

Selecting Applicants

Alex Frankel*

November 6, 2016

Abstract

A firm selects applicants to hire, or a school selects applicants to admit, based on hard information about their quality – call this a test score – along with a biased agent’s privately observed soft information. A contract between the principal and the agent can be thought of as a restriction on acceptance rates as a function of test scores. When the agent’s type – her mix of information and biases – is known to the principal, I give a general characterization of the optimal acceptance rate. When her type is unknown, I solve for the optimal menu of acceptance rates under particular distributional assumptions. In some cases the contracts can be implemented by allowing the agent to accept any applicants she wants subject to a given average test score, or a minimum average test score.

1 Introduction

A firm tends to receive two distinct kinds of information about the quality of its job applicants. First, there is objectively measurable or “hard” information. Each applicant’s education history and years of experience at previous job titles are listed on his or her CV. The firm sees the applicant’s scores on any pre-employment tests that it may have administered. Second, the firm’s hiring managers learn additional “soft” information – privately observed and not directly verifiable – from interviewing the prospective employees and subjectively judging their 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.

*I thank Manuel Amador, Kyle Bagwell, Ben Brooks, Ging Cee Ng, Lisa Kahn, Danielle Li, 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 larger organization. For instance, 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, even though they perform no better if hired. Or the manager may have a more systematic bias. She may be 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 set of applicants from the pool, her choices would be distorted by these biases. The firm might consider eliminating the manager’s discretion entirely. That is, the firm could require that the manager hire those with the most favorable hard information, e.g., the highest scores on a pre-employment test. 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 with a sufficiently biased manager, no discretion would improve on full discretion. Indeed, in the empirical setting of Hoffman, Kahn, and Li (2015), managers who observed pre-employment test scores of job applicants and then hired their favorites did select worse applicants than if they had just hired the ones with the highest test scores.

Of course, the firm is not limited to the two extreme policies of full discretion or no discretion. The firm can use the hard information to constrain the manager’s hiring decisions in an arbitrary manner. In this paper I search for the optimal contracts: the rules which maximize the expected quality of a set of applicants selected from a large pool.

I begin my analysis by supposing that the firm knows the manager’s “type” – her exact mix of information and bias across applicants. For concreteness, call the realizations of hard information “test scores”. Here, the optimal policy takes the form of a specified acceptance rate as a function of the test score. Subject to this acceptance rate, the manager chooses which applicants to hire at each score. I provide a general approach for finding the optimal acceptance rate function, and then I apply the results to a normal specification: normal distribution of applicant quality, normal signals of quality from hard and soft sources, and normally distributed idiosyncratic biases. Under the normal specification, the acceptance rate function should follow a normal CDF – that is, an S-shape. A larger share of applicants are accepted at higher test scores, with acceptance rates approaching zero at the lowest scores and one hundred percent at the highest scores. A manager with either stronger biases or less information faces a steeper acceptance rate: her hiring depends more on the test score and less on her personal judgment.

There is also an alternative implementation of these optimal contracts. The contract al-

lows the manager to accept any applicants she wants, subject to pinning down some moment of the distribution of test scores of those who are accepted. Under the normal specification, this alternative implementation is particularly simple: the contract specifies the average test score.

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 score across the applicants that the manager accepts.

The applied contribution of this paper is to characterize the implementation and the comparative statics of contracts that combine hard and soft information in selecting applicants. 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. In the employment context, hard information such as education and work histories has always been available to employers. But a big change in recent years has been the growth in pre-employment tests.¹ Autor and Scarborough (2008) and Hoffman, Kahn, and Li (2015) present data showing how the introduction of pre-employment tests has affected the hiring processes at certain firms. In the market for consumer loans, Einav, Jenkins, and Levin (2013) look at the impact of improved hard information from the adoption of automated credit scoring. Of these papers, Autor and Scarborough (2008) and Einav, Jenkins, and Levin (2013) are motivated by a model in which there is no agency problem between the managers and the firm and hence no contracting problem to solve. New information simply lets the organization make better decisions. Hoffman, Kahn, and Li (2015) do address the tradeoff of bias and information;

¹According to the Wall Street Journal (2015), the number of employers using pre-hire assessments rose from 26% in 2001 to 57% in 2013. From the article:

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... Results fed to hiring managers can be as simple as a green, yellow or red light indicating the scoring algorithm's recommendations—or run into pages of detail about a candidate's performance.

the current paper embeds their theoretical model. That paper compares full discretion (the policy used by the firm in their data set) to a hypothetical of no discretion, whereas I search for an optimal contract.

A deeper theoretical contribution of this paper is to show how to think about and solve for optimal contracts without transfers in a new principal-agent environment. Here, there are many binary decisions to be made, that of accepting or rejecting each applicant. There is also hard information, a test score, attached to each decision.²

The bulk of the work on contracting over actions, in what is called the delegation literature, involves a single decision to be made. The agent has private information on a one-dimensional state of the world which affects both players' preferences over a one-dimensional action. 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). What is novel in the current paper, both in terms of the tradeoffs it generates and in the levers it makes available to the principal, is the existence of hard information at each decision.

Che, Dessein, and Kartik (2013) looks at the role of hard information about different job applicants – asymmetric priors on quality – when a firm will hire at most one of them. The key distinction between the model of that paper and the current one is the form of the bias. Their misalignment is over *how many* candidates to hire, zero or one, whereas mine is over *which* candidates to hire. The agent in Che, Dessein, and Kartik (2013) has the same ranking of candidates as does the principal, but has a bias towards hiring. I fix the number of candidates to be hired, but take the agent's preferences over candidates to be imperfectly correlated with those of the principal.³

In my model, in which hard information is available for a large number of applicants, the (highly dimensional) analog of the delegation literature's "action" is the function mapping test scores to acceptance rates.⁴ The "state" corresponds to the agent's type, i.e., the parameters describing her bias and information. When the agent's type is known, the principal

²Stein (2000) looks at tradeoffs over different organizational forms when a manager's information may either be hard or soft. That paper focuses on the incentives of a manager to invest in information-gathering; I take the manager's information as exogenous.

³Armstrong and Vickers (2010) and Nocke and Whinston (2013) consider a different sort of 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.

⁴Formally, as described in Section 2.2, this action space would describe deterministic contracts. Stochastic contracts may be more general.

specifies a single acceptance rate function – a single action – for the agent. This acceptance rate function equalizes the quality of the marginal accepted applicant at each test score. When the agent’s type is unknown, the principal offers a menu to screen across agent types. To derive the optimal menu under the normal specification, I formally map my model into a delegation setting with one-dimensional states and actions, plus “money-burning” – an auxiliary action which hurts both players. Conditions in the one-dimensional literature for the optimality of an interval delegation set in which money is not burnt can be translated into sufficiency conditions for the principal to allow a particular one-dimensional set of acceptance rate functions. Specifically, I apply extensions of results from Amador and Bagwell (2013).⁵ In deriving this map, the current paper extends the domain of the delegation literature to a new type of contracting problem.

I now move into the analysis. Omitted proofs can be found in Appendix F.

2 The Model

2.1 Players, Payoffs, and Information

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

The utilities to the principal and agent of hiring applicant $i \in \mathcal{I}$ are u_P^i and u_A^i , respectively. Their realized payoffs in the game – call these V_P and V_A – will be the average values of u_P^i and u_A^i across the population of all hired workers i .

Utilities are determined by an applicant’s *quality* $q^i \in \mathbb{R}$ along with the agent’s *bias*

⁵Versions of delegation problems with money-burning also appear in Kováč and Mylovanov (2009), Ambrus and Egorov (2013), and Amador and Bagwell (2016). Similarly to the current paper, Guo (2014) solves a contracting problem without money by reducing a highly dimensional action space into a one-dimensional frontier, plus money-burning.

towards this applicant $b^i \in \mathbb{R}$.⁶

$$\begin{aligned} u_P^i &= q^i \\ u_A^i &= q^i + b^i. \end{aligned}$$

The distribution of applicant quality q in the population is F_Q .⁷ Each applicant’s quality is imperfectly observed by the players through two signals, a public one and a private one. The public signal realization or “test score” for applicant i is denoted by $t^i \in \mathcal{T}$ and the private signal realization by $s^i \in \mathcal{S}$. \mathcal{T} and \mathcal{S} are subsets of finite-dimensional real coordinate spaces.⁸ The test score t^i is *hard information* – it is observable to both players and can be contracted upon. The private signal s^i is *soft information* – it is observed only by the agent and can never be externally verified or audited. Given quality q for an applicant, the test score t is distributed according to $F_T(\cdot|q)$. The distribution of s conditional on q and t is $F_S(\cdot|q, t)$.

The agent privately observes her bias b^i for each applicant i . The distribution of the agent’s bias is $F_B(\cdot|s, t)$. This distribution may depend on the signals but does not depend directly on quality. In other words, the realization of b is uninformative about q ; the agent’s private information on quality is fully captured by s .

Let me highlight that, in this model, the agent does not need to make any costly investments to acquire her s and b realizations. In many applications related to hiring and admissions, an agent would have to invest time or effort to learn about each applicant. In that case hard information might be gathered first, to be used as a first screen to decide which applicants the agent should evaluate.⁹

The model primitives described above consist of one parameter and four distribution functions: the share k of applicants hired, the quality distribution F_Q , the public and private signal distributions F_T and F_S , and the bias distribution F_B . The three objects k , F_Q , and F_T are properties of the applicant pool, and how it is judged by the firm. I assume throughout

⁶Prior to any distributional assumptions, the assumption that bias and quality are additive in the agent’s utility is fully general. The bias term b^i simply designates the difference between the principal and agent utilities for applicant i . Under distributional assumptions such as independence of b and q , additivity becomes economically meaningful.

⁷As a matter of notation, q represents a generic quality realization, with q^i the realization for applicant i ; the quality distribution is labeled with the capital letter Q . Similar conventions are used for other variables.

⁸In mathematical examples and in informal discussion, I often treat signals as single-dimensional.

⁹As described in the Wall Street Journal (2014), “As the hiring process gets more automated and employers begin incorporating more data into hiring, [pre-employment] tests are used more often and earlier in the process to winnow applicants for specific jobs.” At the firms using these tests, low scoring applicants “rarely get the chance to interview for a job.”

that there is common knowledge over these “principal fundamentals.”

Assumption 1. The quality distribution F_Q , the test score distribution F_T , and the share of applicants to be hired k are commonly known at the start of the game.

Common knowledge of F_Q establishes that there is no aggregate uncertainty over the distribution of quality in the applicant pool. Taking F_T to be commonly known yields two additional implications. First, there is no prior uncertainty over the empirical distribution of test scores. Second, at each test score, there is no uncertainty over the distribution of applicant qualities at that score. The assumption that k is known at the start of the game implies that the firm will never give an agent flexibility over how many applicants to hire. I discuss a relaxation of Assumption 1 in Section 5.2.

The two distribution functions F_S and F_B relate to how the agent evaluates applicants. Call this pair of functions the agent’s *type*, describing the extent of her information and her biases. In the pool of potential hiring managers, some may be better or worse than others at evaluating job applicants; and some may care more or less about hiring the right applicants for the firm rather than the ones they personally like. The agent’s prior information about F_B is not important, but I assume throughout that the agent knows F_S and hence knows how to interpret her private signals. 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 is uncertain of the agent’s type (Section 4).

2.2 Contracting

The principal seeks a contract or set of rules that incentivize the agent to use her private information to select high quality applicants. The timing of the game is as follows.

- i. The principal gives the agent a contract.
- ii. The agent’s type is realized.
- iii. The agent has the opportunity to send an initial message.
- iv. For each applicant i , the test score t^i is realized and is publicly observed by the principal and the agent. At the same time, the agent privately observes each s^i and b^i .
- v. The agent has the opportunity to send an interim message.
- vi. The contract determines which applicants are hired, possibly stochastically, as a function of the two messages and the vector of applicant test scores.

The principal chooses the contract, and the agent sends messages within the contract, so as to maximize their expected payoffs – $V_P = \mathbb{E}[q^i | i \text{ hired}]$ and $V_A = \mathbb{E}[q^i + b^i | i \text{ hired}]$ – over all relevant uncertainty. The agent has no choice of whether to participate. The contract determines the message spaces as well as the mapping from messages and test scores into (probabilities of) acceptances for each applicant. Because applicants are ex ante identical, it is without loss of generality in searching for an optimal contract to restrict attention to anonymous contracts, i.e., those which treat applicants identically regardless of their index i .

Under any contract, exactly k applicants must be accepted. The only output of the contract is the determination of which applicants are to be hired; there are no monetary transfers and no other rewards or punishments to the agent. The principal commits to follow this contracting outcome.

The contract timing as described above is fully general in that it allows for direct revelation mechanisms: the agent can report her type in the initial message (if this was not already known to the principal), and report her observations of s and b for each applicant in the interim message. Note that test scores are publicly observed, and so the principal does not have the option of withholding the test scores from the agent before asking her to report her soft information. As such, the agent knows how her interim message will be translated by the contract into hiring outcomes. The interim message can thus be thought of as “indirectly” assigning acceptances, or probabilities of acceptances, to each applicant subject to constraints imposed by the contract and the initial message.

In fact, thanks to Assumption 1, it will actually be without loss of generality to eliminate the initial message stage (step iii). There is no useful screening that can occur in between the agent’s type realization and the arrival of the applicants, because Assumption 1 ensures that the agent can predict in advance the joint distribution over (t, s, b) in the applicant pool. The only uncertainty is over the mapping from applicant identities to these realizations.

Let me make precise this argument for the redundancy of initial messages. The agent’s messages in a contract determine the distribution of acceptance probabilities at each test score. Fixing $t \in \mathcal{T}$, let $\pi(\cdot; t)$ denote the CDF of acceptance rates of those applicants with test score t : $\pi(x; t)$ is the share of applicants with test score t who are accepted with probability x or less, for $x \in [0, 1]$.¹⁰ I assume that any $\pi(x; t)$ function that can be induced by a contract is jointly measurable in (x, t) .

¹⁰For every t , it holds that $\pi(x; t)$ is weakly increasing in x on the range $[0, 1]$, with $\pi(1; t) = 1$. $\pi(0; t)$ indicates the share of applicants with test score t that are rejected with certainty, and $1 - \lim_{x \rightarrow 1} \pi(x; t)$ the share that are accepted with certainty.

Now consider a contract of the form above, steps i-vi, with both an initial and an interim message. By the same logic implying that contracts can be taken to be anonymous, contracts can also be taken to be anonymous conditional on the agent’s initial message: all applicants are ex ante identical at the time that report is made. Anonymity means that the agent must be able to “permute” her interim message across any applicants with the same observable test score, and thereby permute their probabilities of acceptance. In other words, the contract at the interim stage is a restriction only on the set of feasible π functions. Every initial message leads to some set of feasible π functions, one of which will be induced by the agent’s interim message. Given some induced π , the agent has full flexibility to assign applicants to probabilities within each test score.

By Assumption 1, the agent can perfectly predict which π function she will end up inducing conditional on any initial message, and what her corresponding payoff will be. The agent will send the initial message which lets her choose the highest-payoff π at the interim stage. The outcome of the contract is unchanged if we allow the agent to choose from the union of π functions across all initial messages at the interim message stage. Hence, the initial message is redundant.

Going forward, I take contracts to have no initial message – no step iii – and I refer to “the message” as the interim message of step v. A contract can be thought of as a feasible set of π functions. The message both selects one such π and assigns applicants to probabilities.

I say that a contract is deterministic if each applicant is accepted with probability one or zero, conditional on the test scores and the message. Stochastic contracts involve randomized acceptances.^{11,12} In the notation above, a deterministic contract is one in which, for any message and any t , $\pi(x; t)$ is constant over $x \in [0, 1)$. Deterministic contracts can be more simply described by the induced acceptance rate at each test score than by the full function $\pi(x; t)$. Denote a representative such acceptance rate function by $\alpha : \mathcal{T} \rightarrow [0, 1]$, with $\alpha(t)$

In the case that test scores are continuously distributed, one should calculate the conditional CDF $\pi(\cdot; t)$ in the standard manner: at each t , evaluate the CDF in a band of test scores around t , and then take the limit as the bandwidth goes to 0.

¹¹For a stochastic contract, the correlation structure over acceptances is immaterial, given that principal and agent payoffs are additive across hired applicants. The constraint that exactly k applicants are hired at every realization can be satisfied as long as the expected number of hires, given any message, is k .

¹²From an ex ante perspective, the assignment of applicant indices i to realizations (q^i, t^i, s^i, b^i) is essentially random. So a deterministic non-anonymous contract may correspond to a stochastic anonymous contract. For instance, label the set of applicants \mathcal{I} as the interval $[0, 1]$. Deterministically accepting applicant i if and only if $i \in [0, 1/2]$ is equivalent, in terms of the joint distribution of (q, t, s, b) of hired applicants, to anonymously accepting each applicant with probability $1/2$.

indicating the acceptance rate at test score t :

$$\alpha(t) = \text{Prob}[\text{Applicant } i \text{ accepted} | t^i = t] = 1 - \pi(0; t).$$

Repeating the discussion from above, any deterministic contract can be thought of as a restriction on the set of acceptance rate functions. Suppose one set of k applicants can be deterministically selected, and this set of applicants implies an acceptance rate function of α . Then by anonymity, any other set of k applicants inducing the same acceptance rate at each test score can also be selected.

The distinction between deterministic and stochastic contracts helps drive home the role of having a mass of applicants rather than a single applicant in the model. Letting the agent deterministically select a share $\alpha(t)$ of applicants of those with test score t , from a large population, induces different expected payoffs than does stochastically selecting one applicant with probability $\alpha(t)$ if he or she gets test score t .

2.3 Examples

To make the model more concrete, I now introduce a pair of examples which illustrate the kinds of biases and information structures that may be come up in different applications. In the *normal specification*, the agent has an “idiosyncratic” bias for each applicant, independent of all other terms. The *two-factor model* shows how one might capture a “systematic” bias: the soft and hard information are informative about different aspects of a job candidate, and the agent values these aspects differently than does the principal. In that case the agent’s bias will be correlated with her soft information. In Appendix C.2, I describe how to capture some other forms of systematic bias. For instance, the principal may be in favor of affirmative action for job applicants with certain observable qualities while the agent disagrees, or vice versa.

2.3.1 Normal Specification

Under the *normal specification*, assume that

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

$$t \sim \mathcal{N}(q, \sigma_T^2), \tag{F_T}$$

$$s \sim \mathcal{N}(q, \sigma_S^2) \text{ for all } t, \tag{F_S}$$

$$b \sim \mathcal{N}(0, \sigma_B^2) \text{ for all } s, t \tag{F_B}$$

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

This specification lets us capture the key forces of the model with a small number of parameters – one parameter for each distribution function. The parameter σ_Q^2 is the variance of quality in the population, with mean quality normalized to 0. Then σ_T^2 and σ_S^2 describe how informative each signal is about quality: a variance going to 0 would be perfectly informative, and infinity would be uninformative. Finally, σ_B^2 tells us the strength of the agent’s biases.¹³ The agent’s utility for an applicant is $q + b$, so an agent with a higher σ_B^2 is more biased in the sense that her utility comparisons across applicants depend less on variation in q and more on variation in b . I interpret the independence of b with other terms – specifically, t and s – as representing an agent’s “idiosyncratic” bias.

Assumption 1 states that principal fundamentals σ_Q^2 and σ_T^2 are common knowledge. The agent’s type consists of the two dimensions (σ_S^2, σ_B^2) – information and bias.

2.3.2 Two-Factor Model

Now consider the possibility that the test score and the private signal may be informative about different aspects of a candidate. In the *two-factor model*, let the principal’s utility for applicant i be given by

$$u_P^i = q_1^i + q_2^i.$$

The two quality factors q_1 and q_2 follow some joint distribution F_{Q_1, Q_2} . Let t follow a distribution $F_T(\cdot | q_1, q_2)$ and s follow a distribution $F_S(\cdot | q_1, q_2, t)$. Assume that these signal distributions are such that the private signal s does not add any information about the first

¹³The normalization of the mean of b to 0 is without loss of generality. Adding a constant to the agent’s utility of hiring each applicant would not change her ranking of applicants, or sets of hired applicants.

factor:

$$\mathbb{E}[q_1|s, t] = \mathbb{E}[q_1|t].$$

In other words, the agent gets private information only about the “second factor.” Think of the public test score as measuring a job candidate’s technical ability, while the private interview with a hiring manager yields soft information about the candidate’s social skills. Or, in college admissions, an applicant’s public SAT score and GPA good indicators of academic skills, whereas the admissions officer subjectively assesses “holistic” aspects such as leadership and “fit.” In the same vein, the agent might be an expert brought in to evaluate candidates on features related to her specialty: a coach evaluates athletic ability, or a writing instructor reads and scores the application essay.

Now, let the agent’s utility for applicant i be

$$u_A^i = q_1^i + \lambda q_2^i, \text{ for } \lambda > 0.$$

This utility function introduces a “systematic” bias for $\lambda \neq 1$: the principal and agent value the factors differently. We might expect that a hiring manager would overemphasize the importance of social skills, a writing instructor would overemphasize writing ability, and a coach would overemphasize athletics. These biases, in which the agent values the skill that she is better at evaluating more highly than does the principal, correspond to $\lambda > 1$. I call such an agent an “advocate”. 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.

Let us now rewrite the utilities in the notation of Section 2.1. The principal’s utility is $u_P^i = q^i$, so we have $q^i = q_1^i + q_2^i$. Next, we seek to write the agent’s utilities in the form of $u_A^i = q^i + b^i$. The agent maximizes the expectation of $q_1^i + \lambda q_2^i = q^i + (\lambda - 1)q_2^i$ across hired applicants. Subject to her information, she maximizes the expectation of $u_A^i = q^i + (\lambda - 1)\mathbb{E}[q_2^i|t^i, s^i]$. This expression is equal to $q^i + b^i$ for

$$b^i = (\lambda - 1) \mathbb{E}[q_2^i|t^i, s^i].$$

As required by the formulation of the model, the (degenerate) distribution of the bias depends only on the realizations of the signals.¹⁴

¹⁴I have also written the signal distributions as functions of q_1 and q_2 , but this formulation is embedded in the model with $q \sim F_Q$, $t \sim F_T(\cdot|q)$ and $s \sim F_S(\cdot|t, q)$. First, for any distribution F_{Q_1, Q_2} , we can find the induced distribution F_Q of their sum $q = q_1 + q_2$. Next, we find the joint distribution of q , t , and s , which

An agent’s type in the two-factor model corresponds to a systematic bias parameter λ and a signal structure F_S . We see that systematic biases manifest themselves as a correlation between bias b and signals (t, s) . An advocate with $\lambda > 1$ will be biased in favor of those for whom the private signal reveals positive news on q_2 . A cynic with $\lambda < 1$ will be biased in the reverse manner, and will have bias b^i negatively correlated with her signal of quality.

As written, there is no idiosyncratic bias. One could of course add an idiosyncratic bias term, an independent “epsilon,” to the agent’s utility for each applicant. In Appendix D I write down and analyze a model which combines the systematic biases of the two-factor model with the idiosyncratic biases of the normal specification. There the agent’s type comprises three dimensions: the amount of information, the systematic bias, and the strength of the idiosyncratic bias.

Another distinction to highlight between the two-factor model and the normal specification is that under the normal specification, signals were conditionally independent given q . In the two-factor model, the fact that signals are differentially informative about distinct “quality factors” leads to conditional dependence. Suppose that the distribution of t depended only on q_1 ; the distribution of s depended only on q_2 ; and q_1 and q_2 were realized independently. Then s and t would be *unconditionally* independent. But we would expect s and t 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 Preliminaries

Denote the agent’s belief on expected quality of applicant i by \tilde{q}^i and her expected utility from hiring this applicant by \tilde{u}_A^i .¹⁵

$$\tilde{q}^i \equiv \mathbb{E}_A[q^i | t^i, s^i] \tag{1}$$

$$\tilde{u}_A^i \equiv \mathbb{E}_A[u_A^i | t^i, s^i, b^i] = \tilde{q}^i + b^i. \tag{2}$$

For the analysis that follows, I will maintain the following technical assumption which states that the agent almost surely has a strict preference between any two applicants with the same test score.

Assumption 2. For each t , the distribution of \tilde{u}_A^i conditional on $t^i = t$ has no atoms.

lets us calculate $F_T(\cdot | q)$ and $F_S(\cdot | q, t)$, the signal distributions.

¹⁵I write \mathbb{E}_A to indicate that the expectation is taken with respect to the agent’s information. In particular, the agent knows F_S (even if it is not commonly known) and thus knows how to interpret the signal s^i .

This condition can arise from continuously distributed biases at each t , or a continuously distributed belief on quality arising from the agent’s private signals.

Consider the agent’s choices under some given contract. The agent’s message induces some CDF $\pi(\cdot; t)$ of acceptance probabilities at test score t . But which applicants with this test score are assigned to which acceptance probabilities? Of course, the agent will assign weakly higher acceptance probabilities to applicants with higher \tilde{u}_A . That is, at any fixed test score, acceptance probabilities will be *monotonic* in agent utilities.

We can combine monotonicity with Assumption 2 to simplify the expression of which applicants are assigned to which acceptance probabilities. Instead of keeping track of individual applicants, we need only keep track of the share of applicants accepted at each (t, \tilde{u}_A) pair because applicants with the same (t, \tilde{u}_A) will (almost always) be treated identically. In particular, suppose that different applicants with the same (t, \tilde{u}_A) were accepted with different probabilities from one another. Then the principal’s payoff would depend on the \tilde{q} values of the accepted versus rejected applicants at (t, \tilde{u}_A) , not just the total share accepted. But, thanks to monotonicity, at each t there can be only countably many \tilde{u}_A at which applicants with the same \tilde{u}_A have different acceptance probabilities.¹⁶ And Assumption 2 says that this countable set of \tilde{u}_A must occur with zero probability conditional on t . So behavior at these points is payoff-irrelevant, and can be disregarded.

In other words, two applicants with the same test score t and same agent utility \tilde{u}_A are *indistinguishable* by any contract. Without loss of generality, in any contract and under any equilibrium message, these two applicants will be accepted with the same probability. Summarizing this discussion:

Observation 1. Fix any contract. Take any pair of applicants i and j with $t^i = t^j$. It is without loss of generality to suppose that under any equilibrium message:¹⁷

1. **Distinguishability.** If $\tilde{u}_A^i = \tilde{u}_A^j$ then applicant i and j have the same probability of acceptance.
2. **Monotonicity.** If $\tilde{u}_A^i > \tilde{u}_A^j$ then applicant i has a weakly higher probability of acceptance than applicant j .

¹⁶Look at the minimum and maximum acceptance probabilities across applicants at each (t, \tilde{u}_A) pair. Both of these functions are monotonic in \tilde{u}_A and hence, at each t , have at most countably many discontinuities in \tilde{u}_A . Moreover, any point at which (t, \tilde{u}_A) at which the minimum is strictly below the maximum must be a point at which these functions are discontinuous in \tilde{u}_A .

¹⁷The results are without loss in the sense that any equilibrium message yields identical payoffs for both parties as one which satisfies these properties. But the agent can always “deviate” on a set of applicants of measure 0 without affecting payoffs.

Given distinguishability, the outcome of a message in a contract can be summarized by a acceptance rule $\chi : \mathcal{T} \times \mathbb{R} \rightarrow [0, 1]$, where $\chi(t, \tilde{u}_A)$ indicates the probability that an applicant with test score t and agent utility \tilde{u}_A is accepted. Monotonicity establishes that $\chi(t, \tilde{u}_A)$ is weakly increasing in \tilde{u}_A for every t . A deterministic contract would be one for which $\chi(t, \tilde{u}_A) \in \{0, 1\}$ for all (t, \tilde{u}_A) under every equilibrium message.

Because all applicants with the same test score and same agent utility are treated identically, the principal may as well “lump them all together.” Some of them may have high quality \tilde{q} and low bias b , while others have low quality and high bias. But distinguishability states that no contract can induce the agent to treat these applicants differently. Define $\hat{u}_P(t, \tilde{u}_A)$ to be the average principal utility – that is, the average quality – of applicants with test score t (observed by the principal) and agent utility \tilde{u}_A (unobserved).¹⁸

$$\hat{u}_P(t, \tilde{u}_A) \equiv \mathbb{E}[u_P | t, \tilde{u}_A] = \mathbb{E}[q | t, \tilde{u}_A]. \quad (3)$$

I sometimes write the value of this function for applicant i as $\hat{u}_P^i = \hat{u}_P(t^i, \tilde{u}_A^i)$.

The principal’s payoff under a selection of applicants described by χ can now be written as $\frac{1}{k}$ times the expectation of $\chi(t, \tilde{u}_A) \cdot \hat{u}_P(t, \tilde{u}_A)$ over realizations of t, \tilde{u}_A .

3 Common Knowledge of Agent Type

In this section I consider optimal contracts under *common knowledge of the agent’s type*: the distributions F_B and F_S are commonly known prior to contracting. Combining common knowledge of agent type with common knowledge of F_Q and F_T from Assumption 1, the principal knows the induced joint distribution of (t, \tilde{u}_A) across applicants; denote this distribution by Ω . The principal knows the function \hat{u}_P mapping (t, \tilde{u}_A) to expected utilities. And, given any contract, the principal also knows in advance the acceptance rule χ describing the accepted applicants that will be induced by an agent’s optimal message.

It must hold that the induced χ selects a total of k applicants, and that χ is monotonic

¹⁸The use of a hat rather than a tilde highlights that the expectation is taken with respect to different information than \tilde{u}_A . This expectation is evaluated conditional on the agent’s type, which the principal does not necessarily know. Note that the value $\hat{u}_P(t, \tilde{u}_A)$ is not necessarily well-defined for \tilde{u}_A outside the support of those values that can arise under t .

in agent utility. Letting $\text{Supp}_{U_A}(t)$ denote the support of \tilde{u}_A conditional on t , under Ω :

$$\mathbb{E}_\Omega[\chi(t, \tilde{u}_A)] = k \tag{4}$$

$$\text{For all } t, \chi(t, \tilde{u}_A) \text{ is weakly increasing in } \tilde{u}_A \text{ over } \text{Supp}_{U_A}(t). \tag{5}$$

The notation \mathbb{E}_Ω emphasizes that (t, \tilde{u}_A) is drawn according to Ω .

In fact, any χ satisfying the above conditions can be implemented by some contract. Take some such acceptance rule χ . This χ induces a (commonly known) CDF of acceptance probabilities $\pi(\cdot; t)$ at each test score t . A contract can specify that the agent select applicants according to that induced π . And given such a contract, the agent's behavior of monotonically assigning higher acceptance probabilities to higher utilities at each test score recovers χ .

So the principal's problem under common knowledge can be stated as maximizing

$$\frac{1}{k} \mathbb{E}_\Omega[\chi(t, \tilde{u}_A) \cdot \hat{u}_P(t, \tilde{u}_A)] \tag{6}$$

over the choice of functions $\chi : \mathcal{T} \times \mathbb{R} \rightarrow [0, 1]$, subject to (4) and (5).

3.1 Solving a relaxed problem

One upper bound on the principal's payoff would result from maximizing the objective (6) subject to (4), but without imposing (5): accept k applicants but do not require monotonicity. Let this *upper bound acceptance rule*, or UBAR, be described mathematically by the acceptance rule $\chi^{\text{UBAR}}(t, \tilde{u}_A)$. 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 \hat{u}_P^c such that a share of k agents have $\hat{u}_P^i \geq \hat{u}_P^c$; formally,

$$\hat{u}_P^c \equiv \sup \{x \in \mathbb{R} \mid \text{Prob}_\Omega[\hat{u}_P(t, \tilde{u}_A) \geq x] \geq k\}.$$

UBAR accepts all applicants with \hat{u}_P above the cutoff and rejects all of those below: $\chi^{\text{UBAR}}(t, \tilde{u}_A) = 1$ if $\hat{u}_P(t, \tilde{u}_A) > \hat{u}_P^c$ and $\chi^{\text{UBAR}}(t, \tilde{u}_A) = 0$ if $\hat{u}_P(t, \tilde{u}_A) < \hat{u}_P^c$. If there is a mass of applicants with $\hat{u}_P = \hat{u}_P^c$, then there is flexibility over which of these specific applicants are accepted in order to get k applicants to be hired in total. Without loss, over

any region of flexibility let χ^{UBAR} take values in $\{0, 1\}$ and let it be monotonic in \tilde{u}_A .¹⁹

UBAR is deterministic by construction: χ^{UBAR} takes values only in $\{0, 1\}$. So if UBAR is monotonic, and therefore yields an optimal contract, then this optimal contract can be implemented by specifying the appropriate acceptance rates at each test score. Let $\alpha^{\text{UBAR}}(t)$ be the share of applicants accepted at test score t under χ^{UBAR} .

The following condition guarantees that UBAR will in fact be monotonic:

Definition. Utilities are *aligned up to distinguishability* if for all t , the principal’s expected utility $\hat{u}_P(t, \tilde{u}_A)$ is weakly increasing in \tilde{u}_A over $\text{Supp}_{U_A}(t)$.

That is, utilities are aligned up to distinguishability if better applicants from the agent’s perspective – having higher quality plus bias of $\tilde{u}_A = \tilde{q} + b$ – also tend to be of higher quality \tilde{q} in expectation.²⁰ Under this alignment, at every test score the ordering of applicants by agent utility \tilde{u}_A is the same as by principal utility \hat{u}_P . Therefore UBAR will be monotonic: it accepts those with higher \tilde{u}_A over those with lower \tilde{u}_A .

Figure 1 illustrates the upper bound acceptance rule and alignment up to distinguishability. It is certainly possible to construct distributions under which the condition is violated, even when biases b are independent of s and t . I confirm below, however, that this condition does hold for the normal specification (Section 3.3), the two-factor model (Section C.1), and for a mixture of the two (Section D).

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 score t of $\alpha(t) = \alpha^{\text{UBAR}}(t)$ implements the upper bound acceptance rule. This contract is an optimal contract.

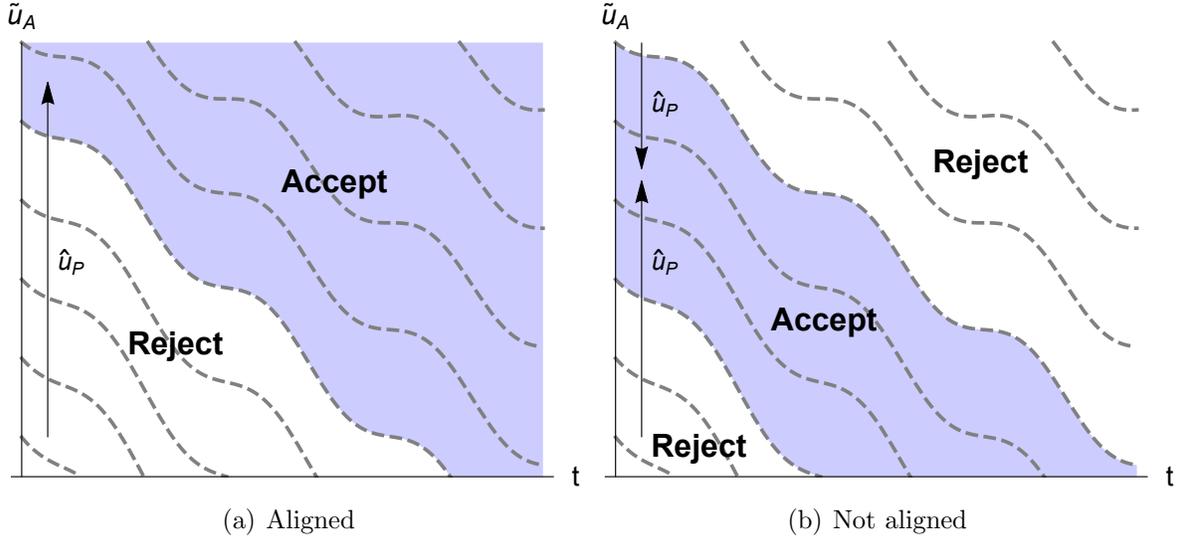
An alternative implementation. The optimal contract of Proposition 1, implementing UBAR, describes the acceptance rate at every test score. There is another implementation of UBAR which may in some cases be simpler to express. The contract will (i) require the agent to select exactly k applicants, and (ii) require that $\mathbb{E}[C(t^i)|i \text{ hired}] = K$, for

¹⁹That is, define some test-score-specific cutoff agent utilities. For those values \tilde{u}_A for which $\hat{u}_P(t, \tilde{u}_A) = \hat{u}_P^c$, let $\chi^{\text{UBAR}}(t, \tilde{u}_A) = 0$ for \tilde{u}_A below the cutoff and let $\chi^{\text{UBAR}}(t, \tilde{u}_A) = 1$ for \tilde{u}_A above the cutoff.

²⁰This condition is sufficient but not strictly 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 $k \in (0, 1)$.

One sufficient condition for utilities to be aligned up to distinguishability is that, for each t , the conditional distribution of \tilde{u}_A (over realizations of s and b) satisfies the monotone likelihood ratio property in q .

Figure 1: The upper bound acceptance rule and alignment up to distinguishability



The dashed curves show possible principal indifference (iso- \hat{u}_P) curves in (t, \tilde{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).

some “moment function” $C(\cdot)$ and some constant K . One can think of such a contract as restricting one moment of the distribution of test scores of those who are hired, rather than the full distribution of scores.

When χ^{UBAR} is monotonic, every applicant with \tilde{u}_A above some test-score-specific cutoff is accepted and every applicant with \tilde{u}_A below is rejected. Under a monotonic χ^{UBAR} , let $\tilde{u}_A^c(t)$ be the cutoff at test score t : take $\tilde{u}_A^c(t) \in \mathbb{R} \cup \{-\infty, \infty\}$ such that for $t \in \mathcal{T}$ and $\tilde{u}_A \in \text{Supp}_{U_A}(t)$, $\chi^{\text{UBAR}}(t, \tilde{u}_A) = 1$ if $\tilde{u}_A > \tilde{u}_A^c(t)$ and $\chi^{\text{UBAR}}(t, \tilde{u}_A) = 0$ if $\tilde{u}_A < \tilde{u}_A^c(t)$. In other words, $\tilde{u}_A^c(t)$ is the function describing the cutoff indifference curve of Figure 1, Panel (a).²¹

We can now express an alternative implementation for the optimal contract UBAR.

Proposition 2. Under common knowledge of the agent’s type, suppose that utilities are

²¹If all applicants are accepted at some test score t , then $\tilde{u}_A^c(t)$ can be set as $\inf \text{Supp}_{U_A}(t)$, which is $-\infty$ if $\text{Supp}_{U_A}(t)$ has no lower bound in \mathbb{R} . Likewise if all applicants are rejected, with $\tilde{u}_A^c(t)$ equal to $\sup \text{Supp}_{U_A}(t)$, possibly ∞ . Note that if $\tilde{u}_A^c(t) = -\infty$ (the agent’s utility can be unboundedly negative, and the agent is to accept all such applicants) on a set of test scores that are realized with positive probability, then the hypothesis of Proposition 2 will not hold. Specifically, $C(t)$ would not have a finite expectation over applicants hired under UBAR.

aligned up to distinguishability. Let $C(t)$ be any affine transformation of the function $\tilde{u}_A^c(t)$. Suppose that the expected value of $C(t)$ of those applicants hired under UBAR exists and is finite, and let K be equal to this expectation: $K = \frac{1}{k} \mathbb{E}_\Omega[\chi^{\text{UBAR}}(t, \tilde{u}_A) \cdot C(t)]$. Then a contract which asks the agent to select any k applicants such that $\mathbb{E}[C(t^i)|i \text{ hired}] = K$ will implement the upper bound acceptance rule. This contract is an optimal contract.

For intuition, consider the following informal Lagrangian argument. Say that the moment function is given by $C(t) = a_0 + a_1 \tilde{u}_A^c(t)$, for $a_1 \neq 0$. Let λ_0 be the multiplier representing the shadow cost of hiring more applicants, and λ_1 the one representing the shadow cost of increasing the average $C(t)$ of those who are hired. At the optimum, the agent hires applicant i if $\tilde{u}_A^i > \lambda_0 + \lambda_1 C(t^i)$. Plugging in the multiplier values $\lambda_0 = -\frac{a_0}{a_1}$ and $\lambda_1 = \frac{1}{a_1}$, the agent hires i if $\tilde{u}_A^i > \tilde{u}_A^c(t)$ – exactly the condition defining UBAR, and thus giving behavior which satisfies the constraints.

If the constraint $\mathbb{E}[C(t^i)|i \text{ hired}] = K$ were relaxed, then on the margin the agent would prefer to hire applicants with higher \tilde{u}_A . When C is a *positive* affine transformation of \tilde{u}_A^c , these preferred applicants are the ones with higher $C(t)$. So the agent interprets the equality as a ceiling: $\mathbb{E}[C(t^i)|i \text{ hired}] = K$ yields the same outcome as $\mathbb{E}[C(t^i)|i \text{ hired}] \leq K$. We can interpret $C(\cdot)$ as a “cost” of hiring an applicant with a given test score, and K as a “budget”. Likewise, when $C(t^i)$ is a *negative* affine transformation, the agent pushes for lower $C(t^i)$. We can then equivalently implement the contract as a floor, with $\mathbb{E}[C(t^i)|i \text{ hired}] \geq K$.

In Sections 3.3, C.1, and D.1 I apply Propositions 1 and 2 to characterize the implementation of the upper bound acceptance rule – the optimal contract under common knowledge of the agent’s type– for the normal specification, the two-factor model, and for a mixture of the two.

3.2 The general problem

In Appendix A, I consider the general problem of solving for the optimal contract – maximizing (6) subject to (4) and (5) – in which alignment up to distinguishability need not hold, and thus UBAR may not be implementable.

The solution involves “ironing.” For instance, at some test score the agent may be required to treat all applicants from the 20th through 30th percentiles of \tilde{u}_A identically. If the agent is to accept 19% of the applicants at this test score, she accepts her favorite 19%. If she is to accept 21%, she accepts her favorite 20% deterministically; and then gives each of the 10% of applicants in the pooling range a a 1/10 chance of acceptance. If she is to accept

more than 30% of applicants, she deterministically accepts or rejects each of them. After pooling together applicants in an appropriate manner, the problem can be transformed into one in which alignment does hold and one can solve it using an approach similar to that in Section 3.1.

The Appendix section gives the general algorithm for deriving the optimal contract. It also shows that, despite the new benefit of randomized acceptances, randomization will only actually be needed at a single test score. So in the case of continuously distributed test scores, in which behavior at any single test score is irrelevant, we still get a deterministic optimal contract (though one that is worse than UBAR).

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 \tilde{q}^i , the agent's expectation on applicant i 's quality given s^i and t^i , as

$$\tilde{q}^i = \frac{\frac{t^i}{\sigma_T^2} + \frac{s^i}{\sigma_S^2}}{\frac{1}{\sigma_Q^2} + \frac{1}{\sigma_T^2} + \frac{1}{\sigma_S^2}} = \frac{\sigma_S^2 \sigma_Q^2 t^i + \sigma_T^2 \sigma_Q^2 s^i}{\sigma_S^2 \sigma_T^2 + \sigma_Q^2 \sigma_S^2 + \sigma_Q^2 \sigma_T^2}. \quad (7)$$

The agent's expected utility for this applicant is $\tilde{u}_A^i = \tilde{q}^i + b^i$. Because I have specified the joint distributions of s , t , and b , I can now find $\hat{u}_P(t, \tilde{u}_A)$, which represents the average quality \tilde{q} of those applicants with test score t and agent utility \tilde{u}_A . Working through the conditional distributions of joint normals (with details for all of the calculations in this subsection given in Appendix F.1),

$$\hat{u}_P(t, \tilde{u}_A) = \beta_T \cdot t + \beta_{U_A} \cdot \tilde{u}_A, \text{ for} \quad (8)$$

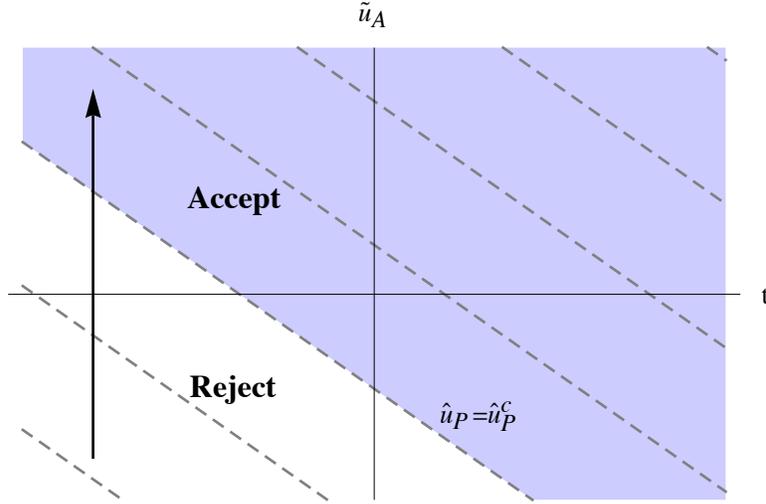
$$\beta_T = \frac{\sigma_B^2 \sigma_Q^2}{(\sigma_Q^2 + \sigma_T^2)(\eta + \sigma_B^2)}, \quad (9)$$

$$\beta_{U_A} = \frac{\eta}{\eta + \sigma_B^2}, \quad (10)$$

$$\text{and } \eta \equiv \frac{\sigma_T^4 \sigma_Q^4}{(\sigma_Q^2 + \sigma_T^2)(\sigma_T^2 \sigma_Q^2 + \sigma_S^2 \sigma_T^2 + \sigma_S^2 \sigma_Q^2)}. \quad (11)$$

(See the discussion after Equation (14) for an interpretation of the term η .) The belief \hat{u}_P is linear in both t and \tilde{u}_A , with respective coefficients $\beta_T > 0$ and $\beta_{U_A} > 0$. We also see that

Figure 2: Upper Bound Acceptance Rule for the Normal Specification



The dashed curves are the linear principal indifference curves, each with slope $-\beta_T/\beta_{U_A}$. The arrow pointing upwards indicates that higher indifference curves represent higher principal utilities, confirming that utilities are aligned up to distinguishability. Under UBAR, applicants are accepted above some cutoff principal utility \hat{u}_P^c chosen so that a mass of k are accepted in total. The acceptance region is shaded.

$\beta_{U_A} < 1$. Under the normalization that the agent and manager 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. This is because an increased agent utility is expected to be partly due to higher quality, and partly due to higher bias. The principal only cares about the first.

More important than its magnitude, the fact that β_{U_A} is positive means that utilities are aligned up to distinguishability. At every test score, average principal utility is increasing in agent utility. So the upper bound acceptance rule will be implementable:

Lemma 1. Under the normal specification with common knowledge of the agent's type, utilities are aligned up to distinguishability. The optimal contract can be implemented as described in Propositions 1 and Proposition 2.

The fact that \hat{u}_P is linear in t and \tilde{u}_A means that the principal indifference curves are linear in (t, \tilde{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 chosen so that k are accepted in total. See Figure 2.

In order to implement UBAR as in Proposition 1, we want to solve for the acceptance rate as a function of test score for a specified \hat{u}_P^c . First observe that t^i and \tilde{u}_A^i follow a joint

normal distribution. The conditional distribution of \tilde{u}_A^i given t^i can be calculated as:

$$\text{Conditional on } t^i = t : \tilde{u}_A^i \sim \mathcal{N}(\mu_{U_A}(t), \sigma_{U_A}^2), \text{ for} \quad (12)$$

$$\mu_{U_A}(t) = \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} t \quad (13)$$

$$\sigma_{U_A}^2 = \eta + \sigma_B^2. \quad (14)$$

Equation (14) now gives an economic interpretation of η , previously defined in (11). It is the variance in the agent's beliefs on quality arising from her private signal. That is, η is the variance of \tilde{q}^i conditional on t^i but unconditional on s^i .

Restating (12),

$$\text{Conditional on } t^i = t : \frac{\tilde{u}_A^i - \mu_{U_A}(t)}{\sigma_{U_A}} \sim \mathcal{N}(0, 1). \quad (15)$$

For any t and any \hat{u}_P^c , we can now calculate the acceptance rate under UBAR. An applicant is accepted under UBAR if

$$\begin{aligned} \beta_T t + \beta_{U_A} \tilde{u}_A &\geq \hat{u}_P^c \\ \Leftrightarrow \frac{\tilde{u}_A - \mu_{U_A}(t)}{\sigma_{U_A}} &\geq \frac{\frac{\hat{u}_P^c - \beta_T t}{\beta_{U_A}} - \mu_{U_A}(t)}{\sigma_{U_A}}. \end{aligned}$$

The LHS of the last line is distributed according to a standard normal. Plugging in $\mu_{U_A}(t)$ from (13) on the RHS and collecting terms, the acceptance condition can be rewritten as

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

for $\gamma_0^* = \frac{\hat{u}_P^c}{\beta_{U_A} \sigma_{U_A}}$ and $\gamma_T^* = \frac{\beta_T}{\beta_{U_A} \sigma_{U_A}} + \frac{\sigma_Q^2}{(\sigma_Q^2 + \sigma_T^2) \sigma_{U_A}}$. I will not calculate γ_0^* explicitly as a function of primitives (\hat{u}_P^c itself being a function of k), but plugging in (9), (10), and (14) to the expression for γ_T^* and simplifying yields

$$\gamma_T^* = \frac{\sigma_Q^2 \sqrt{\eta + \sigma_B^2}}{\eta(\sigma_Q^2 + \sigma_T^2)}. \quad (16)$$

So, indicating the CDF of a standard normal distribution by Φ , the acceptance rate under

UBAR at test score t is $1 - \Phi(\gamma_0^* - \gamma_T^* t) = \Phi(\gamma_T^* t - \gamma_0^*)$:

$$\alpha^{\text{UBAR}}(t) = \Phi(\gamma_T^* t - \gamma_0^*). \quad (17)$$

We now have the implementation of UBAR according to Proposition 1. The acceptance rate at test score t is given by $\alpha^{\text{UBAR}}(t)$ of (17), with $\gamma_T^* > 0$ as specified by (16) and γ_0^* set so that a total of k applicants are accepted across test scores.

We can also express this contract in the alternative manner of Proposition 2: the agent can pick any k applicants subject to $\mathbb{E}[C(t^i)|i \text{ hired}] = K$, for some $C(\cdot)$ and K . The first step is to derive $\tilde{u}_A^c(t)$ as the solution to $\hat{u}_P(t, \tilde{u}_A^c(t)) = \hat{u}_P^c$ for a given \hat{u}_P^c :

$$\begin{aligned} \hat{u}_P(t, \tilde{u}_A^c(t)) &= \hat{u}_P^c \\ \xrightarrow{\text{Eq (8)}} \quad \beta_T t + \beta_{U_A} \tilde{u}_A^c(t) &= \hat{u}_P^c \\ \implies \quad \tilde{u}_A^c(t) &= \frac{\hat{u}_P^c - \beta_T t}{\beta_{U_A}}. \end{aligned}$$

Now we set $C(t)$ to be any affine transformation of $\tilde{u}_A^c(t)$. Because $\tilde{u}_A^c(t)$ is affine in t , we can simply take $C(t) = t$. That gives a very simple implementation of the optimal contract: the agent can choose any k applicants subject to the average test score, $\mathbb{E}[t^i|i \text{ hired}]$, being equal to some level K .²² The agent will select the same set of applicants whether we specify the contract as an acceptance rate function or as an average test score. Fixing k and fixing the empirical distribution of test scores, an acceptance rule $\alpha(t) = \Phi(\gamma_T t - \gamma_0)$ with a higher value of γ_T (and γ_0 adjusted to accept k applicants) corresponds to a higher average test score.

This alternative implementation has the very simple representation of fixing the *average* test score exactly because the principal's indifference curves are *linear* in (t, \tilde{u}_A) -space (Figure 2). A different shape of the indifference curves would lead to a different moment to be fixed.

The following Proposition sums up this discussion.

Proposition 3. Under the normal specification with common knowledge of the agent's type,

²²Because $C(t) = t$ is a negative affine transformation of $\tilde{u}_A^c(t)$, we could also write the contract as a floor: $\mathbb{E}[t^i|i \text{ hired}] \geq K$. The agent would prefer to push the expected test score lower than K , and so would always choose exactly at the cap.

An explicit formula for the principal's optimal choice of K is given by Equation (42) in Appendix E, plugging in for $\sigma_{U_P}^2$ from (20) and solving for τ^* : $K = \frac{\sigma_Q^2 \sqrt{\sigma_Q^2 + \sigma_T^2} \cdot R(k)}{\sqrt{\sigma_Q^2 + \frac{\eta^2}{\eta + \sigma_B^2}}}$, with $R(k)$ as defined in (24).

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:

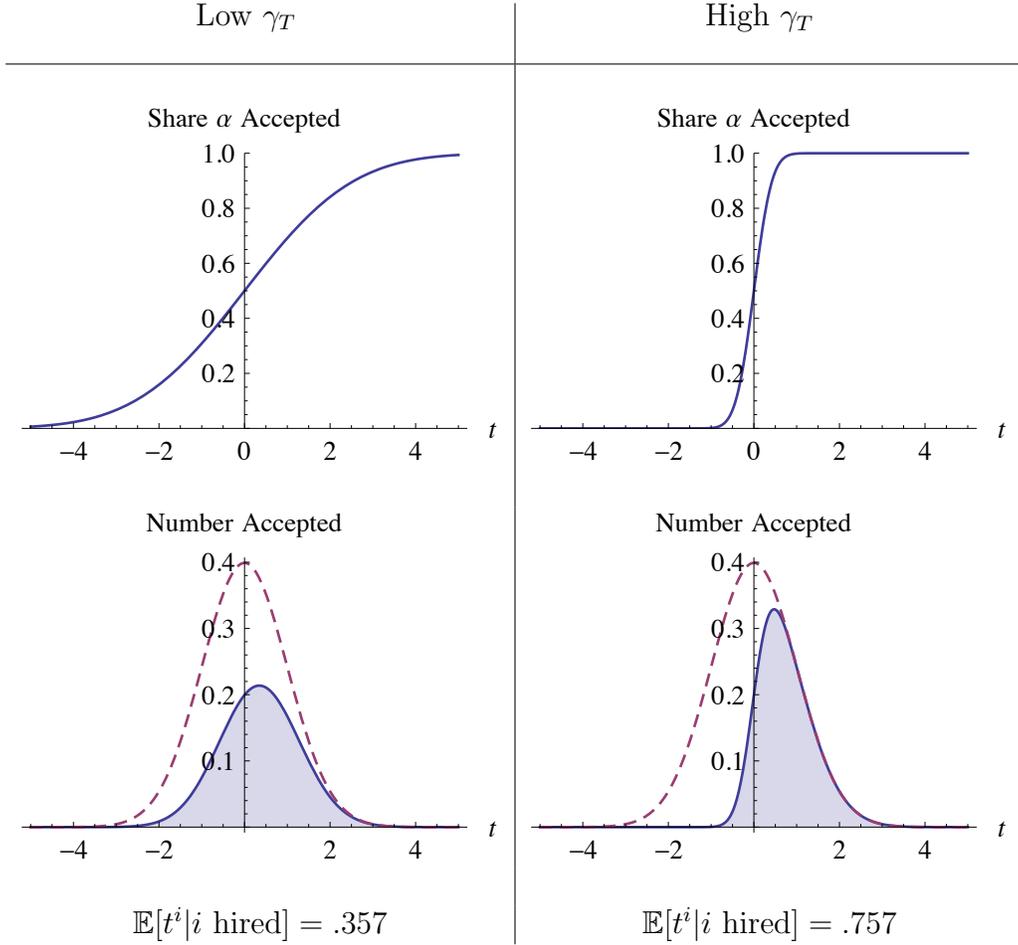
1. An acceptance rate of $\alpha(t) = \Phi(\gamma_T^* t - \gamma_0)$ at test score t , for γ_T^* given by (16) and γ_0 set so that a total of k applicants are accepted; or,
2. An average test score of accepted applicants equal to some value τ^* .

Focus on the first implementation, through a specified acceptance rate at each test score. 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 scores, with the share of applicants accepted approaching 0 as $t \rightarrow -\infty$ and approaching 1 as $t \rightarrow \infty$. Given that γ_0 is to be adjusted to accept k applicants, these contracts are characterized by a one-dimensional sufficient statistic, the steepness γ_T . See Figure 3 to see an illustration of such contracts, and how they vary with the contracting coefficient γ_T . A higher γ_T means that a one unit increase in test scores has a larger effect on acceptance rates. Loosely speaking, a steeper contract – a higher γ_T – can be interpreted as giving the agent less discretion to overrule the public test results. Hiring would depend more on test scores and less on the agent’s input. Taking γ_T to infinity would give us a “no discretion” contract; there would be a test score cutoff below which all applicants were rejected, and above which all were accepted. On the other hand, $\gamma_T \rightarrow 0$ gives a flat contract with a constant acceptance rate at every test score.

In Appendix B I show that the “full discretion” contract, in which the agent selects her favorite applicants after observing their test scores, would also induce a normal CDF acceptance rate. The full discretion steepness γ_T^{FD} satisfies $0 < \gamma_T^{\text{FD}} < \gamma_T^* < \infty$. The agent places some weight on quality, as measured imperfectly by the test score. So with full discretion she still accepts a greater share of applicants at higher test scores. But the agent prefers a flatter acceptance rate than does the principal because she also places weight on idiosyncratic factors that are independent of quality. Equivalently, the agent prefers a lower average test score than does the principal (as discussed in footnote 23).

Let us now explore comparative statics on how the optimal steepness γ_T^* varies with the underlying parameters. One qualification to the interpretation is as follows: as we vary σ_Q^2 or σ_T^2 , the empirical variance $\sigma_Q^2 + \sigma_T^2$ of the test scores changes as well. So the meaning of the coefficient on test scores may change. For instance, if we were to keep the coefficient γ_T on test scores fixed as we increased the variance of the observed 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. To see the impact of a one standard deviation increase in test scores, we could renormalize the contracting coefficient to $\gamma_T^* \sqrt{\sigma_Q^2 + \sigma_T^2}$, which is

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 share of $\Phi(\gamma_T t - \gamma_0)$, for two values of the coefficient γ_T . The second row shows the distribution of test scores for those accepted, with dashed lines indicating the population distribution of test scores. The average test score of the hired applicants is given in the last row; steeper contracts with higher γ_T yield higher average test scores. In this example the empirical distribution of test scores is $\mathcal{N}(0, 1)$; the low value of γ_T is .5 and the high value is 3; and γ_0 is set so that the total number of applicants accepted is $k = .5$. Adjusting γ_0 would translate the α functions in the top row left or right without changing their shape.

the coefficient on the z -score of test results.²³ In the following Proposition I also look at comparative statics on the renormalized coefficient, when relevant.

Proposition 4. In the contract of Proposition 3 part 1, the contracting parameter γ_T^* given by (16) has the following comparative statics and limits:

1. γ_T^* is independent of k .
2. $\frac{d\gamma_T^*}{d\sigma_S^2} > 0$; $\lim_{\sigma_S^2 \rightarrow 0} \gamma_T^* \in (0, \infty)$; and $\lim_{\sigma_S^2 \rightarrow \infty} \gamma_T^* = \infty$.
3. $\frac{d\gamma_T^*}{d\sigma_B^2} > 0$; $\lim_{\sigma_B^2 \rightarrow 0} \gamma_T^* \in (0, \infty)$; and $\lim_{\sigma_B^2 \rightarrow \infty} \gamma_T^* = \infty$.
4. $\frac{d\gamma_T^*}{d\sigma_T^2} < 0$ and $\frac{d}{d\sigma_T^2} \left(\gamma_T^* \sqrt{\sigma_Q^2 + \sigma_T^2} \right) < 0$; $\lim_{\sigma_T^2 \rightarrow 0} \gamma_T^* = \lim_{\sigma_T^2 \rightarrow 0} \gamma_T^* \sqrt{\sigma_Q^2 + \sigma_T^2} = \infty$; and $\lim_{\sigma_T^2 \rightarrow \infty} \gamma_T^* = \lim_{\sigma_T^2 \rightarrow \infty} \gamma_T^* \cdot \sqrt{\sigma_Q^2 + \sigma_T^2} = 0$.
5. $\frac{d\gamma_T^*}{d\sigma_Q^2}$ and $\frac{d}{d\sigma_Q^2} \left(\gamma_T^* \sqrt{\sigma_Q^2 + \sigma_T^2} \right)$ can have either sign; $\lim_{\sigma_Q^2 \rightarrow 0} \gamma_T^* = \lim_{\sigma_Q^2 \rightarrow 0} \gamma_T^* \sqrt{\sigma_Q^2 + \sigma_T^2} = \infty$; and $\lim_{\sigma_Q^2 \rightarrow \infty} \gamma_T^* \in (0, \infty)$ while $\lim_{\sigma_Q^2 \rightarrow \infty} \gamma_T^* \sqrt{\sigma_Q^2 + \sigma_T^2} = \infty$.

All of these results can be calculated straightforwardly from the formula for γ_T^* in (16). Explicit formulas for interior limits are given in the proof.

Part 1 reiterates that the coefficient on test scores does not depend on the number of people to be hired. Changing k just affects γ_0 , shifting the acceptance rate function left or right.

Part 2 states that as the agent becomes better informed, the contract gets flatter – less weight is placed on the test scores. With a fully uninformed agent ($\sigma_S^2 \rightarrow \infty$), the contract relies only on the test scores ($\gamma_T^* \rightarrow \infty$) and not at all on the agent’s input. Part 3 shows that as the agent becomes more biased – more variance in her preferences comes from idiosyncratic sources – the contract gets steeper. In the limit where her preferences are entirely idiosyncratic the contract again depends only on test scores.

One economic question arising from the analysis of pre-employment tests in Hoffman, Kahn, and Li (2015) is how much “discretion” a hiring manager should be given. They look at the limit contracts and show 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 2 and 3 point to similar tradeoffs in the optimal contracts. We use a flatter contract – one in which the agent has “more discretion” to pick low test-score applicants, whom she prefers on the margin – when bias is low or information is high.

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.

Part 4 finds that as σ_T^2 goes up and the public test becomes less informative, we reduce the weight on both the absolute and the z-score of test results. As test scores become perfectly informative, we hire entirely based on test results. As tests become perfectly uninformative, we use a constant acceptance rate across all scores.

Part 5 shows that γ_T^* can vary nonmonotonically with σ_Q^2 , the variance of quality in the population.²⁴ For completeness, I include the characterization of the limiting contracting parameter as the quality variance goes to 0 or infinity.

In Appendix B I replicate the comparative statics analysis for the full discretion contract, rather than the optimal one. 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. The acceptance rate function in the optimal and the full discretion contract both become flatter when the agent has better private information (lower σ_S^2). 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 while the optimal contract becomes steeper when the agent is more biased (higher σ_B^2), the full discretion contract becomes flatter; the agent wants to accept more low quality applicants who tested poorly, and the principal wants to prevent her from accepting them.

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 will need to make parametric and distributional assumptions. Accordingly, this section focuses only on the normal specification. Here, the agent's type θ describes the two dimensions of how informed she is, and how strong is her idiosyncratic bias: $\theta \equiv (\sigma_S^2, \sigma_B^2)$. Let G be the principal's prior distribution over the agent's type θ . In Appendix D.2, I extend the results of this section to a mixture of the normal specification with the two-factor model. In that case, agents may have heterogeneity on the three dimensions of their types: information, idiosyncratic bias, and systematic bias.

I will ultimately solve for the optimal contract by transforming this problem into one that is isomorphic to a one-dimensional delegation problem, and then applying results from that literature. In a one-dimensional problem, the agent observes a one-dimensional state, which determines principal and agent preferences over a one-dimensional action. The principal then

²⁴Numerically, the coefficient γ_T^* appears to either be decreasing, or to be decreasing and then increasing with σ_Q^2 ; the coefficient on the z-score of test results appears to be decreasing and then increasing.

gives the agent a delegation set over the actions she may choose. A number of papers (see the Introduction for a short list) find functional form and distributional assumptions which imply that an interval delegation set is optimal – for instance, a cap or a floor on actions. Amador and Bagwell (2013) also consider the possibility of mutual “money-burning”: an auxiliary action that uniformly reduces the payoffs of both players. That paper finds conditions under which the agent is still given an interval delegation set, and in which money-burning will not be used even if it is available.

In the current paper, the problem is posed a much higher dimensional one. The action effectively corresponds to the entire function mapping test scores to acceptance rates. The players’ preferences over this action are determined by the two-dimensional type of the agent. However, I will find that 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 both players the payoffs of a normal CDF acceptance rate, minus some “money burning” in which both players are made worse off. I also show that, while the agent’s type is two-dimensional, behavior in any contract is determined by a one-dimensional projection of this type. Hence, I can reduce my problem into a one-dimensional delegation problem. Conditions in Amador and Bagwell (2013) yielding a floor on actions and no money-burning translate into a contract in my setting that specifies a minimum steepness of a normal CDF acceptance rate. This contract can also be implemented as a contract which allows the agent to select any applicants subject to a floor on their average test score.

4.1 Rewriting payoffs

Under the normal specification, Equations (12) - (14) describe the distribution of agent utilities \tilde{u}_A^i given $t^i = t$ as $\tilde{u}_A^i \sim \mathcal{N}(\mu_{U_A}(t), \sigma_{U_A}^2)$. The mean $\mu_{U_A}(t)$ is affine in the test score t , but does not depend on the agent’s type θ . The variance $\sigma_{U_A}^2$ is constant in t , but does depend on the type θ ; throughout this section, I write it as $\sigma_{U_A}^2(\theta)$ to emphasize this dependence, and similarly for other terms. The principal’s utility \hat{u}_P^i is $\hat{u}_P(t^i, \tilde{u}_A^i) = \beta_T(\theta)t^i + \beta_{U_A}(\theta)\tilde{u}_A^i$ (Equations (8) - (10)). We can calculate the distribution of \hat{u}_P^i given $t^i = t$ but not \tilde{u}_A^i as

$$\text{Conditional on } t^i = t : \hat{u}_P^i \sim \mathcal{N}(\mu_{U_P}(t), \sigma_{U_P}^2(\theta)), \text{ for} \quad (18)$$

$$\mu_{U_P}(t) = \mu_{U_A}(t) = \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2}t \quad (19)$$

$$\sigma_{U_P}^2(\theta) = \beta_{U_A}^2(\theta) \cdot \sigma_{U_A}^2(\theta) = \frac{\eta^2}{\eta + \sigma_B^2}. \quad (20)$$

The means are the same because the agent's bias is uncorrelated with the test score; the agent's and principal's average utility across applicants at a test score is equal to the expected quality at that score. 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 $\beta_{U_A}^2(\theta)$. Recall that $0 < \beta_{U_A}(\theta) < 1$, implying that $0 < \sigma_{U_P}(\theta) < \sigma_{U_A}(\theta)$.

If i has test score t^i and induces agent utility \tilde{u}_A^i , we can say that applicant i has an agent utility z-score, conditional on the test result, of z^i :

$$z^i \equiv \frac{\tilde{u}_A^i - \mu_{U_A}(t^i)}{\sigma_{U_A}(\theta)}. \quad (21)$$

The z^i term captures how good an applicant appears to the agent, conditional on the public information. Since \hat{u}_P is increasing in \tilde{u}_A , this term also describes the applicant's relative quality for the principal (up to distinguishability). An applicant with a high z-score and low test score looks very good to the agent relative to other applicants with the same test score, but might be lower utility to her than an average applicant (z-score of 0) with a high test score.

With this new definition, we can rewrite the utilities that the principal and agent get from hiring applicant i :

$$\begin{aligned} \tilde{u}_A^i &= \mu_{U_A}(t^i) + \sigma_{U_A}(\theta) \cdot z^i = \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} \cdot t^i + \sigma_{U_A}(\theta) \cdot z^i \\ \hat{u}_P^i &= \mu_{U_P}(t^i) + \sigma_{U_P}(\theta) \cdot z^i = \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} \cdot t^i + \sigma_{U_P}(\theta) \cdot z^i. \end{aligned}$$

For a given set of k hired applicants, let $\zeta \equiv \mathbb{E}[z^i | i \text{ hired}]$ be the average agent utility z-score and let $\tau \equiv \mathbb{E}[t^i | i \text{ hired}]$ be the average test score. We now see that the payoffs of the principal and agent from a set of hired applicants is

$$V_A(\tau, \zeta; \theta) = \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} \cdot \tau + \sigma_{U_A}(\theta) \cdot \zeta \quad (22)$$

$$V_P(\tau, \zeta; \theta) = \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} \cdot \tau + \sigma_{U_P}(\theta) \cdot \zeta. \quad (23)$$

The principal and agent payoffs have been reduced to a linear function of only two moments of the set: the average test score τ , and the average "z-score" ζ of agent utility.

Recall that the agent's type is $\theta = (\sigma_S^2, \sigma_B^2)$. The agent's preferences over (τ, ζ) depend

on the induced value of $\sigma_{U_A}(\theta)$ and the principal's depend on $\sigma_{U_P}(\theta)$. Part 1 of Lemma 2 confirms that the effect of increasing the agent's private information (lower σ_S^2) is to increase both σ_{U_A} and σ_{U_P} . 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 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 range of possible pairs of σ_{U_A} and σ_{U_P} .

Lemma 2.

1. Fixing σ_B^2 , it holds that $\frac{d\sigma_{U_A}}{d\sigma_S^2}(\sigma_S^2, \sigma_B^2) < 0$ and $\frac{d\sigma_{U_P}}{d\sigma_S^2}(\sigma_S^2, \sigma_B^2) < 0$.
2. Fixing σ_S^2 , it holds that $\frac{d\sigma_{U_A}}{d\sigma_B^2}(\sigma_S^2, \sigma_B^2) > 0$ and $\frac{d\sigma_{U_P}}{d\sigma_B^2}(\sigma_S^2, \sigma_B^2) < 0$.
3. Across $\theta = (\sigma_S^2, \sigma_B^2) \in \mathbb{R}_{++}^2$, the image of $\sigma_{U_A}(\theta)$ is \mathbb{R}_{++} . Given $\sigma_{U_A}(\theta) = \tilde{\sigma}_{U_A}$, the image of $\sigma_{U_P}(\theta)$ over θ satisfying $\sigma_{U_A}(\theta) = \tilde{\sigma}_{U_A}$ is the interval $\left(0, \min\left\{\tilde{\sigma}_{U_A}, \frac{\sigma_T^2 \sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} \frac{1}{\tilde{\sigma}_{U_A}}\right\}\right)$.

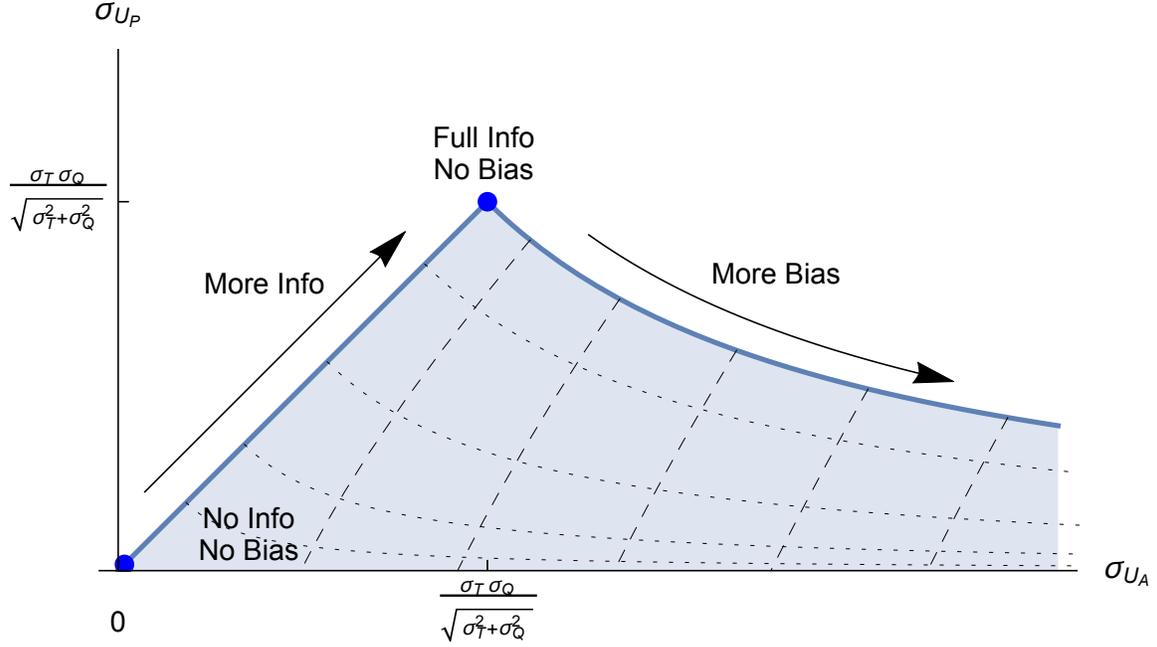
Interpreting Part 3, with no information ($\sigma_S^2 \rightarrow \infty$) and no bias ($\sigma_B^2 \rightarrow 0$), we have $\sigma_{U_P} = \sigma_{U_A} = 0$. Improving information moves us up the $y = x$ line in $(\sigma_{U_A}, \sigma_{U_P})$ -space. The maximum possible value of σ_{U_P} is achieved when the agent has no bias ($\sigma_B^2 \rightarrow 0$) and perfect information ($\sigma_S^2 \rightarrow 0$), with $\sigma_{U_P} = \sigma_{U_A} = \frac{\sigma_T \sigma_Q}{\sqrt{\sigma_Q^2 + \sigma_T^2}}$. The value of σ_{U_A} can then be increased without bound by increasing the bias, but increasing σ_B^2 lowers σ_{U_P} . So the maximum possible σ_{U_P} given $\sigma_{U_A} = \tilde{\sigma}_{U_A}$ first increases linearly with $\tilde{\sigma}_{U_A}$, then after a cutoff falls at a rate of $\frac{1}{\tilde{\sigma}_{U_A}}$. See Figure 4.

4.2 Rewriting the contracting space

Across all possible messages for a given contract, the agent will choose the message that maximizes a weighted sum of average test scores τ and utility z-scores ζ . Players' preferences over (τ, ζ) depend on the agent's type θ , but the possibility set given a contract does not depend on type. In other words, any contract reduces to a set of possible (τ, ζ) from which the agent may choose. The test scores and their averages, of course, are directly observable and contractible. The possible z-scores can be inferred from the rules of the contract.²⁵ What pairs of average test scores τ and average z-scores ζ are possible, across all contracts?

²⁵Suppose that we have a deterministic contract in which the agent sends a message inducing acceptance rate function α . Let us calculate the average z-score, ζ^α . (One could perform a similar exercise for messages in stochastic contracts.) The principal infers that the agent has selected her favorite applicants at each test

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.

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

$$R(x) = \frac{\phi(\Phi^{-1}(1-x))}{x}. \quad (24)$$

score. Because the agent's utility \tilde{u}_A is normally distributed conditional on any test score, at score t it must be that the mean z-score is $R(\alpha(t))$, for R given below in (24). We can then take a weighted average across mean z-scores at each t to find the average z-score of all accepted applicants. In particular, test scores are distributed normally with mean 0 and variance $\sigma_Q^2 + \sigma_T^2$ in the population. Given a deterministic contract with acceptance rate function α , the average z-score ζ^α across all accepted applicants would be given by:

$$\zeta^\alpha = \frac{1}{k} \int_{t=-\infty}^{\infty} \frac{\phi\left(\frac{t}{\sqrt{\sigma_Q^2 + \sigma_T^2}}\right)}{\sqrt{\sigma_Q^2 + \sigma_T^2}} \alpha(t) R(\alpha(t)) dt.$$

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

Now, the highest possible average test score for any set of k applicants comes from a contract which simply selects the k applicants with the highest test scores. Since the empirical distribution of test scores is normal with mean 0 and variance $\sigma_Q^2 + \sigma_T^2$, the top k share of test scores have a mean of R_T for

$$R_T \equiv \sqrt{\sigma_Q^2 + \sigma_T^2} \cdot R(k). \quad (25)$$

So the highest possible average test score for any set of k applicants is R_T , and the lowest possible average score – if we select the applicants with the lowest test scores – is $-R_T$. In either case, selecting entirely on test scores means that the average z-score ζ would be 0.

The highest possible average z-score ζ would arise from selecting the k applicants with the highest z-scores z^i . Recalling that the distribution of agent utility z-scores at every test score is the same, a standard normal, we would achieve that maximum with a deterministic contract that fixes a constant acceptance rate $\alpha(t) = k$ across all t . This gives an upper bound of R_Z on the average z-score ζ , with

$$R_Z \equiv R(k). \quad (26)$$

At the same time, the contract with a constant acceptance rate gives an average test score τ of 0. Note that if the agent were to (suboptimally) play the same a contract by selecting her least favorite applicants, she would would get ζ of $-R_Z$.

In fact, the set of possible pairs of (τ, ζ) across all sets of k applicants is the interior of an ellipse centered at $(0, 0)$, with principal axes of R_T and R_Z . The elliptical shape of the frontier comes from the elliptical distribution of the joint normal over t^i and z^i in the population. See Figure 5.

Lemma 3. Let $\overline{W} \subseteq \mathbb{R}^2$ be the set of (τ', ζ') such that

$$\frac{(\tau')^2}{R_T^2} + \frac{(\zeta')^2}{R_Z^2} \leq 1.$$

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

A contract then induces a subset of this ellipse. The contract can specify any subset of possible τ in $[-R_T, R_T]$, since test scores are directly observable. Then, for each allowed

τ , the contract implies some range of possible ζ that the agent can achieve. Note that any contract that allows for an average z-score of ζ at τ must also allow for $-\zeta$, and for anything in between – the agent can always “permute” her report to select applicants with lower rather than higher values of the nonverifiable z^i .

Lemma 4.

1. Under any contract, the set of achievable average test scores and z-scores (τ, ζ) across messages is independent of agent type.
2. Take some contract with a set W of achievable (τ, ζ) . If $(\tau', \zeta') \in W$, then for any $x \in [-1, 1]$ it also holds that $(\tau', x \cdot \zeta') \in W$.
3. Take $W \subseteq \mathbb{R}^2$. There exists a contract such with set of achievable (τ, ζ) equal to W if and only if (i) $W \subseteq \overline{W}$ (defined in Lemma 3) and (ii) W is consistent with Part 2 above.

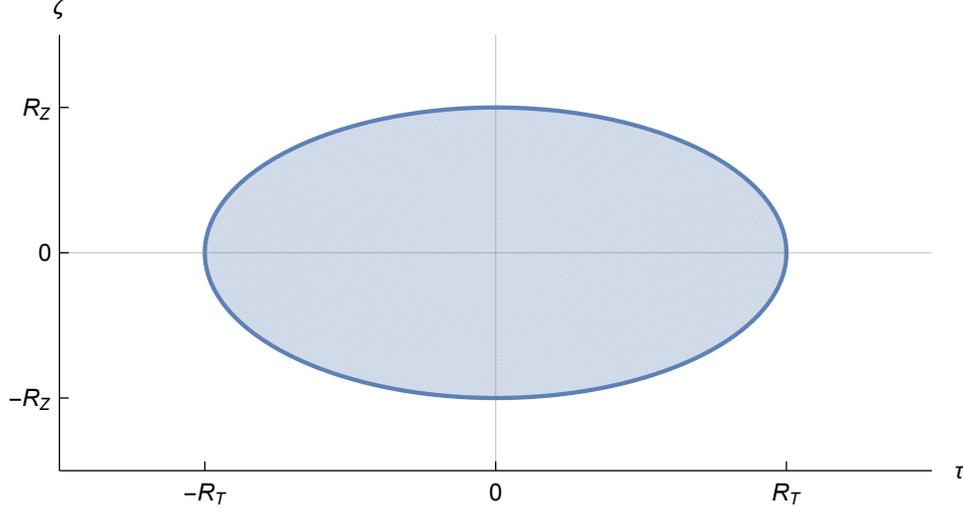
To reiterate, in this contracting notation, a contract specifies a set of achievable (τ, ζ) given by $W \subseteq \overline{W}$, and the agent selects (τ', ζ') to maximize (22) over W . The principal’s payoff under (τ', ζ') is given by (23). Only the upper-right frontier of W in (τ, ζ) -space is relevant, since an agent of any type places a positive weight on both terms. She always picks the highest possible τ for any given ζ , and the highest possible ζ for any given τ (a restatement of monotonicity, from Observation 1, part 2). Moreover, the highest possible ζ given τ must be weakly positive (Lemma 4, part 2).

As we have previously discussed, if a contract asks an agent to select any set of applicants subject to a given expected test score, that contract induces a normal CDF acceptance rate. But if the only restriction on the agent’s choices is the fixed τ , then the agent must be choosing a ζ on the upper frontier of the possibility set \overline{W} , i.e., $\zeta = R_Z \sqrt{1 - \frac{\tau^2}{R_T^2}}$. In other words, the (τ, ζ) values on the upper boundary of the ellipse in Figure 5 are those induced by deterministic contracts with acceptance rates of the normal CDF functional 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 of $\gamma_T = 0$ across test scores; and $\tau < 0$ is achieved by $\gamma_T < 0$.

4.3 Projecting the type space to one dimension

The principal and agent preferences over the average test score τ and the average z-score ζ depend on the agent’s two-dimensional type, $\theta = (\sigma_S^2, \sigma_B^2)$, consisting of bias and information.

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



The shaded ellipse shows the feasible values of $\tau = \mathbb{E}[t^i | i \text{ hired}]$ and $\zeta = \mathbb{E}[z^i | i \text{ hired}]$ across sets of k applicants (Lemma 3). From Lemma 4, any contract is equivalent to a subset W of the shaded region satisfying the following condition: if (τ', ζ') is in W , then so too must be $(\tau', x \cdot \zeta')$ for any $x \in [-1, 1]$.

Given such a contract, the agent sends a message that selects her favorite (τ, ζ) in W . The principal's and agent's payoffs are both increasing as we move up-and-to-the-right, with downward sloping linear indifference curves whose slopes depend on the agent's type; see (22) and (23). The agent's indifference curves have slope $-\frac{1}{\sigma_{U_A}(\theta)} \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2}$, while the principal's have slope $-\frac{1}{\sigma_{U_P}(\theta)} \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2}$. The agent has flatter indifference curves, because $\sigma_{U_P}(\theta) < \sigma_{U_A}(\theta)$.

But from (22) and (23), we see that the principal and agent preferences over (τ, ζ) depend only on θ through the induced 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 σ_{U_J} , for $J = A, P$, leads to a higher weight on ζ relative to τ – flatter indifference curves. 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. But because $\sigma_{U_P}(\theta) < \sigma_{U_A}(\theta)$ for every θ , the agent necessarily has flatter indifference curves than the principal; her ideal point from \overline{W} has higher ζ and – as we have already seen – a lower average test score τ .

Because the agent's preferences depend on θ only through $\sigma_{U_A}(\theta)$, the principal can never separate any two agent types with the same $\sigma_{U_A}(\theta)$. They act identically, in the sense of having the same preferences over (τ, ζ) given any contract. Even though the principal has different preferences over (τ, ζ) across these types (different $\sigma_{U_P}(\theta)$), the types are indistinguishable.

So for each possible realization $\tilde{\sigma}_{U_A}$ of $\sigma_{U_A}(\theta)$, the principal need only look at his average

payoffs over (τ, ζ) across types θ such that $\sigma_{U_A}(\theta) = \tilde{\sigma}_{U_A}$. Indeed, because the principal's payoffs are linear in $\sigma_{U_P}(\theta)$ (Equation (23)), the principal's preferences over types with $\sigma_{U_A}(\theta) = \tilde{\sigma}_{U_A}$ are determined just by the average of $\sigma_{U_P}(\theta)$ over these types. Denote this average by $\hat{\sigma}_{U_P}$:

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

This average must of course be in the range of the possible values described by Lemma 2, part 3. That is, $\hat{\sigma}_{U_P}$ is a function taking values in the shaded region of Figure 4 (with $\tilde{\sigma}_{U_A}$ rather than σ_{U_A} on the x axis).

We now effectively have a one-dimensional type where the payoff-relevant uncertainty for the principal – the state of the world – is captured by $\tilde{\sigma}_{U_A}$, the realization of $\sigma_{U_A}(\theta)$. Let H denote the distribution of $\tilde{\sigma}_{U_A}$ over \mathbb{R}_{++} induced by G . The two-dimensional distribution G determines both the one-dimensional distribution H and the function $\hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A})$.

Payoffs can now be written as

$$V_A(\tau, \zeta; \tilde{\sigma}_{U_A}) = \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} \cdot \tau + \tilde{\sigma}_{U_A} \cdot \zeta \quad (27)$$

$$V_P(\tau, \zeta; \tilde{\sigma}_{U_A}) = \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} \cdot \tau + \hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A}) \cdot \zeta \quad (28)$$

Given a contract inducing a set W of possible (τ, ζ) values, the agent observes $\tilde{\sigma}_{U_A} = \sigma_{U_A}(\theta)$ and chooses $(\tau', \zeta') \in W'$ to maximize V_A . The principal chooses the set W (subject to the conditions of Lemma 4, part 3) to maximize $\mathbb{E}_{\tilde{\sigma}_{U_A} \sim H}[V_P]$, taking into account predictions of the agent's behavior at each type.

4.4 Projecting the action space to one dimension plus money-burning

No equilibrium message in a contract will ever induce (τ, ζ) with $\zeta < 0$, so we can focus on the top half of the ellipse of Figure 5. 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, from both the principal and agent perspectives. We could find another pool of applicants in \bar{W} that strictly increased τ and/or ζ , improving 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

$$\bar{\tau}(\zeta) \equiv R_T \cdot \sqrt{1 - \frac{\zeta^2}{R_Z^2}} = \sqrt{(\sigma_Q^2 + \sigma_T^2) \cdot (R(k)^2 - \zeta^2)}. \quad (29)$$

From Lemma 3, the maximum possible τ at a given ζ is $\bar{\tau}(\zeta)$, and the minimum is $-\bar{\tau}(\zeta)$.

Now take any (τ, ζ) in the upper half of the ellipse. Principal and agent payoffs are equal to the payoff from an applicant pool with (τ, ζ) projected to the right edge of the ellipse (that is, replacing τ with $\bar{\tau}(\zeta)$); minus an amount proportional to the projected distance $\bar{\tau}(\zeta) - \tau$. Specifically, from (22) and (23), the agent and principal ($J = A, P$) get payoffs $V_J = \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} \cdot \tau + \sigma_{U_J}(\theta) \cdot \zeta$. This payoff can be rewritten as

$$\begin{aligned} V_J(\tau, \zeta; \theta) &= \left[\frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} \cdot \bar{\tau}(\zeta) + \sigma_{U_J}(\theta) \cdot \zeta \right] - \left[\frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} \cdot (\bar{\tau}(\zeta) - \tau) \right] \\ &= \left[\frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} \cdot \bar{\tau}(\zeta) + \sigma_{U_J}(\theta) \cdot \zeta \right] - \delta(\tau, \zeta), \end{aligned} \quad (30)$$

for $\delta(\tau, \zeta) \equiv \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} \cdot (\bar{\tau}(\zeta) - \tau)$. As with the transformation from (22) and (23) to (27) and (28), we can always replace $\sigma_{U_A}(\theta)$ with $\tilde{\sigma}_{U_A}$ and $\sigma_{U_P}(\theta)$ with $\hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A})$ in (30).

The first bracketed terms give the payoff from an applicant pool with the same ζ , but projected to the upper-right frontier of the ellipse. We then subtract the “money burning cost” in the second brackets, rewritten as δ , which is positive when $\tau < \bar{\tau}(\zeta)$.

Observe that 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 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 agent sends a message in a contract that determines $\zeta \in [0, R_Z]$ as well as the level of money-burning $\delta \geq 0$.²⁶

4.5 Optimal contracts

Treat ζ as the one-dimensional “action” to be taken, and allow for the possibility of money-burning $\delta \geq 0$. Payoffs over actions are determined by a one-dimensional “state” variable,

²⁶Given 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, and so the upper limit of δ will not be relevant.

$\tilde{\sigma}_{U_A}$, which is determined by the agent's type. The agent is biased towards higher actions than the principal – her ideal ζ is larger for every realization of $\tilde{\sigma}_{U_A}$, because $\hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A}) < \tilde{\sigma}_{U_A}$. This is the setting in one-dimensional delegation under which – under appropriate regularity conditions – *action ceilings* are often found to be optimal. The agent pushes towards higher actions, so the principal restricts her from going too high. Specifically, Amador and Bagwell (2013) give conditions for a ceiling on actions, and no money-burning, to be optimal in a one-dimensional delegation problem – even if it is possible for the principal to enforce unrestricted money-burning $\delta \geq 0$ as a function of the action ζ . Their results can be translated into conditions on H and $\hat{\sigma}_{U_P}(\cdot)$ which guarantee that an optimal contract can be expressed as a choice over any ζ less than or equal to a ceiling, and with $\tau = \bar{\tau}(\zeta)$.

Translating such a contract into a more meaningful (and more implementable) restriction, a *ceiling* on ζ is exactly equivalent to a *floor* on the average test score τ of accepted applicants. In either case, the agent picks (τ, ζ) from an interval of 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. Let the distribution H have bounded support with pdf h . 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, then the optimal contract is deterministic and can be characterized in either of the following two ways:

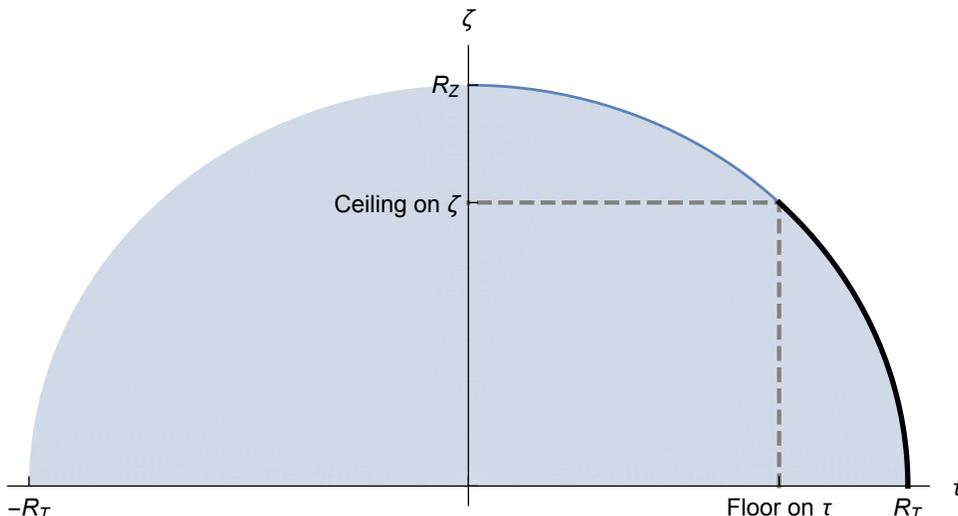
1. The agent is given a floor on the average test score. She may select any k applicants she wants, subject to the average test score of hired applicants τ being at or above some specified level $K > 0$.
2. The agent is given a floor on the steepness of the acceptance rate function. She may select any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

Proposition 9 in Appendix D.2 derives these same contract forms as optimal in a model which combines the idiosyncratic biases of the normal specification with the systematic biases of the two-factor model.

As in Amador and Bagwell (2013) and other papers on one-dimensional delegation, the floor is set to a level that is correct, on average, for the agents who are bound by the floor.²⁷

²⁷By itself, this condition on how the floor is set only implies that there is no benefit from making the

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



Here I highlight on the upper-half of the ellipse \bar{W} from Figure 5. Both the principal and agent value τ and ζ positively, so any outcome not on the upper-right edge of this ellipse would be dominated for both players: there would be another set of applicants with the same ζ and higher τ . From a payoff perspective, it would be as if a set of applicants with the same ζ but with $\tau = \bar{\tau}(\zeta)$ on the frontier had been chosen, and then both players had “burned money.”

The contract in Proposition 5 is first derived as a ceiling on ζ . The agent can pick any set of applicants subject to the inferred average z-score ζ being below some level; hence, she only considers (τ, ζ) points on the thick black curve. In terms of more directly observed statistics, the contract is equivalent to (and is described in Proposition 5 part 1 as) a floor on the average test score τ .

It is possible that the floor always binds – that all agent types choose the same average test score of K . Or the principal might “screen” agent types by setting a low enough floor that agents sometimes exceed it. Indeed, any agent with $\sigma_{U_A}(\theta) \simeq 0$ – one who is approximately uninformed and unbiased – would select applicants purely by the test score, i.e., an acceptance rate function with $\gamma_T \simeq \infty$. If the support of H extends to $\tilde{\sigma}_{U_A} = 0$, then the contract of Proposition 5 would always be nonbinding for those agents with $\sigma_{U_A}(\theta)$ close enough to 0.

acceptance rate steeper or shallower than $\Phi(\Gamma t - \gamma_0)$ 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 scores 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 (one 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 \bar{W} (see Lemmas 3 and 4, as well as footnote 26) – the same pair (τ, ζ) for each of the agents. Any (τ, ζ) is weakly dominated for the principal (independently of agent type) 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. So, averaged over this set of agents, the normal CDF acceptance rate $\Phi(\Gamma t - \gamma_0)$ is preferred to any other specified acceptance rate α .

There are two reasons why the agent would prefer high test scores: because she has worse information (higher σ_S^2), or because her bias σ_B^2 is lower. A less informed agent wants to follow the test more closely, selecting applicants with higher average scores, because she is bad at picking out high quality applicants with low test scores. And the principal agrees that a less informed agent should follow the test more closely. The two players are aligned on this dimension. But the principal is misaligned with the agent on the impact of bias. A less biased agent wants to pick low quality applicants less often, and so tends to pick applicants with higher test scores. The principal, on the other hand, wants a less biased agent to pick applicants with lower average test scores – he trusts her judgment more and wants to let her to overrule the test more often.

Loosely speaking, then, the reason why the principal might give the agent flexibility over the average test score is that heterogeneous agent preferences are driven by uncertainty over information. If agents were only heterogeneous in their biases, then any time an agent wanted higher test scores, the principal would want lower test scores. Let us consider now what happens there is uncertainty over only one of bias or information, but not both.

Commonly known information, uncertain bias. When the agent’s information level σ_S^2 is commonly known, the principal simply fixes the normal CDF acceptance rate, or the average test score, in advance. Additional distributional restrictions such as those in Proposition 5 are not needed. Under these contracts, the agent still has “discretion” or “flexibility” to select her favorite applicants within each test score. But she has no flexibility over the number of applicants selected at each score, or over the average test score.

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 given a specified average test score. She may select any k applicants she wants, subject to the average test score of hired applicants τ being equal to some specified level $K > 0$.
2. The agent is allowed to choose any k applicants she wants as long as the induced acceptance rate $\alpha(t)$ is $\Phi(\gamma_T t - \gamma_0)$, with γ_T equal to some specified level $\Gamma > 0$.

In fact, I show in Appendix F that Proposition 6 follows from a slightly stronger result. Proposition 10 in Appendix F confirms that there is never a benefit of flexibility over τ if the principal’s minimum ideal τ over all possible types is above the agent’s maximum ideal τ – that is, if the maximum $\hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A})$ in the support is below the minimum $\tilde{\sigma}_{U_A}$. 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 implies that $\hat{\sigma}_{U_P}$ is decreasing (as seen in Lemma 2, part 2 and in Figure 4), but it is not a necessary condition.

Commonly known bias, uncertain information. Now let’s look at the reverse case, with commonly known bias but uncertain information. Here there is a potential benefit from flexibility. I do still require the distributional assumptions of Proposition 5 in order to derive the optimal mechanism, a possibly binding floor on steepness or the average test score. But I can give some simpler sufficient conditions for these assumptions to hold.

Lemma 5. Let σ_B^2 be commonly known, and let the distribution H have bounded support with pdf h . If $h(\tilde{\sigma}_{U_A})$ is increasing in $\tilde{\sigma}_{U_A}$ over the support, then the hypothesis of Proposition 5 is satisfied: $H(\tilde{\sigma}_{U_A}) + (\tilde{\sigma}_{U_A} - \hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A}))h(\tilde{\sigma}_{U_A})$ is nondecreasing.

I show this by proving that, while $(\tilde{\sigma}_{U_A} - \hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A}))$ is decreasing, it is decreasing sufficiently slowly so that the expression $H(\tilde{\sigma}_{U_A}) + (\tilde{\sigma}_{U_A} - \hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A}))h(\tilde{\sigma}_{U_A})$ remains nondecreasing. Indeed, the hypothesis of Proposition 5 is satisfied when (i) h increasing, and (ii) the slope of $\hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A})$ is less than or equal to 2; see Lemma 9 in Appendix F. When σ_B^2 is commonly known, the slope of $\hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A})$ is always in $[1, 2]$ – these are the dashed lines in Figure 4.

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 studies how subjective opinions should be incorporated into an algorithm when the agent may be biased. That is, I take the machine learning or statistics problem of optimal prediction from a variety of information sources as a solved problem – I assume that players simply calculate expectations as Bayesians with fully specified models of the environment. Instead, I focus on a strategic issue. If an agent is biased, then any

mechanism that allows for her information to influence outcomes must be allowing her biases to do so as well. The soft information that one recovers can depend on the mechanism that will be used to make decisions.

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

5.1 Statistical Discrimination and Commitment

Work on “statistical” versus “taste-based” discrimination suggests a test for bias – taste-based discrimination – when an agent makes a number of binary decisions. To be concrete, think of the setting of Knowles et al. (2001) in which police officers pull over drivers and search for contraband. For each driver who is pulled over, there is “hard information” corresponding to the driver’s race; and there is an ex post “quality” realization corresponding to the amount of contraband discovered. The agent is demonstrated to be biased, as opposed to just statistically discriminating, if the quality of the marginal driver – the one she was just indifferent about pulling over or not – varies across races. The current paper is motivated by the acceptance and rejection of job applicants with different observable “test scores”. A corresponding test for bias would allow the agent to accept any applicants she wanted and then measure the quality of the marginal accepted applicant at each test score. An unbiased agent would equalize marginal quality across all test scores.

In this paper, I begin with the assumption that an agent is in fact biased. I search for an optimal contract restricting her behavior. Unsurprisingly, though, there is a connection to the problem of testing for bias. When the bias and information structure are known, the “upper bound acceptance rule” of Propositions 1 and 2 equalizes marginal qualities across test scores. In other words, it “de-biases” the agent by inducing her to select applicants in a manner that passes the bias test. Interestingly, however, agents are *not* fully de-biased by the optimal contract when their types are unknown.

The screening contract put forth in Proposition 5 proposes a minimum steepness of the acceptance rate. Some agent types will find this floor binding: a mix of those with better information, a stronger bias, or both. Those with a stronger bias will be picking worse marginal applicants at low test scores, and those with better information will be picking better marginal applicants at low test scores. Averaged across all agents at this floor, the quality of the marginal applicant is indeed equalized across test scores.²⁸ But some agents

²⁸ Recall from footnote 28 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.

may choose steeper acceptance rates than required: those with worse information and/or less 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 concept of commitment by the principal. Under some specific contract, if an agent’s choices pass the bias test, then the principal does not want to adjust the acceptance rates ex post. When the bias test is failed, on the other hand, the principal is tempted to intervene ex post. He wants the agent to accept more applicants at test scores with high marginal quality, and fewer at test scores with low marginal quality. In particular, if the agent chooses an acceptance rate steeper than what is required, the principal wants to force the agent to go back and choose an even steeper acceptance rate ex post. 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 underlying commitment logic is identical to that explored in one-dimensional delegation problems. Suppose an agent is always biased towards a one-dimensional 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 Uncertain distribution of applicant quality

What if the principal does not know the distribution of quality in the applicant population? In the body of the paper, Assumption 1 ruled out this possibility. The quality distribution F_Q was assumed to be common knowledge. But if F_Q were not known, the principal might want the hiring policy to respond to the realized distribution. For instance, under normal specification, Section 3.3 described how the principal’s acceptance rate as a function of test scores depended on the variance of applicant quality σ_Q^2 . Moreover, the principal might want to hire fewer applicants if the pool happens to be of lower quality. But the same assumption stated that the share of applicants hired, k , was predetermined as well.

To analyze the effect of aggregate uncertainty over applicant quality, think first of the normal specification. In the body of the paper I assumed that quality was drawn from a normal distribution with mean 0 and known variance σ_Q^2 . Now suppose that quality is instead drawn from a normal distribution with mean μ_Q and variance σ_Q^2 , both unknown to

the principal. As before, test scores of an applicant with quality q are drawn from a normal distribution with mean q and commonly known variance σ_T^2 .

In fact, it turns out that adding this aggregate uncertainty to the normal specification leaves the contracting problem essentially unchanged. The two parameters μ_Q and σ_Q^2 will be perfectly revealed by the distribution of applicant test scores. Specifically, the empirical test score distribution has mean μ_Q and variance $\sigma_Q^2 + \sigma_T^2$; the parameter σ_Q^2 can be calculated as the empirical variance minus σ_T^2 . So the principal can look at the test score distribution; infer μ_Q and σ_Q^2 ; choose his preferred k at this quality distribution; and then implement the contract going forward as if μ_Q had been normalized to 0, and as if σ_Q^2 and k had been known in advance.^{29,30}

Of course, moving outside of this specification, it will not necessarily be possible to fully infer the distribution of applicant quality from the realized test score distribution. The test itself may simply be too coarse. The principal may be jointly uncertain over both the distribution of applicant quality, and the map from quality to test score. Or, tests may measure only one out of multiple quality dimensions. In the two-factor model, for instance, the test might reveal the distribution of technical skills while being completely uninformative about the distribution of social skills.

Even so, this exercise points out that uncertainty over the quality of the applicant pool will not necessarily qualitatively change the induced contracting problem. The distribution of test scores at the interim stage will often reveal much of what was initially unknown. The principal can then proceed to implement the types of contracts studied previously.

5.3 Inference from performance data

Throughout this paper, I've assumed that the principal either knows an agent's type, or has priors over the distribution of this type. In the normal specification, for instance, a principal who knows an agent's type would fix the average test score of the applicants she accepts, or the steepness of the acceptance rate function (Proposition 3). If the principal instead has a prior distribution over types that satisfies appropriate regularity conditions, he would put a

²⁹Note that the argument in Section 2.2 for the irrelevance of an initial message stage no longer holds once F_Q is not known in advance by the agent. Initial messages are still irrelevant if the agent's type is common knowledge. But if the agent's type is unknown, it may be possible to screen across agent types by asking for an initial message prior to the arrival of applicants. So contracts with only an interim message can still replicate the earlier contracts; but it will be possible that some new contract form can do even better.

³⁰Likewise, if the number of applicants is ex ante uncertain, then for a given number of job openings the share hired k will be inversely proportional to the realized mass of applicants. But the principal can observe the number of applicants prior to the agent's report, set k , and proceed from there.

possibly binding floor on the average test score or steepness (Propositions 5 and 6).

Now suppose that the principal seeks to infer the optimal contract from data on the past performance of applicants. The principal may be unwilling to declare a prior over a given agent’s type. But he has seen this agent, or other ones, make hiring decisions in the past. The principal knows the contracts that agents faced, and the public test scores of each applicant who was hired. Moreover, the principal also observes the ex post performance – the quality realization – of each previously hired applicant. How can this information be used to adjust contracts going forward? To remain consistent with the previous analysis, I assume that performance data is only used to design new contracts. There is still no way to reward agents for hiring applicants who end up performing better.³¹

The simplest exercise is as follows. Suppose that (i) we are in the environment described by the normal specification; (ii) an agent had originally been given full discretion to hire her favorite k applicants, regardless of test score; 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 observes the agent’s acceptance rate, as well as the distribution of realized quality of the accepted applicants, at each test score. In fact, going forward, the principal can implement the optimal contract of Proposition 3 from two moments of the data. The principal need only calculate the *average test score* of accepted applicants, and their *average quality*. An agent who chose a higher average test score tends to be less biased and/or less informed; an agent who chose a higher average quality tends to be less biased and/or more informed. In Appendix E, I give an explicit formula for the optimal contract as a function of these two moments.

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 the outcomes will affect later periods. Second, performance data for the agent in question might only be available after a long amount of time. A more reasonable exercise may be to suppose that we do not use any individual agent’s past performance to update her own contract. Instead, we gather performance data from a pool of agents and use the aggregate results to adjust future contracts. In other words, we use the distribution of population outcomes to set the contract for an agent whose type is drawn from the same population.

Sticking with the normal specification, if we take the model literally then we can give

³¹I 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 E.

a sample of agents full discretion; observe the average test score and average performance of each agent; and then use this information to infer the joint distribution of bias and information (σ_S^2, σ_B^2) across the population of agents. Given the joint distribution, we can solve for the optimal contract. Alternately, we can take more of a first-order approach. Propositions 5 and 6 highlight a one-dimensional parametric class of contracts, those which give a floor on the average test score or the average 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 applicants.

We can heuristically apply a similar first-order approach well beyond the normal specification. The key qualitative conclusion was that we want to look for contracts which specify a floor on the steepness of the acceptance rate function, or on some moment of the distribution of test scores of accepted applicants. The principal can search for contracts with similar floors even if he doesn't restrict acceptance rate functions to be normal CDFs, or the moment function to be the mean. And it is much easier to search across contracts with these features than to search across the space of all possible contracts. Once he has come up with some parametric form for the steepness of a contract, for instance, he can evaluate performance at different parameters and iterate towards the best contract in this class.

The results of this paper also provide guidance for the acceptance rate or moment functions that one might use as a floor. Given functional form assumptions on information and bias, the principal can apply Propositions 1 and 2 to see what the optimal acceptance rate or moment functions would look like under common knowledge of type. These common-knowledge acceptance rate and moment functions are natural as proposed floors; under the normal specification, I showed that they could indeed be the optimal floors.

References

- 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.
- AMBRUS, A. AND G. EGOROV (2013): "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.
- CHAKRABORTY, A. AND R. HARBAUGH (2007): “Comparative cheap talk,” *Journal of Economic Theory*, 132, 70–94.
- 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.
- GUO, Y. (2014): “Dynamic delegation of experimentation,” Tech. rep., mimeo, Yale University.
- HOFFMAN, M., L. B. KAHN, AND D. LI (2015): “Discretion in Hiring,” .
- 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.
- KNOWLES, J., N. PERSICO, AND P. TODD (2001): “Racial Bias in Motor Vehicle Searches: Theory and Evidence,” *Journal of Political Economy*, 109.
- KOVÁČ, E. AND T. MYLOVANOV (2009): “Stochastic mechanisms in settings without monetary transfers: the regular case,” *Journal of Economic Theory*, 114, 1373–1395.
- 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.

STEIN, J. C. (2000): "Information production and capital allocation: Decentralized vs. hierarchical firms," Tech. rep., National Bureau of Economic Research.

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 was 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, \tilde{u}_A) -space through \hat{u}_P : $\hat{u}_P(t, \tilde{u}_A) \equiv \mathbb{E}[q|t, \tilde{u}_A]$. Now define a similar function, l , which tells us the expected quality for any *quantile* of \tilde{u}_A at a given test score. Specifically, at each test score t , there is a continuous conditional distribution of \tilde{u}_A over $\text{Supp}_{U_A}(t)$ which can be rewritten in terms of its quantiles (i.e., by going from a CDF to an inverse CDF): \tilde{u}_A increases in quantile at each t , with quantile 0 at the minimum of $\text{Supp}_{U_A}(t)$ and quantile 1 at the maximum. For $x \in [0, 1]$ and $t \in \mathcal{T}$, let $l(x; t)$ be the value $\hat{u}_P(t, \tilde{u}'_A)$ for \tilde{u}'_A at the $1 - x$ quantile of the distribution of \tilde{u}_A conditional on test score t . Higher x gives lower \tilde{u}'_A . If the agent were freely able to choose a share x' of the applicants at test score t , she would choose the ones with $x \in [0, x']$. Alignment up to distinguishability would be equivalent to the statement that $l(x; t)$ is weakly decreasing in x for every t .

When alignment up to distinguishability fails, there exist test scores for which $l(x; t)$ is not weakly decreasing. At these test scores, define an *ironed* version of the function l as follows. Let $L(x'; t) \equiv \int_0^{x'} l(x; t) dx$ for $x \in [0, 1]$ and $t \in \mathcal{T}$. Now “iron” L by defining $\bar{L}(\cdot; t)$ to be the concave hull of $L(\cdot; t)$. Finally, let the ironed l be defined as $\bar{l}(x; t) \equiv d\frac{\partial \bar{L}}{\partial x}(x; t)$. The function \bar{l} is defined almost everywhere by concavity of $\bar{L}(x; t)$ in x , and $\bar{l}(x; t)$ is weakly decreasing in x at every t . At any t for which $l(x; t)$ is weakly decreasing in x , it holds that $L(\cdot; t)$ is concave and so $\bar{l}(\cdot; t)$ is identically equal to $l(\cdot; t)$.

Now the optimizing acceptance rule – in quantile rather than \tilde{u}_A space – chooses the values (x, t) that lead to a mass of k acceptances, and which maximize the expectation of $\bar{l}(\cdot; t)$ on this mass. Monotonicity (acceptance rate decreasing in x) will follow from the fact that $\bar{l}(x; t)$ has been constructed to be weakly decreasing in t – it is the derivative of a concave function. The “ironing” constraint says that all applicants on an “ironing” interval at a given t (in which $\bar{L}(x; t) < L(x; t)$) must be treated identically, with the same acceptance

rate.

This acceptance rule amounts to finding a cutoff value l^c and accepting all applicants with $\bar{l}(x; t) > l^c$ and rejecting all those with $\bar{l}(x; t) < l^c$. One can choose arbitrarily on those with $\bar{l}(x; t) = l^c$ so long as we respect the ironing constraint. The cutoff l^c is chosen so k applicants are accepted in total.

Satisfying the ironing constraint is automatic if there are no “flat intervals” in which $\bar{l}(\cdot; t)$ is constant at l^c . So, suppose that there are such flat regions at some t . In that case we satisfy the ironing constraint by taking the acceptance probability to be constant over the flat interval at every such t .

One way of satisfying the ironing constraint is to choose the single acceptance probability in $[0, 1]$ for all applicants (x, t) with $\bar{l}(x; t) = l^c$, where the acceptance probability is set so that a total of k applicants are accepted. This does indeed give an optimal acceptance rule. It may also involve randomization at many test scores – any test score at which $\bar{l}(\cdot; t)$ has a flat interval at l^c .

Alternatively, we can satisfy the ironing constraint by choosing different acceptance probabilities for the flat regions at different test scores. It is always possible to order the (possibly multidimensional) test scores in \mathcal{T} in such a manner that the acceptance probability over the relevant flat region is set at 1 for test scores below a threshold t^* ; 0 for test scores above t^* ; and some intermediate level in $[0, 1]$ for the single threshold test score t^* . The threshold t^* is chosen so that there is an appropriate total probability of flat regions at l^c at lower test scores to get the total number of acceptances to k .

Importantly, in this alternative way of satisfying the ironing constraint, we see that there is at most a single test score 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 score. When test scores are continuous, of course, behavior at any single test score can be disregarded. So with continuously distributed test scores – no atom of probability at any single score – 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 \tilde{u}_A^i above some fixed level. Working through the normal specification algebra of Section 3.3, but replacing the UBAR acceptance cutoff line $\tilde{u}_A^c(t)$ with a constant in t , the agent with full discretion

chooses an acceptance rate of $\Phi(\gamma_T^{\text{FD}}t - \gamma_0)$ for

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

From (16), γ_T^* can be rewritten as $\gamma_T^{\text{FD}} + \frac{\sigma_Q^2\sigma_B^2}{\eta(\sigma_Q^2 + \sigma_T^2)\sqrt{\sigma_B^2 + \eta}}$. So we see that $0 < \gamma_T^{\text{FD}} < \gamma_T^*$.

Proposition 7. In the full discretion contract, γ_T^{FD} from (31) has the following comparative statics and limits:

1. γ_T^{FD} is independent of k .
2. $\frac{d\gamma_T^{\text{FD}}}{d\sigma_S^2} > 0$; $0 < \lim_{\sigma_S^2 \rightarrow 0} \gamma_T^{\text{FD}} < \lim_{\sigma_S^2 \rightarrow 0} \gamma_T^*$; and $\lim_{\sigma_S^2 \rightarrow \infty} \gamma_T^{\text{FD}} \in (0, \infty)$.
3. $\frac{d\gamma_T^{\text{FD}}}{d\sigma_B^2} < 0$; $\lim_{\sigma_B^2 \rightarrow 0} \gamma_T^{\text{FD}} = \lim_{\sigma_B^2 \rightarrow 0} \gamma_T^*$; and $\lim_{\sigma_B^2 \rightarrow \infty} \gamma_T^{\text{FD}} = 0$.
4. $\frac{d\gamma_T^{\text{FD}}}{d\sigma_T^2} < 0$ and $\frac{d}{d\sigma_T^2} \left(\gamma_T^{\text{FD}} \sqrt{\sigma_Q^2 + \sigma_T^2} \right) < 0$; $\lim_{\sigma_T^2 \rightarrow 0} \gamma_T^{\text{FD}} \in (0, \infty)$ and $\lim_{\sigma_T^2 \rightarrow 0} \gamma_T^{\text{FD}} \sqrt{\sigma_Q^2 + \sigma_T^2} \in (0, \infty)$; and $\lim_{\sigma_T^2 \rightarrow \infty} \gamma_T^{\text{FD}} = \lim_{\sigma_T^2 \rightarrow \infty} \gamma_T^{\text{FD}} \sqrt{\sigma_Q^2 + \sigma_T^2} = 0$.
5. $\frac{d\gamma_T^{\text{FD}}}{d\sigma_Q^2} > 0$ and $\frac{d}{d\sigma_Q^2} \gamma_T^{\text{FD}} \sqrt{\sigma_Q^2 + \sigma_T^2} > 0$; $\lim_{\sigma_Q^2 \rightarrow 0} \gamma_T^{\text{FD}} = \lim_{\sigma_Q^2 \rightarrow 0} \gamma_T^{\text{FD}} \sqrt{\sigma_Q^2 + \sigma_T^2} = 0$; and $0 < \lim_{\sigma_Q^2 \rightarrow \infty} \gamma_T^{\text{FD}} < \lim_{\sigma_Q^2 \rightarrow \infty} \gamma_T^*$ while $\lim_{\sigma_Q^2 \rightarrow \infty} \gamma_T^{\text{FD}} \sqrt{\sigma_Q^2 + \sigma_T^2} = \infty$.

There are a few main observations to make. The most important one is that, relative to Proposition 4, the sign changes on the derivative with respect to σ_B^2 . As discussed in the main text, we see that the principal (in the optimal contract with steepness γ_T^*) and agent (in the Full Discretion contract with steepness γ_T^{FD}) disagree about the effect of bias. The agent prefers flatter contracts when she is more biased, and the principal prefers steeper contracts. Compare that to the other dimension of the agent's type, σ_S^2 , parametrizing her information level. The principal and agent agree that a more informed agent should have a flatter acceptance rate function.

Second, the only other (potential) sign change is that we now get a clean comparative static on σ_Q^2 , rather than having an ambiguous sign. The full discretion acceptance rate gets steeper with respect to test scores (in absolute and relative senses) when the variance of population quality increases. With no quality variance, the agent's preferences become entirely driven by bias, and so the full discretion contract becomes flat even as the principal-optimal contract becomes arbitrarily steep.

Finally, this analysis confirms that as the agent's bias goes away, the agent's preferred outcome goes to that of the principal's optimal contract: $\gamma_T^{\text{FD}} \rightarrow \gamma_T^*$. Without bias, the incentives of the two parties are perfectly aligned.

C Additional Analysis of Systematic Biases

C.1 Two-factor model under Common Knowledge of Agent Type

Under the two-factor model, we have $\tilde{u}_A^i = \tilde{q}_1^i + \lambda \tilde{q}_2^i$. Rearranging, we get

$$\tilde{q}_2^i = \frac{\tilde{u}_A^i - \tilde{q}_1^i}{\lambda}.$$

Recall that by the assumption that $\mathbb{E}[q_1|s, t] = \mathbb{E}[q_1|t]$, \tilde{q}_1^i is observable to the principal: $\tilde{q}_1^i = \tilde{q}(t^i)$. We can now calculate the principal's expected utility \hat{u}_P given t and \tilde{u}_A as

$$\hat{u}_P(t, \tilde{u}_A) = \mathbb{E}[q_1 + q_2|t, \tilde{u}_A] = \tilde{q}_1(t) + \frac{\tilde{u}_A - \tilde{q}_1(t)}{\lambda} = \frac{(\lambda - 1)\tilde{q}_1(t) + \tilde{u}_A}{\lambda}.$$

The coefficient $\frac{1}{\lambda}$ on \tilde{u}_A is always positive, implying that utilities are aligned up to distinguishability and that UBAR can be implemented.

Lemma 6. Under the two-factor model with common knowledge of agent type, utilities are aligned up to distinguishability and so the optimal contract can be implemented as described in Propositions 1 and Proposition 2.

In fact, because of the assumed absence of idiosyncratic biases, UBAR actually implements the principal's first-best outcomes given all public and private information – regardless of misalignment parameter λ . If we added idiosyncratic biases (as in Appendix D), we would need distributional assumptions to guarantee the implementability of UBAR, and UBAR would only achieve a second-best solution.

The sign of the coefficient $\frac{\lambda-1}{\lambda}$ on $\tilde{q}_1(t)$ in \hat{u}_P depends on the agent's bias term λ . Supposing that test scores are normalized so that higher t yields higher $\tilde{q}_1(t)$, the sign determines whether indifference curves in (t, \tilde{u}_A) -space are sloped upwards or downwards. (Under a further normalization to $\tilde{q}_1(t) = t$, we would get linear indifference curves.) For the advocate with $\lambda > 1$, there is a positive coefficient on $\tilde{q}_1(t)$, 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 a candidate with a low test score indicating $\tilde{q}_1 = 0$, and one with a high test score indicating $\tilde{q}_1 = 1$. For the agent to be indifferent, it must mean that the agent observes that \tilde{q}_2 is one-half a unit lower for the candidate with a high test score. So the principal prefers the candidate with the high test score: one unit higher \tilde{q}_1 , one-half unit lower \tilde{q}_2 . For the cynic with $\lambda < 1$, however, we get upward-sloping indifference curves. Think now of an agent with $\lambda = \frac{1}{2}$ facing the same low and high test score candidates. If

this cynical agent is indifferent, then the principal prefers the candidate with the lower test score: one unit lower \tilde{q}_1 , two units higher \tilde{q}_2 .

One implementation of UBAR is to fix the acceptance rate $\alpha(t)$ as a function of the test score. We see that the acceptance rate is determined by the joint distribution of \tilde{q}_1 and \tilde{q}_2 ; the principal selects the k applicants with the highest $\tilde{q}_1 + \tilde{q}_2$ as if these were both observable. The acceptance rate at test score t is just the share of applicants with $\tilde{q}_1 + \tilde{q}_2$ in the top k , conditional on $\tilde{q}_1 = t$. Moreover, while $\alpha(t)$ depends on the details of the joint distribution of q_1 and q_2 and the agent’s information structure, the acceptance rate does not depend on the bias term λ . The same acceptance rate implements first-best for all levels of bias λ , no difference across cynics or advocates. Restating this point, fixing the acceptance rate in this manner yields the optimal contract for a principal with *any* beliefs on λ , not just for λ commonly known.

We can also implement UBAR in the alternative manner by fixing the average first-factor quality score $\mathbb{E}[\tilde{q}_1(t^i)|i \text{ hired}]$ at some level K . One distinction between agents who are cynics ($\lambda < 1$) and advocates ($\lambda > 1$) does follow from the different signs of the indifference curves. Advocates would prefer to push the test score below the principal’s preferred level K . They view the constraint as a binding floor. Cynics want to push average test scores higher, and so for them the constraint is a ceiling.

C.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. That is, the publicly observable characteristics are valued independently of their information content. 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 as long as he or she looks better on paper. The admission 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. This induces a bias – misaligned objectives – if the racial preferences are not perfectly shared by both parties.

Let “true quality” be denoted q_1^i for applicant i . We have distribution $q_1 \sim F_{Q1}$ of true quality, and then signal distributions $t \sim F_T(\cdot|q_1)$ and $s \sim F_S(\cdot|q_1, t)$ that are correlated with

q_1 . Let the expected value of true quality given all information be $\tilde{q}_1^i = \mathbb{E}[q_1^i | s^i, t^i]$.³²

Then there is an addition to the principal utility, $q_{2P}^i = q_{2P}(t^i)$, and an addition to agent utility, $q_{2A}^i = q_{2A}(t^i)$, where $q_{2P}(\cdot)$ and $q_{2A}(\cdot)$ are arbitrary functions of the realization of hard information. So utilities are as follows:

$$\begin{aligned} u_P^i &= q^i \equiv q_1^i + q_{2P}(t^i) = q^i \\ u_A^i &= q_1^i + q_{2A}(t^i) = q^i + b^i, \text{ for } b^i = q_{2A}(t^i) - q_{2P}(t^i) \end{aligned}$$

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 for this example: at any test score, applicants preferred by the agent are preferred by the principal. Formally, given t^i and $\tilde{u}_A^i = \tilde{q}_1^i + q_{2A}(t^i)$, we can solve for \tilde{q}_1^i :

$$\tilde{q}_1^i = \tilde{u}_A^i - q_{2A}(t^i).$$

So the induced principal utility \hat{u}_P^i at test score t^i and agent utility \tilde{u}_A^i is

$$\hat{u}_P^i(t^i, \tilde{u}_A^i) = \tilde{q}_1^i + q_{2P}(t^i) = \tilde{u}_A^i - q_{2A}(t^i) + q_{2P}(t^i)$$

which is increasing in \tilde{u}_A^i . Hence, we can apply the results of Section 3 to solve for the optimal acceptance rate. Just as with the two-factor model, this acceptance rate does not depend on the agent's bias function $q_{2A}(\cdot)$. So this acceptance rate in fact implements the optimal contract for any agent bias $q_{2A}(\cdot)$, and any beliefs that the principal may have on the bias. (Due to the lack of idiosyncratic bias, it actually implements the first-best payoff for the principal.)

D 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 in the body of the paper all extend to this model as well.

³²For the case of a college which cares about its ranking, the test score could be one-dimensional; the fact that colleges are ranked on their average test scores leads them to overweight this factor in admissions. For the case of racial preferences, we would instead think about the hard information as having two or more dimensions that might or might not be correlated: race, plus a signal of true quality.

For this section, I write signals and utilities in a somewhat reduced form manner for clarity. It is straightforward to “microfound” these utilities through appropriate joint-normal priors on the two quality factors and normally distributed signals; see footnote 34, below.

Let the public signal t^i reveal everything that can be inferred about quality factor q_1 ; normalize this signal realization so that

$$\tilde{q}_1^i = \tilde{q}_1(t^i) = t^i.$$

Then the private signal s^i gives additional information about a second quality factor q_2 . The expected value of q_2 given s^i and t^i is assumed to be

$$\tilde{q}_2^i = \tilde{q}_2(t^i, s^i) = rt^i + s^i,$$

where $s^i \sim \mathcal{N}(0, l \cdot \sigma_2^2)$ for $l \in (0, 1)$. The interpretation is that the underlying distribution of q_2^i , conditional on the observation of $t^i = t$, has mean rt and residual 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 with who perfectly observed the realization of q_2 would have $l \rightarrow 1$, and one who received no private information would have $l \rightarrow 0$.³³

Principal and agent utilities of hiring applicant i are as follows:

$$\begin{aligned} \text{Principal:} & \quad \tilde{q}^i = \tilde{q}_1^i + \tilde{q}_2^i = t^i(1 + r) + s^i \\ \text{Agent:} & \quad \tilde{u}_A^i = \tilde{q}_1^i + \lambda \tilde{q}_2^i + \epsilon_B^i = t^i(1 + \lambda r) + \lambda s^i + \epsilon_B^i \end{aligned}$$

for $\lambda > 0$ and $\epsilon_B^i \sim \mathcal{N}(0, \sigma_B^2)$. We have a systematic bias induced by λ , if $\lambda \neq 1$, just as in the two-factor model. Then we have an additional idiosyncratic bias introduced by ϵ_B as in

³³As stated, I do not commit to the details of the updating model that would get us these posteriors. But here is one such model. 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 empirical 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 (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).

the normal specification. The agent's utility can be rewritten in terms of her net bias b as

$$\tilde{u}_A^i = \tilde{q}^i + b^i \text{ for } b^i = (\lambda - 1)(rt^i + s^i) + \epsilon_B^i.$$

The agent's type θ contains 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 – the new term relative to the normal specification – $\lambda \in (0, \infty)$ for systematic bias.³⁴

For any fixed t , it holds that \tilde{q} and \tilde{u}_A conditional on t but unconditional on s and ϵ_B have a joint normal distribution. Specifically,

$$\text{Conditional on } t^i = t : \tilde{u}_A^i \sim \mathcal{N}(\mu_{U_A}(t), \sigma_{U_A}^2), \text{ for} \quad (32)$$

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

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

Likewise, conditional on t , \tilde{q} has mean $t(1 + r)$ and variance $l\sigma_2^2$. Moreover, still conditional on t , the covariance of \tilde{q} with \tilde{u}_A is $\lambda l\sigma_2^2$. We can iteratively calculate the posterior mean of \tilde{q} given t and \tilde{u}_A , indicated $\hat{u}_P(t, \tilde{u}_A) = \mathbb{E}[\tilde{q}|t, \tilde{u}_A]$, as the expectation of \tilde{q} given \tilde{u}_A , conditional

³⁴Without going into the algebraic details, the combined model here embeds the normal specification after we make the following notational adjustment. Call s and t the signals in the normal specification, and s' and t' the signals in the combined model. Then the normal specification maps into the combined model if we take $\lambda = 1$ (indicating that there is no systematic bias), along with

$$s' = \left(s - t \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} \right) \cdot \frac{\sigma_T^2 \sigma_Q^2}{\sigma_S^2 \sigma_T^2 + \sigma_Q^2 \sigma_S^2 + \sigma_Q^2 \sigma_T^2} \quad t' = t \cdot \frac{\sigma_S^2 \sigma_Q^2}{\sigma_S^2 \sigma_T^2 + \sigma_Q^2 \sigma_S^2 + \sigma_Q^2 \sigma_T^2}$$

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

on t :³⁵

$$\hat{u}_P(t, \tilde{u}_A) = \beta_T t + \beta_{U_A} \tilde{u}_A, \text{ for} \quad (35)$$

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

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

Utilities are aligned up to distinguishability: the coefficient β_{U_A} on agent utility is positive. Adding idiosyncratic biases reduces the coefficient, making beliefs on quality less responsive to agent utilities, but does not change the sign. Let us look next at the coefficient β_T on test scores. Without idiosyncratic shocks (plugging in $\sigma_B^2 = 0$), β_T reduces to $\frac{\lambda-1}{\lambda}$ as in the two-factor model. Adding idiosyncratic shocks through σ_B^2 pulls the coefficient β_T towards $1 + r$ (and also makes the coefficient depend on parameters other than λ). We see that adding idiosyncratic preference shocks takes the principal's belief, at any given test score and utility realization, in the direction of $(1 + r)t$ – the estimate of quality given t , unconditional on \tilde{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, sufficiently large idiosyncratic biases σ_B^2 switches the sign of β_T for cynical agents ($\lambda < 1$) from negative to positive.

From the equation for \hat{u}_P , we see that \hat{u}_P conditional on t (but across realizations of \tilde{u}_A) has mean and variance of

$$\begin{aligned} \mu_{U_P}(t) &= t(1 + r) \\ \sigma_{U_P}^2 &= \beta_{U_A}^2 \sigma_{U_A}^2 = \frac{(\lambda \sigma_2^2)^2}{\lambda^2 l \sigma_2^2 + \sigma_B^2}. \end{aligned}$$

In the original normal specification, the principal and agent had equal mean utilities at a given test score. 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 $(1 + \lambda r)$. For positive correlation ($r > 0$), 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.

³⁵Note that I have not assumed that t is jointly normally distributed with \tilde{q} and \tilde{u}_A (although such a result would follow from the “microfoundation” of footnote 34). So I cannot directly apply the formula for conditional means of a joint normal distribution, as in Equation (48) of Appendix (F.1). However, I can still calculate $\hat{u}_P(t, \tilde{u}_A)$ from the more direct assumptions on the joint normal distribution of \tilde{q} and \tilde{u}_A at each given t .

D.1 Common knowledge of 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 principal chooses a cutoff utility \hat{u}_P^c and finds a policy in which the agent will accept all applicants with $\hat{u}_P \geq \hat{u}_P^c$. This is implemented by a normal CDF acceptance rate, $\alpha(t) = \Phi(\gamma_T t - \gamma_0)$, at some appropriate steepness $\gamma_T = \gamma_T^*$. We can solve for this optimal coefficient (from the equations for the distribution of \hat{u}_P conditional on t) as

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

The acceptance rate is increasing in the test score (positive γ_T^*) 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 sufficiently strong negative correlation with $r < -1$, then higher quality applicants tend to have lower test scores (they are lower on the first quality factor), and fewer are accepted.³⁶

Thanks to the linearity of \hat{u}_P in both t and \tilde{u}_A , the policy could also be implemented by fixing the average test score of hired workers. Gathering together these observations:

Proposition 8. Under the combined model with common knowledge of agent type, utilities are aligned up to distinguishability. 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 of $\alpha(t) = \Phi(\gamma_T^* t - \gamma_0)$ at test score t , for γ_T^* given by (38) and γ_0 set so that a total of k applicants are accepted; or,
2. An average test score of accepted applicants equal to some value τ^* .

D.2 Unknown agent type

With uncertainty over the agent's type $\theta = (l, \lambda, \sigma_B^2)$, we can also replicate much of the analysis of the normal specification in solving for the optimal policy. Say that θ follows distribution function G . It now holds that μ_{U_A} , σ_{U_A} , and σ_{U_P} all depend on θ , so I will write those as functions of θ from here on out. For this analysis let us assume that the empirical

³⁶While we would expect t to be normally distributed in some microfounded normal prior/normal signal specification, this analysis did not impose any assumptions on the empirical 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.

distribution of test scores t is normally distributed, with mean normalized to 0 and variance of Var_T . (We could justify the normal distribution as arising from $t = q_1$ with q_1 normally distributed; see footnote 34.)

Define τ , z^i , and ζ as before:

$$\begin{aligned}\tau &\equiv \mathbb{E}[t^i | i \text{ hired}] \\ z^i &\equiv \frac{\tilde{u}_A^i - \mu_{U_A}(t^i; \theta)}{\sigma_{U_A}(\theta)} \\ \zeta &\equiv \mathbb{E}[z^i | i \text{ hired}].\end{aligned}$$

The outcome space in terms of (τ, ζ) is exactly as in Lemma 3, with $R_Z = R(k) = \frac{1}{k}\phi(\Phi^{-1}(1-k))$ 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 $\zeta \in [-R_Z, R_Z]$.

For an agent of type θ hiring applicant i , where applicant i has utility z-score of z^i and test score of t^i , the utilities to the principal and agent are

$$\begin{aligned}\text{P: } \hat{u}_P^i &= \mu_{U_P}(t^i) + \sigma_{U_P}(\theta)z^i = (1+r)t^i + \sigma_{U_P}(\theta)z^i \\ \text{A: } \tilde{u}_A^i &= \mu_{U_A}(t^i; \theta) + \sigma_{U_A}(\theta)z^i = (1+\lambda r)t^i + \sigma_{U_A}(\theta)z^i\end{aligned}$$

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

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

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 $\sigma_{U_A}(\theta)\frac{1+r}{1+\lambda r}\zeta + (1+r)\tau$. Let us define

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

We can now reduce the agent's type determining her behavior from the three-dimensional $\theta = (l, \lambda, \sigma_B^2)$ into the one-dimensional $\rho(\theta)$. All agents with $\rho(\theta) = \tilde{\rho}$ act identically. Let the distribution of $\rho(\theta) = \tilde{\rho}$, induced by $\theta \sim G$, be given by H .

Because the principal can never distinguish agents with the same realization of $\tilde{\rho}$, it is also convenient to define the principal's average value of σ_{U_P} across all agent types with

$\rho(\theta) = \tilde{\rho}$ as $\hat{\sigma}_{U_P}(\tilde{\rho})$:

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

We can now rewrite the principal and agent maximization problems as

$$\text{Agent: } \max \left(\tilde{\rho} \cdot \zeta + (1+r)\bar{\tau}(\zeta) \right) - \delta \quad (39)$$

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

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

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 $\tilde{\rho}$.

We can now give the analog of Proposition 5.

Proposition 9. Let the distribution H have bounded support with pdf h . If $H(\tilde{\rho}) + (\tilde{\rho} - \hat{\sigma}_{U_P}(\tilde{\rho}))h(\tilde{\rho})$ is nondecreasing, then the optimal contract can be characterized in either of the following two ways:

1. 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 $K > 0$.
2. The agent is allowed to choose 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 $\Gamma > 0$.

This result embeds Proposition 5 – up to some changes of notation in the combined model versus the normal specification – 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 to account for a commonly known value of $\lambda \neq 1$, or a distribution of λ across agents. The main takeaway from this result is that the simple contract form of a minimum on the average test score, or a minimum average steepness of the acceptance rate, is robust to the addition of systematic biases in addition to idiosyncratic ones.

E 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_B^2 and σ_S^2 , is not known. (Below, I address how one could also infer σ_Q^2 and σ_T^2 from performance data.) I proceed in a prior-free manner and do not specify the principal’s prior beliefs over the agent’s type.

In the “first period” the principal gives the agent a full discretion contract. For each applicant i that is hired, the principal observes not only their public test score t^i , but also their realized performance – their quality q^i . Then in the “second period” the principal solves the problem in the body of the paper, giving the agent a contract that maximizes his payoff. The principal has the opportunity to incorporate the information about the agent’s first-period behavior when choosing the second-period contract. Assume that the agent acts myopically in the sense that her behavior in the first period maximizes her first period payoff; 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 a considerable amount of data: the acceptance rate as a function of test scores, plus the entire distribution of realized qualities for the accepted applicants at each score. However, it is enough for the agent to look at two moments of this data in setting up the optimal contract. Let τ_1 be the average test score of the applicants accepted in the first period, and let ξ_1 be the average of the realized quality.³⁷ We can use τ_1 and ξ_1 to set the contract in the second period as the optimal contract as if the agent’s type were known. As in Proposition 3 part 2, the optimal 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 K .

Test scores of first-period applicants follow a normal distribution with mean normalized to 0 and variance $\sigma_Q^2 + \sigma_T^2$. Given that the number of applicants to be accepted is k , the range of possible average test scores of accepted applicants in period 1 is $[-R_T, R_T]$, for $R_T = \sqrt{\sigma_Q^2 + \sigma_T^2} R(k)$, as in (24) and (25). Let $\bar{\zeta}(\tau)$ be the highest possible ζ at an average

³⁷There are many other ways to calculate the parameters for the optimal contract. If the model is “true”, then they all give the same answer. For instance, given any average test score τ_1 , the model predicts that the acceptance rate at each test score t will be $\Phi(\gamma_T \cdot t - \gamma_0)$, where γ_T and γ_0 are parameters that can be calculated from τ_1 combined with σ_Q^2 , σ_T^2 , k . Of course, in any real-world data there will be deviations from a model. In that sense it is convenient that I can do these calculations with two simple means. The means can be calculated even if the data does not look exactly like the model predicts, and they don’t throw out any of the quality or test score observations.

test score of τ , from Lemma 3, plugging in R_T and R_Z in terms of $R(k)$:

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

Given full discretion in period 1, the principal can infer that if an average test score of τ_1 was observed, then the average z-score must have been $\zeta_1 = \bar{\zeta}(\tau_1)$.

Following (22) and (23), the agent and principal payoffs over τ and $\zeta = \bar{\zeta}(\tau)$ are determined by the respective parameters $\sigma_{U_A}(\theta)$ and $\sigma_{U_P}(\theta)$. From (23), the principal's optimal contract specifies that $\tau_2 = \tau^*$, for τ^* solving

$$\begin{aligned} \tau^* &= \arg \max_{\tau} \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} \tau + \sigma_{U_P}(\theta) \bar{\zeta}(\tau) \\ \Rightarrow 0 &= \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} + \sigma_{U_P}(\theta) \bar{\zeta}'(\tau^*) \\ \Rightarrow \tau^* &= \frac{\sigma_Q^2 \sqrt{\sigma_Q^2 + \sigma_T^2} \cdot R(k)}{\sqrt{\sigma_Q^2 + \sigma_{U_P}^2(\theta)}} \end{aligned} \quad (42)$$

Of course, the value of $\sigma_{U_P}(\theta)$ depends on the agent's type, which the principal is trying to learn from the data. 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 (23) (with $\zeta = \bar{\zeta}(\tau_1)$) to infer $\sigma_{U_P}(\theta)$:

$$\begin{aligned} \xi_1 &= \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} \tau_1 + \sigma_{U_P}(\theta) \bar{\zeta}(\tau_1) \\ \Rightarrow \sigma_{U_P}(\theta) &= \frac{\xi_1(\sigma_Q^2 + \sigma_T^2) - \sigma_Q^2 \tau_1}{(\sigma_Q^2 + \sigma_T^2) R(k)^2 - \tau_1^2} \end{aligned} \quad (43)$$

Now plug this value of $\sigma_{U_P}(\theta)$ into (42) to get

$$\tau^* = \frac{\sigma_Q^2 \sqrt{\sigma_Q^2 + \sigma_T^2} \cdot R(k)}{\sqrt{\sigma_Q^2 + \frac{(\xi_1(\sigma_Q^2 + \sigma_T^2) - \sigma_Q^2 \tau_1)^2}{(\sigma_Q^2 + \sigma_T^2) R(k)^2 - \tau_1^2}}}. \quad (44)$$

The optimal contract in the second period lets the agent accept any applicants she wants, subject to requiring the agent to set the period-2 average test score to the value of τ^* in

(44).³⁸

Lemma 7. Over the range of ξ_1 and τ_1 consistent with the model, $\frac{\partial \tau^*}{\partial \xi_1} < 0$ while $\frac{\partial \tau^*}{\partial \tau_1} > 0$.

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 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.

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. We know that (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

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

Now plug in the value of σ_{UP} from (43), replacing all occurrences of $\sigma_Q^2 + \sigma_T^2$ with Var_T :

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

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

$$\sigma_Q^2 = \frac{\bar{q}_1(t) - \frac{\xi_1 \text{Var}_T R(\alpha(t))}{\text{Var}_T R(k)^2 - \tau_1^2}}{\frac{t}{\text{Var}_T} - \frac{\tau_1 R(\alpha(t))}{\text{Var}_T R(k)^2 - \tau_1^2}}. \quad (45)$$

³⁸One could also solve for the optimal contract even if the observable parameters $(k, \sigma_Q^2, \sigma_T^2)$ were to change from period 1 to 2. But one would not simply plug in the period-1 value of $\sigma_{UP}(\theta)$ from (43) into the period-2 expression (42). The value of $\sigma_{UP}(\theta)$ depends on σ_T^2 and σ_Q^2 , which might change from period to period.

Of course, under the 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 (45), we have $\sigma_T^2 = \text{Var}_T - \sigma_Q^2$.

F Proofs

F.1 Proofs for Section 3

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 score t and some agent utilities $\tilde{u}_A^{\text{low}} < \tilde{u}_A^{\text{high}}$. Alignment implies that $\hat{u}_P(t, \tilde{u}_A^{\text{low}}) \leq \hat{u}_P(t, \tilde{u}_A^{\text{high}})$. I seek to confirm that acceptance rates have the ordering $\chi^{\text{UBAR}}(t, \tilde{u}_A^{\text{low}}) \leq \chi^{\text{UBAR}}(t, \tilde{u}_A^{\text{high}})$. If $\hat{u}_P(t, \tilde{u}_A^{\text{low}}) < \hat{u}_P^c$ then acceptance probabilities must be so ordered because $\chi^{\text{UBAR}}(t, \tilde{u}_A^{\text{low}}) = 0$; similarly if $\hat{u}_P(t, \tilde{u}_A^{\text{high}}) > \hat{u}_P^c$, because $\chi^{\text{UBAR}}(t, \tilde{u}_A^{\text{low}}) = 1$. The remaining possibility is that $\hat{u}_P(t, \tilde{u}_A^{\text{low}}) = \hat{u}_P(t, \tilde{u}_A^{\text{high}}) = \hat{u}_P^c$. In that case, acceptance probabilities are ordered due to the selection of χ^{UBAR} as monotonic over the flexible region. \square

Proof of Proposition 2. Take some cutoff function $\tilde{u}_A^c(t)$. C is an affine transformation of \tilde{u}_A^c , so let $C(t) = a_0 + a_1 \tilde{u}_A^c(t)$ for $a_0 \in \mathbb{R}$ and $a_1 \in \mathbb{R} \setminus \{0\}$. Equivalently, $\tilde{u}_A^c(t) = \frac{C(t) - a_0}{a_1}$. We assume that the expectation of $C(t)$ exists and is finite, which implies that $C(\cdot)$ and $\tilde{u}_A^c(\cdot)$ are almost everywhere finite-valued.

The agent's problem can be written as choosing the acceptance rule χ to maximize her objective $\frac{1}{k} \mathbb{E}_\Omega[\chi(t, \tilde{u}_A) \cdot \tilde{u}_A]$, subject to the two constraints of accepting k applicants and of setting the expectation of $C(t)$ conditional on hiring to K :

$$\mathbb{E}_\Omega[\chi(t, \tilde{u}_A)] = k \tag{46}$$

$$\frac{1}{k} \mathbb{E}_\Omega[\chi(t, \tilde{u}_A) \cdot C(t)] = K. \tag{47}$$

I claim that this problem is solved by χ^{UBAR} .

By construction, χ^{UBAR} satisfies the constraints (46) and (47). Suppose for the sake of a contradiction that $\hat{\chi}$ also satisfies the constraints, but yields a strictly higher value of the

objective:

$$\begin{aligned} & \frac{1}{k} \mathbb{E}_\Omega [\hat{\chi}(t, \tilde{u}_A) \tilde{u}_A] > \frac{1}{k} \mathbb{E}_\Omega [\chi^{\text{UBAR}}(t, \tilde{u}_A) \tilde{u}_A] \\ \Rightarrow & \frac{1}{k} \mathbb{E}_\Omega [\hat{\chi}(t, \tilde{u}_A) \cdot \tilde{u}_A] - \frac{K}{a_1} + \frac{a_0}{a_1} > \frac{1}{k} \mathbb{E}_\Omega [\chi^{\text{UBAR}}(t, \tilde{u}_A) \cdot \tilde{u}_A] - \frac{K}{a_1} + \frac{a_0}{a_1}. \end{aligned}$$

Now apply (46) and (47) to both $\hat{\chi}$ and χ^{UBAR} to bring the extra terms inside the expectations:

$$\begin{aligned} & \frac{1}{k} \mathbb{E}_\Omega \left[\hat{\chi}(t, \tilde{u}_A) \cdot \left(\tilde{u}_A - \frac{C(t)}{a_1} + \frac{a_0}{a_1} \right) \right] > \frac{1}{k} \mathbb{E}_\Omega \left[\chi^{\text{UBAR}}(t, \tilde{u}_A) \cdot \left(\tilde{u}_A - \frac{C(t)}{a_1} + \frac{a_0}{a_1} \right) \right] \\ \Rightarrow & \frac{1}{k} \mathbb{E}_\Omega [\hat{\chi}(t, \tilde{u}_A) \cdot (\tilde{u}_A - \tilde{u}_A^c(t))] > \frac{1}{k} \mathbb{E}_\Omega [\chi^{\text{UBAR}}(t, \tilde{u}_A) \cdot (\tilde{u}_A - \tilde{u}_A^c(t))]. \end{aligned}$$

But this last line yields a contradiction, because χ^{UBAR} maximizes $\mathbb{E}_\Omega[\chi(t, \tilde{u}_A)(\tilde{u}_A - \tilde{u}_A^c(t))]$ pointwise over all possible χ taking values in $[0, 1]$. It sets χ as high as possible, to 1, if $\tilde{u}_A > \tilde{u}_A^c$; and as low as possible, to 0, if $\tilde{u}_A < \tilde{u}_A^c$.³⁹ \square

Before I address the proofs of the formally stated results of Section 3.3, let me work out the derivations of formulas (8)-(14) 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 $\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$. Then the conditional distribution of X_1 given $X_2 = x_2$ is given by

$$X_1 | X_2 = x_2 \sim \mathcal{N}(\mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (x_2 - \mu_2), \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}). \quad (48)$$

Lemma 8. The realizations of $(q^i, t^i, \tilde{u}_A^i)$ are joint normally distributed, with means of 0 and covariance matrix of

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

³⁹As described in the main text, the proof approach follows the standard Lagrangian 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, which have already been plugged in, are $\lambda_0 = -\frac{a_0}{a_1}$ on (46) and $\lambda_1 = \frac{1}{a_1}$ on (47).

Given this result, we can apply (48) to calculate $\hat{u}_P(t, \tilde{u}_A)$, defined as the conditional expectation of q^i given $t^i = t$ and $\tilde{u}_A^i = \tilde{u}_A$ – take q^i as X_1 and (t^i, \tilde{u}_A^i) as X_2 . Working out the algebra yields Equations (8)-(11). We then apply (48) with t^i as X_1 and \tilde{u}_A^i as X_2 to find the conditional mean and variance of t^i given \tilde{u}_A^i in Equations (12)-(14). The proof of the Lemma is at the end of this subsection.

Proof of Lemma 1. Follows from arguments in the text. \square

Proof of Proposition 3. Follows from Propositions 1 and 2, given arguments in the text. \square

Proof of Proposition 4. Restating (11) and (16),

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

1. The parameter k does not appear in the formula for γ_T^* .
2. The parameter σ_S^2 appears only in γ_T^* through η . Routine differentiation shows that $\frac{d\gamma_T^*}{d\eta} < 0$ 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,

$$\begin{aligned}\lim_{\sigma_S^2 \rightarrow 0} \gamma_T^* &= \frac{1}{\sigma_T^2} \sqrt{\frac{\sigma_T^2 \sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} + \sigma_B^2}, \text{ because } \lim_{\sigma_S^2 \rightarrow 0} \eta = \frac{\sigma_T^2 \sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} \\ \lim_{\sigma_S^2 \rightarrow \infty} \gamma_T^* &= \infty, \text{ because } \lim_{\sigma_S^2 \rightarrow \infty} \eta = 0.\end{aligned}$$

3. 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,

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

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

4. 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 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 $\frac{d(\gamma_T^* \sqrt{\sigma_Q^2 + \sigma_T^2})}{d\sigma_T^2}$ as $\frac{\partial(\gamma_T^* \sqrt{\sigma_Q^2 + \sigma_T^2})}{\partial \sigma_T^2} + \frac{\partial(\gamma_T^* \sqrt{\sigma_Q^2 + \sigma_T^2})}{\partial \eta} \frac{d\eta}{d\sigma_T^2}$. Once again, $\frac{\partial(\gamma_T^* \sqrt{\sigma_Q^2 + \sigma_T^2})}{\partial \sigma_T^2} < 0$, $\frac{\partial(\gamma_T^* \sqrt{\sigma_Q^2 + \sigma_T^2})}{\partial \eta} < 0$, and $\frac{d\eta}{d\sigma_T^2} > 0$. Therefore $\frac{d(\gamma_T^* \sqrt{\sigma_Q^2 + \sigma_T^2})}{d\sigma_T^2} < 0$.

Taking limits,

$$\lim_{\sigma_T^2 \rightarrow 0} \gamma_T^* = \infty, \text{ because } \lim_{\sigma_T^2 \rightarrow 0} \eta = 0$$

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

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

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

5. Numerical examples show that, depending on parameters, γ_T^* and $\gamma_T^* \sqrt{\sigma_Q^2 + \sigma_T^2}$ can both be either increasing or decreasing in σ_Q^2 .

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

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

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

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

Proof of Lemma 8. From the specified joint distributions of q and t , it follows that $\text{Var}(q) = \sigma_Q^2$, $\text{Var}(t) = \sigma_Q^2 + \sigma_T^2$, and $\text{Cov}(t, q) = \sigma_Q^2$. It remains to calculate $\text{Var}(\tilde{u}_A)$, $\text{Cov}(t, \tilde{u}_A)$, and $\text{Cov}(q, \tilde{u}_A)$.

It will be helpful to note as well that $\text{Cov}(s, q) = \sigma_Q^2$, $\text{Cov}(s, t) = \sigma_Q^2$, and $\text{Var}(s) = \sigma_S^2 + \sigma_Q^2$. The bias term b has variance σ_B^2 , and has 0 covariance with s , t , or q .

From (7),

$$\tilde{u}_A = \tilde{q} + b = \frac{\frac{t}{\sigma_T^2} + \frac{s}{\sigma_S^2}}{\frac{1}{\sigma_Q^2} + \frac{1}{\sigma_T^2} + \frac{1}{\sigma_S^2}} + b.$$

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

$$\text{Cov}(q, \tilde{u}_A) = \frac{\frac{\text{Cov}(t, q)}{\sigma_T^2} + \frac{\text{Cov}(s, q)}{\sigma_S^2}}{\frac{1}{\sigma_Q^2} + \frac{1}{\sigma_T^2} + \frac{1}{\sigma_S^2}} = \frac{\frac{\sigma_Q^2}{\sigma_T^2} + \frac{\sigma_Q^2}{\sigma_S^2}}{\frac{1}{\sigma_Q^2} + \frac{1}{\sigma_T^2} + \frac{1}{\sigma_S^2}} = \frac{\sigma_Q^4(\sigma_S^2 + \sigma_T^2)}{\sigma_S^2\sigma_T^2 + \sigma_T^2\sigma_Q^2 + \sigma_Q^2\sigma_S^2}.$$

$\text{Cov}(t, \tilde{u}_A)$ is given by

$$\text{Cov}(t, \tilde{u}_A) = \frac{\frac{\text{Var}(t)}{\sigma_T^2} + \frac{\text{Cov}(s, t)}{\sigma_S^2}}{\frac{1}{\sigma_Q^2} + \frac{1}{\sigma_T^2} + \frac{1}{\sigma_S^2}} = \frac{\frac{\sigma_Q^2 + \sigma_T^2}{\sigma_T^2} + \frac{\sigma_Q^2}{\sigma_S^2}}{\frac{1}{\sigma_Q^2} + \frac{1}{\sigma_T^2} + \frac{1}{\sigma_S^2}} = \sigma_Q^2 \frac{\frac{1}{\sigma_Q^2} + \frac{1}{\sigma_T^2} + \frac{1}{\sigma_S^2}}{\frac{1}{\sigma_Q^2} + \frac{1}{\sigma_T^2} + \frac{1}{\sigma_S^2}} = \sigma_Q^2.$$

And finally, $\text{Var}(\tilde{u}_A)$ is given by

$$\begin{aligned} \text{Var}(\tilde{u}_A) &= \frac{\frac{\text{Var}(t)}{\sigma_T^4} + \frac{\text{Var}(s)}{\sigma_S^4} + 2\frac{\text{Cov}(s, t)}{\sigma_T^2\sigma_S^2}}{\left(\frac{1}{\sigma_Q^2} + \frac{1}{\sigma_T^2} + \frac{1}{\sigma_S^2}\right)^2} + \sigma_B^2 = \frac{\frac{\sigma_Q^2 + \sigma_T^2}{\sigma_T^4} + \frac{\sigma_Q^2 + \sigma_S^2}{\sigma_S^4} + 2\frac{\sigma_Q^2}{\sigma_T^2\sigma_S^2}}{\left(\frac{1}{\sigma_Q^2} + \frac{1}{\sigma_T^2} + \frac{1}{\sigma_S^2}\right)^2} + \sigma_B^2 \\ &= \sigma_Q^2 \frac{\left(\frac{1}{\sigma_T^2} + \frac{1}{\sigma_S^2} + \frac{1}{\sigma_Q^2}\right) \left(\frac{1}{\sigma_T^2} + \frac{1}{\sigma_S^2}\right)}{\left(\frac{1}{\sigma_Q^2} + \frac{1}{\sigma_T^2} + \frac{1}{\sigma_S^2}\right)^2} + \sigma_B^2 = \sigma_Q^2 \frac{\frac{1}{\sigma_T^2} + \frac{1}{\sigma_S^2}}{\frac{1}{\sigma_Q^2} + \frac{1}{\sigma_T^2} + \frac{1}{\sigma_S^2}} + \sigma_B^2 \\ &= \frac{\sigma_Q^4(\sigma_S^2 + \sigma_T^2)}{\sigma_S^2\sigma_T^2 + \sigma_T^2\sigma_Q^2 + \sigma_Q^2\sigma_S^2} + \sigma_B^2. \end{aligned} \quad \square$$

F.2 Proofs for Section 4

Proof of Lemma 2. Rewriting (11), (14), and (20),

$$\begin{aligned}\sigma_{U_A} &= \sqrt{\eta + \sigma_B^2} \\ \sigma_{U_P} &= \frac{\eta}{\sqrt{\eta + \sigma_B^2}} \\ \text{with } \eta &= \frac{\sigma_T^4 \sigma_Q^4}{(\sigma_Q^2 + \sigma_T^2)(\sigma_T^2 \sigma_Q^2 + \sigma_S^2 \sigma_T^2 + \sigma_S^2 \sigma_Q^2)}\end{aligned}$$

1. First, observe that η decreases in σ_S^2 . Now, fixing σ_B^2 , we see that σ_{U_A} and σ_{U_P} both increase in η .
2. Fixing σ_S^2 , η is constant in σ_B^2 . We see that σ_{U_A} increases in σ_B^2 while σ_{U_P} decreases in σ_B^2 .
3. The value η from is independent of σ_B^2 , and decreases from $\frac{\sigma_T^2 \sigma_Q^2}{\sigma_Q^2 + \sigma_T^2}$ to zero as σ_S^2 increases from zero to infinity. From Parts 1 and 2, we maximize σ_{U_P} at a given $\tilde{\sigma}_{U_A}$ by finding the mixture of the lowest σ_S^2 (most information) and the lowest σ_B^2 (least bias) consistent with $\tilde{\sigma}_{U_A}$. For $\sigma_{U_A} \leq \frac{\sigma_T \sigma_Q}{\sqrt{\sigma_Q^2 + \sigma_T^2}}$, we can get $\tilde{\sigma}_{U_A}$ with $\sigma_S^2 > 0$ and $\sigma_B^2 \rightarrow 0$, and $\sigma_S^2 \rightarrow 0$ implies $\sigma_{U_P} \rightarrow \tilde{\sigma}_{U_A}$. For larger $\tilde{\sigma}_{U_A}$, we take $\sigma_B^2 > 0$ and $\sigma_S^2 \rightarrow 0$, and with $\sigma_S^2 \rightarrow 0$ we have $\sigma_{U_P} \rightarrow \frac{\sigma_T^2 \sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} \frac{1}{\tilde{\sigma}_{U_A}}$. \square

Proof of Lemma 3.

Step 1. Let us start by deriving the “upper-right frontier” of (τ, ζ) for a given k . These are the pairs that maximize $p\tau + (1-p)\zeta$ for some $p \in [0, 1]$.

Fixing $p \in [0, 1]$, we maximize $p\tau + (1-p)\zeta$ by selecting the k applicants with the highest values of $pt^i + (1-p)z^i$. (These are exactly the applicants above a line in (t, \tilde{u}_A) -space – accepting such applicants induces a normal CDF acceptance rate.) Accepting such applicants In the population, t^i and z^i are independently normally distributed with means of 0, and respective variances of $\sigma_Q^2 + \sigma_T^2$ and 1. Therefore $pt^i + (1-p)z^i$ 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 $pt^i + (1-p)z^i$ are those with $\frac{pt^i + (1-p)z^i}{\sigma_{\text{comb}}} > r^*$, for r^* satisfying $\Phi(r^*) = 1 - k$. I seek to calculate the expected value of t^i and of z^i among applicants with $\frac{pt^i + (1-p)z^i}{\sigma_{\text{comb}}} > r^*$.

We have the following joint normal distribution among the three variables of t^i , z^i , and

$$\frac{pt^i + (1-p)z^i}{\sigma_{\text{comb}}}.$$

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

As in the expression (48) of Appendix F.1, we can calculate conditional means of t^i and z^i given a realization of $\frac{pt^i + (1-p)z^i}{\sigma_{\text{comb}}}$:

$$\begin{aligned} \mathbb{E}[t^i | \frac{pt^i + (1-p)z^i}{\sigma_{\text{comb}}} = r] &= \frac{p(\sigma_Q^2 + \sigma_T^2)}{\sigma_{\text{comb}}} r \\ \mathbb{E}[z^i | \frac{pt^i + (1-p)z^i}{\sigma_{\text{comb}}} = r] &= \frac{1-p}{\sigma_{\text{comb}}} r \end{aligned}$$

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

$$\begin{aligned} \mathbb{E}[t^i | \frac{pt^i + (1-p)z^i}{\sigma_{\text{comb}}} > r] &= \frac{p(\sigma_Q^2 + \sigma_T^2)}{\sigma_{\text{comb}}} \mathbb{E}[\frac{pt^i + (1-p)z^i}{\sigma_{\text{comb}}} | \frac{pt^i + (1-p)z^i}{\sigma_{\text{comb}}} > r] \\ \mathbb{E}[z^i | \frac{pt^i + (1-p)z^i}{\sigma_{\text{comb}}} > r] &= \frac{1-p}{\sigma_{\text{comb}}} \mathbb{E}[\frac{pt^i + (1-p)z^i}{\sigma_{\text{comb}}} | \frac{pt^i + (1-p)z^i}{\sigma_{\text{comb}}} > r] \end{aligned}$$

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

$$\begin{aligned} \tau &= \mathbb{E}[t^i | \frac{pt^i + (1-p)z^i}{\sigma_{\text{comb}}} > r^*] = \frac{p(\sigma_Q^2 + \sigma_T^2)}{\sigma_{\text{comb}}} \frac{\phi(r^*)}{1-\Phi(r^*)} = \frac{p(\sigma_Q^2 + \sigma_T^2)}{\sqrt{p^2(\sigma_Q^2 + \sigma_T^2) + (1-p)^2}} R(k) \\ \zeta &= \mathbb{E}[z^i | \frac{pt^i + (1-p)z^i}{\sigma_{\text{comb}}} > r^*] = \frac{1-p}{\sigma_{\text{comb}}} \frac{\phi(r^*)}{1-\Phi(r^*)} = \frac{1-p}{\sqrt{p^2(\sigma_Q^2 + \sigma_T^2) + (1-p)^2}} R(k) \end{aligned}$$

As p goes from 0 to 1, τ goes from 0 to $\sqrt{\sigma_Q^2 + \sigma_T^2} R(k)$ and ζ goes from $R(k)$ to 0.

Writing $R_Z = R(k)$ and $R_T = \sqrt{\sigma_Q^2 + \sigma_T^2} R(k)$, 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 on the upper-right frontier, we trace out the 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 – 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. \square

Proof of Lemma 4. 1. Suppose that for some agent, and some realization of t^i, s^i , and b^i across all applicants i , the agent can send a message that yields an applicant pool with average test score and average z-score of (τ, ζ) . By anonymity, this agent could have sent a different message to achieve the same (τ, ζ) at any other realizations of t^i, s^i , and b^i across applicants – the joint distribution would be the same. In particular, if the agent’s ranking of applicants at each test score had been permuted, she could have sent a message that permuted the acceptances in the same manner. Moreover, the ranking at a given test score is fully determined by the z-score z^i of each applicant, which follows a standard normal distribution at each test score.

Now consider some other agent, possibly of a different type. While the distribution of s^i and b^i at each test score is different if the agent has a different type, the distribution of z^i still follows a standard normal. So this new agent can send a message as if she were the first agent, but with her own ranking of applicants at each test score. This message yields the same (τ, ζ) as above.

2. At a given message, and under a given realization of t^i, z^i across applicants, say that an agent’s message induces average test score and average z-score (τ', ζ') . The same message could have been sent under any permutation of all of the z^i 's. For instance, if the agent’s ranking of applicants at each test score had been flipped – if each z^i were multiplied by -1 – then the same message would yield $(\tau', -\zeta')$. And by anonymity, that means that there is some message under the true realizations of z^i which would have resulted in the same outcome of $(\tau', -\zeta')$.

The same logic shows that $(\tau', x\zeta')$ is achievable for any $x \in [-1, 1]$. At a given message,

some permutation of z^i 's at each test score results in a multiple of x on the average z-score.⁴⁰ That means that at the original realizations of z^i , a different message would have given $(\tau', x\zeta')$.

3. It suffices to show that a contract can specify any subset of feasible τ in $[-R_T, R_T]$; and for each such τ , can specify a maximum possible ζ in $[0, R_Z\sqrt{1 - \frac{(\tau)^2}{R_T^2}}]$. Given the maximum ζ of ζ' at some $\tau = \tau'$, we know from the previous part that the agent can also induce an average test-score and average z-score of $(\tau', x\zeta')$ for any $x \in [-1, 1]$.

First, fix some $\tau' \in [-R_T, R_T]$ and some $\zeta' \in [0, R_Z\sqrt{1 - \frac{(\tau')^2}{R_T^2}}]$. I want to find a contract in which the outcome must have an average test score of τ' , and in which the maximum possible average z-score across an agent's reports is ζ' . Call this the subcontract at τ' . Once we have found the subcontracts at each allowed τ' , we can then set the contract to be the union of all such subcontracts – the agent selects one of the subcontract from a menu, and then plays the induced subgame.

If $\zeta' = 0$, we can take the τ' subcontract to be the one which randomly selects any k applicants with this average test score without the agent's input.⁴¹ If $\zeta' = R_Z\sqrt{1 - \frac{(\tau')^2}{R_T^2}}$ – the maximum achievable average z-score at this average test score – we can set the τ' subcontract in one of two ways. First, we can say that the agent may select any k applicants subject to this average test-score; she will then pick her favorite set, maximizing the average z-score by taking it to ζ' . Alternatively, we can specify the acceptance rate as the appropriate normal CDF acceptance rate which accepts k applicants with this average test score. For $\zeta' \in (0, R_Z\sqrt{1 - \frac{(\tau')^2}{R_T^2}})$, we can simply randomize between the $\zeta' = 0$ and $\zeta' = R_Z\sqrt{1 - \frac{(\tau')^2}{R_T^2}}$ subcontracts above with appropriate probabilities.⁴² \square

⁴⁰For instance, reversing z-scores at all $z^i \in [-y, y]$ for $y \geq 0$ smoothly takes us from $x = 1$ at $y = 0$ to $x = -1$ as $y \rightarrow \infty$.

⁴¹This would be a stochastic contract. There is also a deterministic implementation: require that all applicants with t^i in some interval be accepted, and none outside the interval. To get $\tau' = R_T$ (the highest possible) with $\zeta' = 0$, take the interval to be the highest k quantiles of test scores – $[t, \infty)$ for t satisfying $\Phi(\frac{t}{\sqrt{\sigma_Q^2 + \sigma_T^2}}) = 1 - k$. To get the $\tau' = -R_T$ (the lowest possible) with $\zeta' = 0$, take the interval to be the lowest k quantiles of test scores. As we shift the interval and accept applicants with test score quantiles in $[x, x + k]$ for $x \in [0, 1 - k]$, we trace out average test scores from $-R_T$ to R_T while maintaining an average z-score of 0.

⁴²Once again, this is a stochastic contract, but there also exists a deterministic implementation. At τ' there is a deterministic acceptance rate function corresponding to $\zeta' = 0$ from footnote 42, and a deterministic acceptance rate function corresponding to $\zeta' = R_Z\sqrt{1 - \frac{(\tau')^2}{R_T^2}}$ as the appropriate normal CDF. Without going into details, we can parametrize a smooth path of deterministic acceptance rate functions which goes from the $\zeta' = 0$ function to the $\zeta' = R_Z\sqrt{1 - \frac{(\tau')^2}{R_T^2}}$ function, such that at each acceptance rate along the path (i) the average test score remains at τ' , and (ii) we continue to accept k applicants. The average z-score

Proof of Proposition 5. I will show this result as an application of the one-dimensional delegation problem in Amador and Bagwell (2013), specifically citing Proposition 1(b) of that paper. In this problem, the “state” is $\tilde{\sigma}_{U_A}$, the contractible “action” is $\zeta \in [0, R_Z]$, and the level of joint “money-burning” is δ .⁴³ Employing some of the notation of Amador and Bagwell (2013), let $w(\tilde{\sigma}_{U_A}, \zeta)$ be the principal’s realized payoff over the state and action, prior to money-burning:

$$w(\tilde{\sigma}_{U_A}, \zeta) \equiv \hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A}) \cdot \zeta + b(\zeta), \text{ for } b(\zeta) \equiv \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} R_T \cdot \sqrt{1 - \frac{\zeta^2}{R_Z^2}}. \quad (49)$$

The agent’s payoff prior to money-burning can be written as

$$\tilde{\sigma}_{U_A} \cdot \zeta + b(\zeta). \quad (50)$$

We can now verify the regularity conditions of Assumption 1 of Amador and Bagwell (2013). 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. Condition (iv) is somewhat more involved; first, we solve for the function $\zeta^*(\tilde{\sigma}_{U_A})$ (denoted π_f in Amador and Bagwell (2013)) which is the agent’s ideal action ζ in state $\tilde{\sigma}_{U_A}$. Taking the first order condition of (50), we can solve for this function as

$$\zeta^*(\tilde{\sigma}_{U_A}) = \frac{\tilde{\sigma}_{U_A} R_Z}{\sqrt{\tilde{\sigma}_{U_A}^2 + \frac{\sigma_Q^4}{(\sigma_Q^2 + \sigma_T^2)^2} \frac{R_T^2}{R_Z^2}}}. \quad (51)$$

The function ζ^* is smooth and increasing, satisfying the condition. Finally, for (v) we take the derivative of w with respect to ζ to get

$$w_\zeta(\tilde{\sigma}_{U_A}, \zeta) = \hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A}) - \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} R_T \frac{\zeta}{\sqrt{1 - \frac{\zeta^2}{R_Z^2}}} \quad (52)$$

ζ would move smoothly along this path as well, and so hits any $\zeta \in [0, R_Z \sqrt{1 - \frac{(\tau')^2}{R_T^2}}]$ by the intermediate value theorem.

⁴³The notation of Amador and Bagwell (2013) has state γ , action π , and money-burning t . Their state is distributed according to F , while in this paper the state $\tilde{\sigma}_{U_A}$ is distributed according to H . In the problem of Amador and Bagwell (2013), arbitrary money-burning $t \geq 0$ is allowed, and they find conditions under which no money-burning is used. Here the possible money-burning δ is bounded at $2\tau(\zeta)$; but a condition that guarantees that no money-burning is used when any $\delta \geq 0$ can be implemented will also guarantee that no money-burning is used when we have bounds on δ .

and observe that it is continuous. We have now verified Assumption 1.

The next step is to calculate a few quantities. First, we have $\kappa \equiv \min\{\inf_{\tilde{\sigma}_{U_A}, \zeta} \frac{w_{\zeta\zeta}(\tilde{\sigma}_{U_A}, \zeta)}{b''(\zeta)}, 1\} = 1$ because $w_{\zeta\zeta}$ (indicating the second derivative) is identically equal to b'' . Second, let us evaluate w_{ζ} at the agent's ideal point from (51). We could plug in (51) into (52) and simplify, but it is easier to observe that the agent's ideal point $\zeta^*(\tilde{\sigma}_{U_A})$ is derived from the first order condition $b'(\zeta^*(\tilde{\sigma}_{U_A})) = -\tilde{\sigma}_{U_A}$, and $w_{\zeta}(\tilde{\sigma}_{U_A}, \zeta) = \hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A}) + b'(\zeta)$. Therefore,

$$w_{\zeta}(\tilde{\sigma}_{U_A}, \zeta^*(\tilde{\sigma}_{U_A})) = \hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A}) - \tilde{\sigma}_{U_A}.$$

Next we confirm that the condition (c1) of Amador and Bagwell (2013) holds globally across all $\zeta \in [0, R_Z]$. Global (c1) is sufficient to imply (from their addendum) that the optimal delegation set will be of the interval form. [Details to be added.](#) Condition (c1) states that

$$\kappa H(\tilde{\sigma}_{U_A}) - w_{\zeta}(\tilde{\sigma}_{U_A}, \zeta^*(\tilde{\sigma}_{U_A}))h(\tilde{\sigma}_{U_A})$$

is nondecreasing. Plugging in κ and $w_{\zeta}(\tilde{\sigma}_{U_A}, \zeta^*(\tilde{\sigma}_{U_A}))$ gives the condition in the statement of the proposition.⁴⁴

The last step is to confirm that the interval in question is of the form of a cap (ceiling) on ζ , corresponding to a floor on τ – that is, that we never require the agent to take an action ζ above her ideal point ζ^* . This holds if their condition (c3') holds, that $w_{\zeta}(\tilde{\sigma}_{U_A}, \zeta^*(\tilde{\sigma}_{U_A})) \leq 0$ for $\tilde{\sigma}_{U_A}$ equal to the minimum of the support. Indeed, $w_{\zeta}(\tilde{\sigma}_{U_A}, \zeta^*(\tilde{\sigma}_{U_A})) = \hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A}) - \tilde{\sigma}_{U_A} \leq 0$ for all $\tilde{\sigma}_{U_A}$. \square

Proof of Proposition 6. Follows immediately from Proposition 10, below. \square

Proposition 10. Suppose that, for $\tilde{\sigma}_{U_A}$ in the support of H , the maximum of $\hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A})$ is less than or equal to the minimum of $\tilde{\sigma}_{U_A}$. Then the optimal contract can be characterized in either of the following two ways:

1. The agent is given a specified average test score. She may select any k applicants she wants, subject to the average test score of hired applicants τ being equal to some specified level $K > 0$.

⁴⁴The sufficient conditions for interval delegation to be optimal come from Amador and Bagwell (2013) Proposition 1(b); in my case, because of the form of the agent's bias, any interval delegation set will take the form of a floor on τ (ceiling on ζ). That paper also gives necessary conditions for interval delegation to be optimal in their Proposition 2(b). Because my problem satisfies the assumptions of Proposition 2(b), one can see that the sufficient conditions I provide are very close to being necessary; essentially, the wiggle room is that $H(\tilde{\sigma}_{U_A}) + (\tilde{\sigma}_{U_A} - \hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A}))h(\tilde{\sigma}_{U_A})$ might not need to be globally nondecreasing, but only nondecreasing over the states in which that agent takes actions above the floor.

2. The agent is allowed to choose any k applicants she wants as long as the induced acceptance rate $\alpha(t)$ is $\Phi(\gamma_T t - \gamma_0)$, with γ_T equal to some specified level $\Gamma > 0$.

Lemma 9. Suppose that (i) the distribution H has bounded support with 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.⁴⁵ Then the hypothesis of Proposition 5 is satisfied: $H(\tilde{\sigma}_{U_A}) + (\tilde{\sigma}_{U_A} - \hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A}))h(\tilde{\sigma}_{U_A})$ is nondecreasing.

Proof of Lemma 5. It is sufficient to confirm that $\hat{\sigma}'_{U_P}(\tilde{\sigma}_{U_A}) \leq 2$ when the bias σ_B^2 is commonly known, in which case Lemma 9 above implies the result.

From (14) and (20),

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

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 (14), which states that for any agent type (σ_S^2, σ_B^2) it holds that $\sigma_{U_A}^2 = \eta + \sigma_B^2$, with $\eta > 0$. \square

Proof of Proposition 10. Step 1. Take $\tilde{\sigma}_{U_A}^1 < \tilde{\sigma}_{U_A}^2$, and take points (τ^1, ζ^1) and (τ^2, ζ^2) in \overline{W} . Suppose an agent of type $\tilde{\sigma}_{U_A}^1$ weakly prefers (τ^1, γ^1) to (τ^2, γ^2) , and an agent of type $\tilde{\sigma}_{U_A}^2$ weakly prefers (τ^2, γ^2) to (τ^1, γ^1) . I claim that if $\hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A}) \leq \tilde{\sigma}_{U_A}^1$, then under type $\tilde{\sigma}_{U_A}$ the principal weakly prefers (τ^1, γ^1) to (τ^2, γ^2) .

This claim follows as a straightforward single-crossing argument from (27) and (28). From the agent's two choices, it must be that $\tau^1 \geq \tau^2$ and $\zeta^1 \leq \zeta^2$. Now, writing out the agent's choice at type $\tilde{\sigma}_{U_A}^1$, it holds that

$$\frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} \cdot \tau^1 + \tilde{\sigma}_{U_A}^1 \cdot \zeta^1 \geq \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} \cdot \tau^2 + \tilde{\sigma}_{U_A}^1 \cdot \zeta^2.$$

Because $\zeta^1 \leq \zeta^2$, the latter inequality holds when $\tilde{\sigma}_{U_A}^1$ is replaced by $\hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A}) \leq \tilde{\sigma}_{U_A}^1$.

Step 2. By the above claim, under any contract, the principal prefers the (τ, ζ) pair chosen by the agent with type $\tilde{\sigma}_{U_A}$ equal to the minimum of the support 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

⁴⁵For $\hat{\sigma}_{U_P}$ differentiable, assumption (iii) amounts to $\hat{\sigma}'_{U_P}(\tilde{\sigma}_{U_A}) \leq 2$.

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. \square

Proof of Lemma 9. 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'):

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

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,

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

The first bracketed term is weakly positive by (ii), and the second and third bracketed terms cancel each other out. Therefore the expression is in fact weakly positive.⁴⁶ \square

F.3 Additional Appendix Proofs

Proof of Lemma 6. Follows from arguments in the text. \square

⁴⁶With h and Δ differentiable, one could have more straightforwardly found the derivative of the LHS to be $h(\tilde{\sigma}_{U_A}) \cdot (1 + \Delta'(\tilde{\sigma}_{U_A})) + \Delta(\tilde{\sigma}_{U_A}) \cdot h'(\tilde{\sigma}_{U_A})$. Assumption (iii) implies that $\Delta'(\tilde{\sigma}_{U_A}) \geq -1$, and therefore that the expression is weakly positive.

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

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

1. The parameter k does not appear in the formula for γ_T^{FD} .
2. The parameter σ_S^2 appears only in γ_T^{FD} through η . Routine differentiation shows that $\frac{d\gamma_T^{\text{FD}}}{d\eta} < 0$ and $\frac{d\eta}{d\sigma_S^2} < 0$, and so by the chain rule $\frac{d\gamma_T^{\text{FD}}}{d\sigma_S^2} > 0$.

Taking limits,

$$\begin{aligned}\lim_{\sigma_S^2 \rightarrow 0} \gamma_T^{\text{FD}} &= \frac{\sigma_Q^2}{(\sigma_Q^2 + \sigma_T^2)\sqrt{\frac{\sigma_Q^2 \sigma_T^2}{\sigma_Q^2 + \sigma_T^2} + \sigma_B^2}}, \text{ because } \lim_{\sigma_S^2 \rightarrow 0} \eta = \frac{\sigma_T^2 \sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} \\ \lim_{\sigma_S^2 \rightarrow \infty} \gamma_T^{\text{FD}} &= \frac{\sigma_Q^2}{\sigma_B(\sigma_Q^2 + \sigma_T^2)}, \text{ because } \lim_{\sigma_S^2 \rightarrow \infty} \eta = 0.\end{aligned}$$

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

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

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 $\gamma_T^{\text{FD}} < \gamma_T^*$.

3. 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,

$$\lim_{\sigma_B^2 \rightarrow 0} \gamma_{\text{FD}}^* = \frac{\sigma_Q^2}{(\sigma_Q^2 + \sigma_T^2)\sqrt{\eta}}$$

$$\lim_{\sigma_B^2 \rightarrow \infty} \gamma_T^{\text{FD}} = 0.$$

From the proof of Proposition 4, we see that $\lim_{\sigma_B^2 \rightarrow 0} \gamma_{\text{FD}}^* = \lim_{\sigma_B^2 \rightarrow 0} \gamma_T^*$.

4. 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 $\frac{d(\gamma_T^{\text{FD}} \sqrt{\sigma_Q^2 + \sigma_T^2})}{d\sigma_T^2}$ as $\frac{\partial(\gamma_T^{\text{FD}} \sqrt{\sigma_Q^2 + \sigma_T^2})}{\partial \sigma_T^2} + \frac{\partial(\gamma_T^{\text{FD}} \sqrt{\sigma_Q^2 + \sigma_T^2})}{\partial \eta} \frac{d\eta}{d\sigma_T^2}$. Once again, $\frac{\partial(\gamma_T^{\text{FD}} \sqrt{\sigma_Q^2 + \sigma_T^2})}{\partial \sigma_T^2} < 0$, $\frac{\partial(\gamma_T^{\text{FD}} \sqrt{\sigma_Q^2 + \sigma_T^2})}{\partial \eta} < 0$, and $\frac{d\eta}{d\sigma_T^2} > 0$. Therefore $\frac{d(\gamma_T^{\text{FD}} \sqrt{\sigma_Q^2 + \sigma_T^2})}{d\sigma_T^2} < 0$.

Taking limits,

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

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

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

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

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

$$\begin{aligned} \lim_{\sigma_Q^2 \rightarrow \infty} \gamma_T^{\text{FD}} &= \frac{1}{\sqrt{\sigma_B^2 + \frac{\sigma_T^4}{\sigma_T^2 + \sigma_S^2}}} \\ \lim_{\sigma_Q^2 \rightarrow \infty} \gamma_T^{\text{FD}}\sqrt{\sigma_Q^2 + \sigma_T^2} &= \infty. \end{aligned}$$

We can rewrite $\lim_{\sigma_Q^2 \rightarrow \infty} \gamma_T^{\text{FD}}$ 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 $\lim_{\sigma_Q^2 \rightarrow \infty} \gamma_T^* = \frac{(\sigma_S^2 + \sigma_T^2)\sqrt{\sigma_B^2 + \frac{\sigma_T^4}{\sigma_T^2 + \sigma_S^2}}}{\sigma_T^4}$ from the proof of Proposition 4 part 5. \square

Proof of Proposition 8. Follows from arguments in the text. \square

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

Proof of Lemma 7. 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 $R(k)$ in (24). 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 (23). And $\sigma_{U_P}(\theta)$ must be in the range $(0, \sigma_{U_A}(\theta))$ from Lemma 2 part 3.⁴⁷ 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 (22) 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)

$$\frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} \tau_1 < \xi_1 < \frac{\sigma_Q^2 R(k)}{\tau_1}. \quad (53)$$

Now, return to the comparative statics of τ^* given by (44). τ^* moves in the opposite direction as the fraction $\frac{(\xi_1(\sigma_Q^2 + \sigma_T^2) - \sigma_Q^2 \tau_1)^2}{(\sigma_Q^2 + \sigma_T^2)R(k)^2 - \tau_1^2}$ 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 (53). 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 (53). \square

⁴⁷The upper bound here need not be tight, depending on parameters.