any k applicants as long as the induced acceptance rate $\alpha(T)$ is $\Phi(\gamma_T T - \gamma_0)$, with γ_T equal to some specified level $\Gamma > 0$.*
2. *The agent may select any k applicants as long as the average test score of hired applicants τ is equal to some specified level $\kappa > 0$.*

Proof of Proposition 10. **Step 1.** Take $\tilde{\sigma}^l_{U_A} < \tilde{\sigma}^h_{U_A}$, and take points (τ^l, ζ^l) and (τ^h, ζ^h) in $\overline{W}$. Suppose an agent with $\sigma_{U_A}(\theta) = \tilde{\sigma}^l_{U_A}$ weakly prefers (τ^l, ζ^l) to (τ^h, ζ^h), and an agent with $\sigma_{U_A}(\theta) = \tilde{\sigma}^h_{U_A}$ weakly prefers (τ^h, ζ^h) to (τ^l, ζ^l). I claim that if $\hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A}) \leq \tilde{\sigma}^l_{U_A}$, then conditional on $\sigma_{U_A}(\theta) = \tilde{\sigma}_{U_A}$ the principal weakly prefers (τ^l, ζ^l) to (τ^h, ζ^h).

This claim follows as a straightforward single-crossing argument from (26) and (27). From the two preference orderings, it must be that $\tau^l \geq \tau^h$ and $\zeta^l \leq \zeta^h$. Now, writing out the agent's choice given $\sigma_{U_A}(\theta) = \tilde{\sigma}^l_{U_A}$, it holds that

$$\frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} \cdot \tau^l + \tilde{\sigma}^l_{U_A} \cdot \zeta^l \geq \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} \cdot \tau^h + \tilde{\sigma}^l_{U_A} \cdot \zeta^h.$$

Because $\zeta^l \leq \zeta^h$, the same inequality holds when $\tilde{\sigma}^l_{U_A}$ is replaced by $\hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A}) \leq \tilde{\sigma}^l_{U_A}$.

Step 2. By the claim in Step 1, under any contract, the principal prefers the (τ, ζ) pair chosen by the agent with $\sigma_{U_A}(\theta)$ equal to the minimum of the support of H to that chosen by any other agent type. So the contract is weakly improved by one which requires the agent to always choose that value (τ, ζ). This new contract, in turn, can be improved by one that specifies that the agent always chooses the principal's ex ante preferred (τ, ζ): the value on the payoff frontier which maximizes

$$\frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} \cdot \tau + \mathbb{E}_{\tilde{\sigma}_{U_A} \sim H}[\hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A})] \cdot \zeta.$$

This can be implemented by fixing the appropriate average test score, or by setting the appropriate normal CDF acceptance rate function. □

Proof of Lemma 4. It is sufficient to confirm that $\hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A})$ is differentiable over the support when the bias is commonly known (and is therefore continuous), and that $\hat{\sigma}'_{U_P}(\tilde{\sigma}_{U_A}) \leq 2$. In that case Lemma 7 below implies the result: for $\hat{\sigma}_{U_P}$ differentiable, condition (iii) of Lemma 7 amounts to $\hat{\sigma}'_{U_P}(\tilde{\sigma}_{U_A}) \leq 2$.

From (15) and (18),

$$\hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A}) = \frac{\tilde{\sigma}_{U_A}^2 - \sigma_B^2}{\tilde{\sigma}_{U_A}} = \tilde{\sigma}_{U_A} - \frac{\sigma_B^2}{\tilde{\sigma}_{U_A}}$$

Taking the derivative, $\hat{\sigma}'_{U_P}(\tilde{\sigma}_{U_A}) = 1 + \frac{\sigma_B^2}{\tilde{\sigma}_{U_A}^2}$. To show that $\hat{\sigma}'_{U_P}(\tilde{\sigma}_{U_A}) \leq 2$, it suffices to show that $\tilde{\sigma}_{U_A}^2 > \sigma_B^2$; and this follows directly from (15), which states that for any agent type $\theta = (\sigma_S^2, \sigma_B^2)$ it holds that $\sigma_{U_A}^2(\theta) = \eta + \sigma_B^2$, with $\eta > 0$. □

Lemma 7. *Suppose that (i) the distribution H has pdf h, (ii) $h(\tilde{\sigma}_{U_A})$ is nondecreasing in $\tilde{\sigma}_{U_A}$ over the support of the distribution, and (iii) $(2\tilde{\sigma}_{U_A} - \hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A}))$ is nondecreasing in $\tilde{\sigma}_{U_A}$. Then $H(\tilde{\sigma}_{U_A}) + (\tilde{\sigma}_{U_A} - \hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A}))h(\tilde{\sigma}_{U_A})$ is nondecreasing in $\tilde{\sigma}_{U_A}$.*

Proof of Lemma 7. Let $\Delta(\tilde{\sigma}_{U_A}) \equiv \tilde{\sigma}_{U_A} - \hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A})$. It holds that $\Delta(\tilde{\sigma}_{U_A}) > 0$. Assumption (iii), that $2\tilde{\sigma}_{U_A} - \hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A})$ is nondecreasing, can be equivalently stated as $\Delta(\bar{\sigma}) + \bar{\sigma} \geq \Delta(\underline{\sigma}) + \underline{\sigma}$ for any $\bar{\sigma} > \underline{\sigma}$ in the support of $\tilde{\sigma}_{U_A}$. In other words, (iii) is equivalent to (iii′):

$$\Delta(\bar{\sigma}) - \Delta(\underline{\sigma}) \geq \underline{\sigma} - \bar{\sigma} \text{ for any } \bar{\sigma} > \underline{\sigma}. \qquad \text{(iii}'\text{)}$$

I seek to prove that for any $\bar{\sigma} > \underline{\sigma}$,

$$(H(\bar{\sigma}) + \Delta(\bar{\sigma})h(\bar{\sigma})) - (H(\underline{\sigma}) + \Delta(\underline{\sigma})h(\underline{\sigma})) \geq 0.$$

Rewriting the LHS,

$$
\begin{aligned}
(H(\bar{\sigma})+&\Delta(\bar{\sigma})h(\bar{\sigma})) - (H(\underline{\sigma}) + \Delta(\underline{\sigma})h(\underline{\sigma})) \\
&= H(\bar{\sigma}) - H(\underline{\sigma}) + \Delta(\bar{\sigma})h(\bar{\sigma}) - \Delta(\underline{\sigma})h(\underline{\sigma}) + [\Delta(\bar{\sigma})h(\underline{\sigma}) - \Delta(\bar{\sigma})h(\underline{\sigma})] \\
&= [\Delta(\bar{\sigma})(h(\bar{\sigma}) - h(\underline{\sigma}))] + [H(\bar{\sigma}) - H(\underline{\sigma})] + [h(\underline{\sigma})(\Delta(\bar{\sigma}) - \Delta(\underline{\sigma})] \\
&\geq [\Delta(\bar{\sigma})(h(\bar{\sigma}) - h(\underline{\sigma}))] + [h(\underline{\sigma})(\bar{\sigma} - \underline{\sigma})] + [h(\underline{\sigma})(\Delta(\bar{\sigma}) - \Delta(\underline{\sigma}))] \quad \text{by (ii)} \\
&\geq [\Delta(\bar{\sigma})(h(\bar{\sigma}) - h(\underline{\sigma}))] + [h(\underline{\sigma})(\bar{\sigma} - \underline{\sigma})] + [h(\underline{\sigma})(\underline{\sigma} - \bar{\sigma})] \quad \text{by (iii}'\text{)} \\
&= \Delta(\bar{\sigma})(h(\bar{\sigma}) - h(\underline{\sigma})) \geq 0 \text{ by (ii).}
\end{aligned}
$$

□

G.3 Additional Appendix proofs

Proof of Proposition 7. Restating (11) and (31),

$$
\begin{aligned}
\gamma_T^{\text{FD}} &= \frac{\sigma_Q^2}{(\sigma_Q^2 + \sigma_T^2)\sqrt{\sigma_B^2 + \eta}} \text{ and} \\
\gamma_T^{\text{FD}}\sqrt{\sigma_Q^2 + \sigma_T^2} &= \frac{\sigma_Q^2}{\sqrt{\sigma_Q^2 + \sigma_T^2}\sqrt{\sigma_B^2 + \eta}}, \\
\text{for } \eta &= \frac{\sigma_Q^4 \sigma_T^4}{(\sigma_Q^2 + \sigma_T^2)(\sigma_Q^2\sigma_T^2 + \sigma_Q^2\sigma_S^2 + \sigma_T^2\sigma_S^2)}.
\end{aligned}
$$

1. The parameter k does not appear in the formula for γ_T^{FD}.
2. Taking $\gamma_T^{\text{FD}} = \frac{\sigma_Q^2}{(\sigma_Q^2+\sigma_T^2)\sqrt{\sigma_B^2+\eta}}$ as a function of σ_Q^2, σ_B^2, σ_T^2, and η, we can write $\frac{d\gamma_T^{\text{FD}}}{d\sigma_T^2}$ as $\frac{\partial\gamma_T^{\text{FD}}}{\partial\sigma_T^2} + \frac{\partial\gamma_T^{\text{FD}}}{\partial\eta}\frac{d\eta}{d\sigma_T^2}$. It is easy to confirm that $\frac{\partial\gamma_T^{\text{FD}}}{\partial\sigma_T^2} < 0$, $\frac{\partial\gamma_T^{\text{FD}}}{\partial\eta} < 0$, and $\frac{d\eta}{d\sigma_T^2} > 0$. Therefore $\frac{d\gamma_T^{\text{FD}}}{d\sigma_T^2} < 0$.
   Taking $\gamma_T^{\text{FD}}\sqrt{\sigma_Q^2 + \sigma_T^2} = \frac{\sigma_Q^2}{\sqrt{\sigma_Q^2+\sigma_T^2}\sqrt{\sigma_B^2+\eta}}$ as a function of σ_Q^2, σ_B^2, σ_T^2, and η, we can write $\frac{d\left(\gamma_T^{\text{FD}}\sqrt{\sigma_Q^2+\sigma_T^2}\right)}{d\sigma_T^2}$ as $\frac{\partial\left(\gamma_T^{\text{FD}}\sqrt{\sigma_Q^2+\sigma_T^2}\right)}{\partial\sigma_T^2} + \frac{\partial\left(\gamma_T^{\text{FD}}\sqrt{\sigma_Q^2+\sigma_T^2}\right)}{\partial\eta}\frac{d\eta}{d\sigma_T^2}$. Once again, $\frac{\partial\left(\gamma_T^{\text{FD}}\sqrt{\sigma_Q^2+\sigma_T^2}\right)}{\partial\sigma_T^2} < 0$, $\frac{\partial\left(\gamma_T^{\text{FD}}\sqrt{\sigma_Q^2+\sigma_T^2}\right)}{\partial\eta} < 0$, and $\frac{\partial\eta}{\partial\sigma_T^2} > 0$. Therefore $\frac{d\left(\gamma_T^{\text{FD}}\sqrt{\sigma_Q^2+\sigma_T^2}\right)}{d\sigma_T^2} < 0$.

Taking limits,

$$\lim_{\sigma_T^2 \to 0} \gamma_T^{\mathrm{FD}} = \frac{1}{\sigma_B}, \text{ because } \lim_{\sigma_T^2 \to 0} \eta = 0$$

$$\lim_{\sigma_T^2 \to \infty} \gamma_T^{\mathrm{FD}} = 0, \text{ because } \lim_{\sigma_T^2 \to \infty} \eta = \frac{\sigma_Q^4}{\sigma_Q^2 + \sigma_S^2}$$

$$\lim_{\sigma_T^2 \to 0} \gamma_T^{\mathrm{FD}} \sqrt{\sigma_Q^2 + \sigma_T^2} = \frac{\sigma_Q}{\sigma_B}, \text{ because } \lim_{\sigma_T^2 \to 0} \eta = 0$$

$$\lim_{\sigma_T^2 \to \infty} \gamma_T^{\mathrm{FD}} \sqrt{\sigma_Q^2 + \sigma_T^2} = 0, \text{ because } \lim_{\sigma_T^2 \to \infty} \eta = \frac{\sigma_Q^4}{\sigma_Q^2 + \sigma_S^2}.$$

3. The parameter σ_S^2 appears only in γ_T^{FD} through η. Routine differentiation shows that $\frac{d\gamma_T^{\mathrm{FD}}}{d\eta} < 0$ and $\frac{d\eta}{d\sigma_S^2} < 0$, and so by the chain rule $\frac{d\gamma_T^{\mathrm{FD}}}{d\sigma_S^2} > 0$.
Taking limits,

$$\lim_{\sigma_S^2 \to 0} \gamma_T^{\mathrm{FD}} = \frac{\sigma_Q^2}{(\sigma_Q^2 + \sigma_T^2)\sqrt{\frac{\sigma_Q^2 \sigma_T^2}{\sigma_Q^2 + \sigma_T^2} + \sigma_B^2}}, \text{ because } \lim_{\sigma_S^2 \to 0} \eta = \frac{\sigma_Q^2 \sigma_T^2}{\sigma_Q^2 + \sigma_T^2}$$

$$\lim_{\sigma_S^2 \to \infty} \gamma_T^{\mathrm{FD}} = \frac{\sigma_Q^2}{\sigma_B(\sigma_Q^2 + \sigma_T^2)}, \text{ because } \lim_{\sigma_S^2 \to \infty} \eta = 0.$$

From the proof of Proposition 4, $\lim_{\sigma_S^2 \to 0} \gamma_T^* = \frac{1}{\sigma_T^2}\sqrt{\frac{\sigma_Q^2 \sigma_T^2}{\sigma_Q^2 + \sigma_T^2} + \sigma_B^2}$. On the other hand, $\lim_{\sigma_S^2 \to 0} \gamma_T^{\mathrm{FD}}$ can be written as

$$\lim_{\sigma_S^2 \to 0} \gamma_T^{\mathrm{FD}} = \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2 \left(\sigma_B^2 + \frac{\sigma_Q^2 \sigma_T^2}{\sigma_Q^2 + \sigma_T^2}\right)} \sqrt{\frac{\sigma_Q^2 \sigma_T^2}{\sigma_Q^2 + \sigma_T^2} + \sigma_B^2}.$$

Observing that $\frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2 \left(\sigma_B^2 + \frac{\sigma_Q^2 \sigma_T^2}{\sigma_Q^2 + \sigma_T^2}\right)} = \frac{1}{\sigma_T^2 + \frac{\sigma_B^2}{\sigma_Q^2}(\sigma_Q^2 + \sigma_T^2)} < \frac{1}{\sigma_T^2}$, we see that $\gamma_T^{\mathrm{FD}} < \gamma_T^*$.

4. The value η remains constant as we vary σ_B^2. Taking the derivative of γ_T^{FD} with respect to σ_B^2 gives

$$-\frac{\sigma_Q^2}{2(\sigma_Q^2 + \sigma_T^2)(\eta + \sigma_B^2)^{\frac{3}{2}}} < 0.$$

Taking limits,

$$\lim_{\sigma_B^2 \to 0} \gamma^*_{\text{FD}} = \frac{\sigma_Q^2}{(\sigma_Q^2 + \sigma_T^2)\sqrt{\eta}}$$
$$\lim_{\sigma_B^2 \to \infty} \gamma_T^{\text{FD}} = 0.$$

From the proof of Proposition 4, we see that $\lim_{\sigma_B^2 \to 0} \gamma^*_{\text{FD}} = \lim_{\sigma_B^2 \to 0} \gamma^*_T$.

5. Taking $\gamma_T^{\text{FD}} = \frac{\sigma_Q^2}{(\sigma_Q^2+\sigma_T^2)\sqrt{\sigma_B^2+\eta}}$ as a function of σ_Q^2, σ_B^2, σ_T^2, and η, we can write $\frac{d\gamma_T^{\text{FD}}}{d\sigma_Q^2}$ as

$$\begin{aligned}
\frac{d\gamma_T^{\text{FD}}}{d\sigma_Q^2} &= \frac{\partial \gamma_Q^{\text{FD}}}{\partial \sigma_Q^2} + \frac{\partial \gamma_T^{\text{FD}}}{\partial \eta} \cdot \frac{d\eta}{d\sigma_Q^2} \\
&= \frac{\sigma_T^2}{(\sigma_Q^2 + \sigma_T^2)^2\sqrt{\sigma_B^2 + \eta}} - \frac{\sigma_Q^2}{2(\sigma_Q^2 + \sigma_T^2)(\sigma_B^2 + \eta)^{\frac{3}{2}}} \cdot \frac{\sigma_Q^2 \sigma_T^6 (2\sigma_Q^2\sigma_S^2 + \sigma_Q^2\sigma_T^2 + 2\sigma_S^2\sigma_T^2)}{(\sigma_Q^2 + \sigma_T^2)^2(\sigma_Q^2\sigma_S^2 + \sigma_Q^2\sigma_T^2 + \sigma_S^2\sigma_T^2)} \\
&= \frac{\sigma_T^2}{(\sigma_Q^2 + \sigma_T^2)^2\sqrt{\sigma_B^2 + \eta}} - \frac{\sigma_Q^2}{2(\sigma_Q^2 + \sigma_T^2)(\sigma_B^2 + \eta)^{\frac{3}{2}}} \cdot \frac{\eta \sigma_T^2 (2\sigma_Q^2\sigma_S^2 + \sigma_Q^2\sigma_T^2 + 2\sigma_S^2\sigma_T^2)}{\sigma_Q^2(\sigma_Q^2 + \sigma_T^2)(\sigma_Q^2\sigma_S^2 + \sigma_Q^2\sigma_T^2 + \sigma_S^2\sigma_T^2)} \\
&= \frac{\sigma_T^2}{(\sigma_Q^2\sigma_T^2)^2(\sigma_B^2 + \eta)^{\frac{3}{2}}} \left(\sigma_B^2 + \eta - \eta \cdot \frac{2\sigma_Q^2\sigma_S^2 + \sigma_Q^2\sigma_T^2 + 2\sigma_S^2\sigma_T^2}{2(\sigma_Q^2\sigma_S^2 + \sigma_Q^2\sigma_T^2 + \sigma_S^2\sigma_T^2)} \right) > 0.
\end{aligned}$$

And without doing more algebra, if γ_T^{FD} is positive and increasing in σ_Q^2, and if $\sqrt{\sigma_Q^2 + \sigma_T^2}$ is positive and increasing in σ_Q^2, then clearly their product $\gamma_T^{\text{FD}}\sqrt{\sigma_Q^2 + \sigma_T^2}$ is increasing in σ_Q^2.

Taking limits, as $\sigma_Q^2 \to 0$ it is easy to see from the above formulas that $\lim_{\sigma_Q^2 \to 0} \gamma_T^{\text{FD}} = \lim_{\sigma_Q^2 \to 0} \gamma_T^{\text{FD}}\sqrt{\sigma_Q^2 + \sigma_T^2} = 0$. As $\sigma_Q^2 \to \infty$, we have $\eta \to \frac{\sigma_T^4}{\sigma_T^2+\sigma_S^2}$ and so

$$\lim_{\sigma_Q^2 \to \infty} \gamma_T^{\text{FD}} = \frac{1}{\sqrt{\sigma_B^2 + \frac{\sigma_T^4}{\sigma_T^2+\sigma_S^2}}}$$
$$\lim_{\sigma_Q^2 \to \infty} \gamma_T^{\text{FD}}\sqrt{\sigma_Q^2 + \sigma_T^2} = \infty.$$

We can rewrite $\lim_{\sigma_Q^2 \to \infty} \gamma_T^{\text{FD}}$ as $\frac{(\sigma_Q^2+\sigma_T^2)\sqrt{\sigma_B^2+\frac{\sigma_T^4}{\sigma_T^2+\sigma_S^2}}}{\sigma_T^4+\sigma_B^2(\sigma_T^2+\sigma_S^2)}$ which is less than $\lim_{\sigma_Q^2 \to \infty} \gamma_T^* =$

$\frac{(\sigma_T^2+\sigma_S^2)\sqrt{\sigma_B^2+\frac{\sigma_T^4}{\sigma_T^2+\sigma_S^2}}}{\sigma_T^4}$ from the proof of Proposition 4 part 5. □

Proof of Proposition 8. Follows from arguments in the text. □

Proof of Proposition 9. Given the notation that has been introduced, all of the arguments follow exactly as in Section 4 and Proposition 5. □

Proof of Lemma 5. 1. The optimal period-2 contract under common knowledge of the agent's type sets the average test score to some value κ^*. I first show how to derive κ^* in terms of the period 1 outcome.

As a preliminary step, recall that in the notation of Section 4.2, conditional on any agent type, the payoffs from any set of accepted applicants are determined by the average test score τ and the average z-score ζ. Since the agent accepts k applicants, and applicant test scores have an unconditional distribution that is normal with mean 0 and variance $\sigma_Q^2 + \sigma_T^2$, the range of possible τ is $[-R_T, R_T]$, for $R_T = \sqrt{\sigma_Q^2 + \sigma_T^2} R(k)$, as in (22) and (24). Now let $\bar{\zeta}(\tau)$ be the highest possible ζ at an average test score τ, from Lemma 3, plugging in R_T and R_Z in terms of $R(k)$:

$$\bar{\zeta}(\tau) \equiv \sqrt{R(k)^2 - \frac{\tau^2}{(\sigma_Q^2 + \sigma_T^2)}}.$$

Now suppose that the agent is given full discretion to hire her favorite set of applicants in period 1 and she acts myopically. The average test score $\tau_1 \in [-R_T, R_T]$ is observable to the principal. The average z-score is not directly observable, but the principal can infer that – since the agent's payoff increases in ζ – the average z-score must have been the highest possible level consistent with τ_1, i.e., $\zeta_1 = \bar{\zeta}(\tau_1)$.

If the principal knows the agent type θ, and therefore the induced quantity $\sigma_{U_P}(\theta)$, then the principal's preferences over (τ, ζ) are given by (21). The principal's optimal contract specifies that $\tau_2 = \kappa^*$, where κ^* is the τ component of the pair (τ, ζ) that optimizes (21). Hence, κ^* solves

$$\begin{aligned} \kappa^* &= \arg\max_{\tau} \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2}\tau + \sigma_{U_P}(\theta)\bar{\zeta}(\tau) \\ \Rightarrow\ 0 &= \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} + \sigma_{U_P}(\theta)\bar{\zeta}'(\kappa^*) \\ \Rightarrow\ \kappa^* &= \frac{R(k)\sigma_Q^2\sqrt{\sigma_Q^2 + \sigma_T^2}}{\sqrt{\sigma_Q^2 + \sigma_{U_P}^2(\theta)}}. \end{aligned} \tag{53}$$

Of course, $\sigma_{U_P}(\theta)$ depends on the agent's type, which the principal is trying to learn from the data.[48] But the principal knows his payoff from the first-period choices – this is exactly the average quality level ξ_1. So the principal can plug τ_1 and ξ_1 into (21) (with $V_P = \xi_1$ and $\zeta = \bar{\zeta}(\tau_1)$) to infer $\sigma_{U_P}(\theta)$:

$$\begin{aligned} \xi_1 &= \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2}\tau_1 + \sigma_{U_P}(\theta)\bar{\zeta}(\tau_1) \\ \Rightarrow \sigma_{U_P}(\theta) &= \frac{\xi_1(\sigma_Q^2 + \sigma_T^2) - \sigma_Q^2\tau_1}{\sqrt{(\sigma_Q^2 + \sigma_T^2)((\sigma_Q^2 + \sigma_T^2)R(k)^2 - \tau_1^2)}}. \end{aligned} \tag{54}$$

Now plug this value of $\sigma_{U_P}(\theta)$ into (53) to get (42), i.e.,

$$\kappa^* = \frac{R(k)\sigma_Q^2\sqrt{\sigma_Q^2 + \sigma_T^2}}{\sqrt{\sigma_Q^2 + \frac{(\xi_1(\sigma_Q^2+\sigma_T^2)-\sigma_Q^2\tau_1)^2}{(\sigma_Q^2+\sigma_T^2)((\sigma_Q^2+\sigma_T^2)R(k)^2-\tau_1^2)}}}.$$

The optimal contract in the second period lets the agent accept any k applicants she wants, subject to requiring the period-2 average test score to be κ^* in (42).[49]

2. Now consider the comparative statics on κ^* with respect to τ_1 and ξ_1.

   We know that τ_1 can be any value in $[0, R_T]$, with $R_T = \sqrt{\sigma_Q^2 + \sigma_T^2} \cdot R(k)$ for

[48] Note that one could also solve for the optimal contract even if the observable parameters (k, σ_Q^2, σ_T^2) were to change from period 1 to 2. But one would no longer plug in the period-1 value of $\sigma_{U_P}(\theta)$ into the period-2 payoff expression, since $\sigma_{U_P}(\theta)$ depends on σ_T^2 and σ_Q^2.

[49] One could also solve for the optimal contract even if the observable parameters (k, σ_Q^2, σ_T^2) were to change from period 1 to 2. But one would not simply plug in the period-1 value of $\sigma_{U_P}(\theta)$ from (54) into the period-2 expression (53). The value of $\sigma_{U_P}(\theta)$ depends on σ_T^2 and σ_Q^2, which might change from period to period.

$R(k)$ in (22). Let us bound the range of ξ_1 consistent with an observed τ_1. The principal's payoff in the first period, ξ_1, is equal to $\frac{\sigma_Q^2}{\sigma_Q^2+\sigma_T^2}\tau_1 + \sigma_{U_P}(\theta)\bar{\zeta}(\tau_1)$ from (21). And $\sigma_{U_P}(\theta)$ must be in the range $(0, \sigma_{U_A}(\theta))$ from Lemma 2 part 3.[50] So, given τ_1,

$$\frac{\sigma_Q^2}{\sigma_Q^2+\sigma_T^2}\tau_1 < \xi_1 < \frac{\sigma_Q^2}{\sigma_Q^2+\sigma_T^2}\tau_1 + \sigma_{U_A}(\theta)\bar{\zeta}(\tau_1).$$

Moreover, $\sigma_{U_A}(\theta)$ can be inferred from τ_1: the model predicts that the agent has chosen τ_1 to maximize (20) over τ, with $\zeta = \bar{\zeta}(\tau)$. Taking the first order condition and solving for $\sigma_{U_A}(\theta)$ gives

$$\sigma_{U_A}(\theta) = \frac{\sigma_Q^2}{\tau_1}\sqrt{R(k)^2 - \frac{\tau_1^2}{\sigma_Q^2+\sigma_T^2}}.$$

Plugging this value of $\sigma_{U_A}(\theta)$ along with $\bar{\zeta}(\tau_1)$ into the above sequence of inequalities, we get (after some simplification)

$$\frac{\sigma_Q^2}{\sigma_Q^2+\sigma_T^2}\tau_1 < \xi_1 < \frac{\sigma_Q^2 R(k)}{\tau_1}. \tag{55}$$

Now, return to the comparative statics of κ^* given by (42). κ^* moves in the opposite direction as the fraction $\frac{(\xi_1(\sigma_Q^2+\sigma_T^2)-\sigma_Q^2\tau_1)^2}{(\sigma_Q^2+\sigma_T^2)((\sigma_Q^2+\sigma_T^2)R(k)^2-\tau_1^2}$ as we vary ξ_1 or τ_1. And it is immediate that the fraction is increasing in ξ_1 as long as $\xi_1(\sigma_Q^2+\sigma_T^2) - \sigma_Q^2\tau_1 > 0$, which holds by the left inequality of (55). Next, differentiating, it is straightforward to show that the sign of the derivative of the fraction with respect to τ_1 is equal to the sign of $(\xi_1(\sigma_Q^2+\sigma_T^2) - \sigma_Q^2\tau_1)(\tau_1\xi_1 - \sigma_Q^2 R(k)^2)$. The first parenthetical term is positive, as just described; the second parenthetical term is negative by the second inequality in (55). So κ^* decreases in ξ_1 and increases in τ_1.

□

Proof of Lemma 6. From the specified joint distributions of Q and T, it follows that $\mathrm{Var}(Q) = \sigma_Q^2$, $\mathrm{Var}(T) = \sigma_Q^2 + \sigma_T^2$, and $\mathrm{Cov}(T,Q) = \sigma_Q^2$. It remains to calculate $\mathrm{Var}(U_A)$, $\mathrm{Cov}(T, U_A)$, and $\mathrm{Cov}(Q, U_A)$.

It will be helpful to note as well that $\mathrm{Cov}(S,Q) = \sigma_Q^2$, $\mathrm{Cov}(S,T) = \sigma_Q^2$, and $\mathrm{Var}(S) = \sigma_S^2 + \sigma_Q^2$. The bias term B has variance σ_B^2, and has 0 covariance with S,

[50] The upper bound here need not be tight, depending on parameters.

T, or Q.

From (7),

$$U_A = \mathbb{E}[Q|T,S] + B = \frac{\frac{T}{\sigma_T^2} + \frac{S}{\sigma_S^2}}{\frac{1}{\sigma_Q^2} + \frac{1}{\sigma_T^2} + \frac{1}{\sigma_S^2}} + B.$$

$\mathrm{Cov}(q, \tilde{u}_A)$ is given by

$$\mathrm{Cov}(Q, U_A) = \frac{\frac{\mathrm{Cov}(Q,T)}{\sigma_T^2} + \frac{\mathrm{Cov}(Q,S)}{\sigma_S^2}}{\frac{1}{\sigma_Q^2} + \frac{1}{\sigma_T^2} + \frac{1}{\sigma_S^2}} = \frac{\frac{\sigma_Q^2}{\sigma_T^2} + \frac{\sigma_Q^2}{\sigma_S^2}}{\frac{1}{\sigma_Q^2} + \frac{1}{\sigma_T^2} + \frac{1}{\sigma_S^2}} = \frac{\sigma_Q^4(\sigma_S^2 + \sigma_T^2)}{\sigma_Q^2\sigma_T^2 + \sigma_Q^2\sigma_S^2 + \sigma_T^2\sigma_S^2}.$$

$\mathrm{Cov}(T, U_A)$ is given by

$$\mathrm{Cov}(T, U_A) = \frac{\frac{\mathrm{Var}(T)}{\sigma_T^2} + \frac{\mathrm{Cov}(T,S)}{\sigma_S^2}}{\frac{1}{\sigma_Q^2} + \frac{1}{\sigma_T^2} + \frac{1}{\sigma_S^2}} = \frac{\frac{\sigma_Q^2+\sigma_T^2}{\sigma_T^2} + \frac{\sigma_Q^2}{\sigma_S^2}}{\frac{1}{\sigma_Q^2} + \frac{1}{\sigma_T^2} + \frac{1}{\sigma_S^2}} = \sigma_Q^2 \frac{\frac{1}{\sigma_Q^2} + \frac{1}{\sigma_T^2} + \frac{1}{\sigma_S^2}}{\frac{1}{\sigma_Q^2} + \frac{1}{\sigma_T^2} + \frac{1}{\sigma_S^2}} = \sigma_Q^2.$$

And finally, $\mathrm{Var}(U_A)$ is given by

$$\begin{aligned}
\mathrm{Var}(U_A) &= \frac{\frac{\mathrm{Var}(T)}{\sigma_T^4} + \frac{\mathrm{Var}(S)}{\sigma_S^4} + 2\frac{\mathrm{Cov}(T,S)}{\sigma_T^2\sigma_S^2}}{\left(\frac{1}{\sigma_Q^2} + \frac{1}{\sigma_T^2} + \frac{1}{\sigma_S^2}\right)^2} + \sigma_B^2 = \frac{\frac{\sigma_Q^2+\sigma_T^2}{\sigma_T^4} + \frac{\sigma_Q^2+\sigma_S^2}{\sigma_S^4} + 2\frac{\sigma_Q^2}{\sigma_T^2\sigma_S^2}}{\left(\frac{1}{\sigma_Q^2} + \frac{1}{\sigma_T^2} + \frac{1}{\sigma_S^2}\right)^2} + \sigma_B^2 \\
&= \sigma_Q^2 \frac{\left(\frac{1}{\sigma_Q^2} + \frac{1}{\sigma_T^2} + \frac{1}{\sigma_S^2}\right)\left(\frac{1}{\sigma_T^2} + \frac{1}{\sigma_S^2}\right)}{\left(\frac{1}{\sigma_Q^2} + \frac{1}{\sigma_T^2} + \frac{1}{\sigma_S^2}\right)^2} + \sigma_B^2 = \sigma_Q^2 \frac{\frac{1}{\sigma_T^2} + \frac{1}{\sigma_S^2}}{\frac{1}{\sigma_Q^2} + \frac{1}{\sigma_T^2} + \frac{1}{\sigma_S^2}} + \sigma_B^2 \\
&= \frac{\sigma_Q^4(\sigma_S^2 + \sigma_T^2)}{\sigma_Q^2\sigma_T^2 + \sigma_Q^2\sigma_S^2 + \sigma_T^2\sigma_S^2} + \sigma_B^2. \qquad \square
\end{aligned}$$