Selecting Applicants

Alex Frankel*

April 2, 2020

Abstract

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

KEYWORDS: Principal-agent contracting, Delegation, Hiring

1 Introduction

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

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

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

If the manager were given full discretion to hire any applicants she wanted, her choices would be distorted by these biases. The firm might instead require the manager to hire those with the most favorable hard information. This no-discretion policy would inevitably screen out some "diamonds in the rough," high quality candidates who showed their worth only on the softer measures. But, as pointed out by Hoffman et al. (2018), no discretion would improve on full discretion if the manager's biases were sufficiently strong.

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

Spelling out some key features of the model, there is a large number (continuum) of applicants, of whom some share will be accepted. For each applicant there is public or hard information about quality as well as a private or soft signal observed only by an agent. The agent also has some bias in favor of or against each applicant. The principal writes a contract to determine which applicants the agent may accept. The principal's objective is to maximize the average quality of the selected applicants, whereas the agent cares about quality plus bias.

I begin my analysis by supposing that the firm knows the manager's "type" – the distribution of her information and bias across applicants. Here, the optimal policy takes the form of a specified acceptance rate at each realization of hard information, e.g., at each test result. The manager chooses which applicants to hire subject to this acceptance rate. I provide a general approach for finding the optimal acceptance rate function, essentially by equalizing the quality of the marginal accepted applicant across test results. I then apply the results to a benchmark normal specification: normal distribution of applicant quality, one-dimensional normal signals of quality from both hard and soft sources, and normally distributed idiosyncratic biases. Under the normal specification, the acceptance rate should follow a normal CDF function –

an S-curve, in which a higher share of applicants are accepted at higher test results. A manager with either stronger biases or less information faces a steeper acceptance rate function, corresponding to a contract with less discretion: her hiring depends more on the test and less on her personal judgment.

There is an alternative implementation of these optimal contracts. Realizations of hard information are first mapped into a one-dimensional score. Then the contract specifies that the manager may accept any applicants she wants, subject to a minimum average score of those who are accepted. (This floor will bind.) Under the normal specification, the score can be set equal to the test result: the manager can accept any set of applicants whose average test score is high enough. A higher minimum average score would imply less discretion for the manager.

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 screens 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 (not necessarily binding) minimum average test score. These floors bind only for more informed and/or more biased managers.

The bulk of the work on the delegation of decisions by a principal to a better-informed but biased agent involves a single one-dimensional decision to be made. See, for example, Holmström (1977, 1984) and Melumad and Shibano (1991) for early work, or Alonso and Matouschek (2008) and Amador and Bagwell (2013) more recently. Papers on delegation or cheap talk – commitment or no commitment – over multiple decisions include Chakraborty and Harbaugh (2007), Frankel (2014), and Frankel (2016). What is novel in the current paper, both in terms of the tradeoffs it generates and in the contracting levers it makes available, is the existence of distinct realizations of hard information across decisions.

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

The problem of combining information from hard and soft sources is becoming increasingly relevant as "big data" and the spread of IT supplement traditional subjective evaluations with newly available, or newly quantifiable, hard information. According to the Wall Street Journal (2015), for instance, the number of US employers using pre-employment tests rose from 26% in 2001 to 57% in 2013.[1] Through a mix of theory and data, Autor and Scarborough (2008) and Hoffman et al. (2018) shed light on how pre-employment tests have affected firm hiring, while Einav et al. (2013) looks at the impact of automated credit scoring in the market for consumer loans. In particular, Hoffman et al. (2018) addresses the question of how much discretion to grant to a potentially biased agent. That paper compares full discretion, the policy used by the firm in their data set, to a hypothetical policy of no discretion.

In related theoretical work, Prendergast and Topel (1996) studies the payment and promotion of employees by a firm that receives hard information on performance as well as soft information from a manager who distorts her reports due to favoritism, i.e., a desire to pay employees more. Che et al. (2013) looks specifically at the role of hard and soft information in hiring, where hard information is modeled via asymmetric priors on applicants' quality.[2] Unlike the current paper, Che et al. (2013) considers

[1] The article describes these tests as follows:

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

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

the hiring of a single worker rather than many. A more important distinction is in the modeling of the agent's bias: their misalignment is over *how many* rather than *which* candidates to hire. The agent in that paper has a bias towards hiring someone rather than leaving the position unfilled. I fix the (large) number of candidates to be hired, but take the agent's preferences over candidates to be imperfectly correlated with those of the principal.[3]

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

2 The model

2.1 Players, payoffs, and information

There is a firm (principal), a manager (agent), and a mass 1 of ex ante identical job applicants. An exogenous fraction $k \in (0, 1)$ of the applicants will be hired; I sometimes refer to hired applicants as being accepted or selected. The firm will specify rules determining the process by which the manager makes hiring decisions.

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

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

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

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

[4] If the agent were required to put in a costly investment in order to observe her private signal,

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

The final applicant characteristic is the bias, B. Assume that the distribution of B conditional on T and S is independent of Q, and thus that its conditional distribution can be denoted by $F_{B|Q,T,S} = F_{B|T,S}$. In other words, the bias contains no information on quality beyond what is captured by the hard and soft information.

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

Under a given acceptance rule, we can define a random variable Hired $\in \{0, 1\}$ that describes whether a randomly drawn applicant is hired (1) or not (0). Conditional on Q, T, S, and B, Hired takes value 1 with probability $\chi(T, S, B)$. The principal's and agent's realized payoffs, V_P and V_A, are their average utilities over the hired applicants:[6]

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

Players make decisions in order to maximize expected payoffs. Assume that the expectations of Q and $Q + B$ are finite so that these payoffs are as well.

We see that the model primitives consist of one scalar parameter and four dis-

hard information might be used as a first screen to decide which applicants to evaluate privately. The Wall Street Journal (2014) reports that at many companies, job applicants who do poorly on pre-employment tests will not have their resumes looked at or will not get interviews.

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

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

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

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

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

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

Recall that there is a continuum of applicants, where F_Q and $F_{T|Q}$ are the distributions of Q and $T|Q$. So common knowledge of F_Q establishes that there is no aggregate uncertainty over the distribution of quality in the applicant pool. Common knowledge of $F_{T|Q}$ further implies that there is no aggregate uncertainty over the distribution of test results, nor over the conditional distribution of applicant quality at each test result.

I also now impose a technical assumption that will simplify notation going forward. Recall that, given the agent's observation of the public and private signals, her expectation of an applicant's quality is $\mathbb{E}[Q|T,S]$. Denote her corresponding expectation of her own utility from hiring an applicant by U_A:

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

Assumption 2. For each $t \in \mathcal{T}$, the distribution of $U_A|T = t$ is atomless.

In other words, the agent almost surely has a strict preference between any two applicants with the same test result. This could arise from continuously distributed biases or from a continuously distributed belief on quality arising from the agent's private signals. The assumption will ensure that I do not have to worry about disparate treatment within a mass of applicants at $T = t$ who share the same (S, B) or, more generally, the same U_A.

2.2 Contracting

The principal writes a contract that specifies the rules by which the agent selects applicants. The principal can commit to accept any applicants that the agent selects, and the agent participates in whatever contract the principal writes. The only outcome of this contracting relationship is the determination of which applicants are accepted. That is, there are no monetary transfers (or other extrinsic incentives) that condition on the agent's behavior or on the quality of accepted applicants.

All of the applicants' test results are public. However, the contracting mechanism can only condition on the agent's type or her applicant-specific private signals and biases through her reports. As discussed below, it is without loss of generality to consider a mechanism in which the agent simply reports the probability with which each applicant is to be accepted.[7] The contract will therefore be a restriction on the joint distribution of observable test results and the reported acceptance probabilities.[8] The formal contracting game is as follows.

(1) **The agent's type is realized.**

(2) **Applicants' test results are publicly observed.**

(3) **The principal gives the agent a contract.** The contract is a nonempty set of allowed joint distributions over test results and reported acceptance probabilities. Each element of this set must be feasible given the distribution of test results, and it must satisfy the budget constraint of accepting k applicants.

(4) **The agent observes her private signals and biases for each applicant.**

[7] I omit mention of the correlation structure over probabilistic acceptances because principal and agent payoffs are additive across applicants, so correlation is does not affect expected payoffs.

[8] All such restrictions on the agent's behavior, and any characterizations of the agent's best responses, should always be understood as holding almost surely.

(5) **The agent reports acceptance probabilities.** Formally, the agent chooses an acceptance rule χ mapping test results T, private signals S, and biases B into acceptance probabilities $\chi(T,S,B)$. The acceptance rule must induce a joint distribution over T and $\chi(T,S,B)$ that is allowed by the contract, i.e., is in the set of step 3.[9]

(6) **Applicants are accepted according to the reported probabilities.** Payoffs V_P and V_A are realized.

The principal chooses a contract to maximize his subjective expectation of $V_P = \mathbb{E}[Q|\text{Hired} = 1]$, taking into account his beliefs about the agent's type and behavior when predicting who will be hired (i.e., what the acceptance rule will be). The agent then makes reports to maximize $V_A = \mathbb{E}[Q + B|\text{Hired} = 1]$.

Subject to the budget constraint, the contract may impose arbitrary restrictions on the joint distribution over test results and acceptance probabilities. For instance, at test result $T = t$ the contract might require that anywhere between 10% and 50% of applicants are accepted; that exactly 20% are to be accepted with probability .4 each; or that acceptances at t depend on the share accepted at other test results.

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

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

Any two deterministic reports inducing the same acceptance rate function necessarily have the same joint distribution over test results and acceptance probabilities. So a contract either permits both of them or forbids both. Hence, a deterministic contract

[9] More generally, the model could allow the agent to map T, S, B into mixtures over acceptance probabilities – different acceptance probabilities for applicants at the same (t, s, b). Assumption 2 guarantees the agent would not mix even if allowed, so I simplify notation by ruling out mixtures.

Specifically, as discussed in Section 2.3, at each $T = t$ the agent will monotonically assign higher acceptance probabilities to applicants for whom (S, B) induces a higher U_A. The agent might only consider mixing at an atom of $U_A|T = t$, and Assumption 2 forbids these atoms. (Indeed, without Assumption 2, satisfying the contract could be infeasible without mixtures. That would occur, for instance, if the distribution of $(S, B)|T = t$ had an atom, but the contract required a continuous distribution of acceptance probabilities at $T = t$.)

can be fully characterized by a menu of permitted acceptance rate functions. The agent may then select any applicants (choose any acceptance rule) for which the induced acceptance rate function is an element of the menu.[10]

Now let us return to the issue of why it is sufficient to consider contracts of the form of steps (1) - (6) above, rather than working with direct revelation contracts. A direct revelation contract would differ from the above in two ways. First, in between the agent's type realization at step (1) and the arrival of applicants at step (2) – call this step (1.5) – the agent would be asked to report her type. (This step could be omitted if the agent's type were common knowledge.) Second, at messaging step (5) she would report her private signal and bias for each applicant instead of an acceptance probability. These two reports would then map into acceptances.

The fact that the agent's direct reports at step (5) can be replaced with acceptance probabilities follows from the standard logic of delegation mechanisms: the agent knows how her reports will be translated into acceptances, so she may as well just report the acceptances.[11] The fact that the reporting of type at step (1.5) can be omitted, even when the agent's type is not commonly known, is a consequence of Assumption 1. Specifically, Assumption 1 ensures that once the agent knows her type, she faces no aggregate uncertainty about the future: she knows the joint distribution over T, S, B of the applicants who have yet to arrive. So if at step (5) the agent would ever be able to identify an ex-post profitable deviation to having previously misreported her type, the agent would have already had all of the information at step (1.5) to know that this deviation would be profitable. Hence, the principal has no reason not to delay all agent reports until step (5).

2.3 Preliminary Analysis

The agent's message in a contract assigns applicants to acceptance probabilities based on (T, S, B). At a given test result, the agent prefers to assign higher acceptance probabilities to applicants that give her higher expected utility $U_A = \mathbb{E}[Q|T, S] + B$. That is, at any test result $T = t$, the agent will report acceptance probabilities that

[10] Stochastic contracts may be more complicated. For instance, given acceptance rate function α, a stochastic contract could require that at $T = t$ the agent (deterministically) chooses $\alpha(t)$ applicants to accept with probability 1; or that every applicant at $T = t$ be accepted with probability $\alpha(t)$.

[11] If the principal could withhold test results from the agent, then the agent would not in fact know how her reports would translate into acceptances. I discuss how the principal might benefit from withholding test results in Section 5.3.

are weakly increasing – *monotonic* – in her expected utility U_A.

Combining monotonicity with Assumption 2, take any applicants with the same realizations of both T and U_A. Some may have high perceived quality $\mathbb{E}[Q|T,S]$ and low bias B, while others have low quality and high bias. But all such the applicants will (almost surely) be accepted with the same probability.[12] In other words, applicants with the same test result T and agent's expected utility U_A *cannot be distinguished* by any contract the principal offers. Formalizing the above discussion:

Observation 1. Fix a contract, a test result $t \in \mathcal{T}$, and two pairs of private signal and bias realizations (s^1, b^1) and (s^2, b^2) in $\mathcal{S} \times \mathbb{R}$. For each $i = 1, 2$, let the agent's expected utility U_A for an applicant with $T = t$, $S = s^i$, and $B = b^i$ be denoted by $u_A^i = \mathbb{E}[Q|T = t, S = s^i] + b^i$. Then under any equilibrium acceptance rule χ:

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

Now define $U_P(t, u_A)$ to be the principal's expected utility – that is, the average quality – of applicants with test result $T = t$ (observed by the principal) and agent's expected utility $U_A = u_A$ (unobserved by the principal):

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

Given distinguishability, we can rewrite acceptance rules as mappings from (T, U_A) – rather than (T, S, B) – into acceptance probabilities. As such, going forward, I take acceptance rules to be functions $\chi : \mathcal{T} \times \mathbb{R} \to [0, 1]$ where $\chi(t, u_A)$ indicates the probability of accepting an applicant with $T = t$ and $U_A = u_A$. The principal's payoff V_P under acceptance rule χ can now be written as

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

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

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

2.4 Examples of information and biases

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

2.4.1 Normal Specification

In the *normal specification*, assume that

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

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

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

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

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

This specification lets us capture the key forces of the model with a small number of parameters – one variance parameter for each distribution. The parameter σ_Q^2 is the variance of quality in the population, with mean normalized to 0. Then σ_T^2 and σ_S^2 describe how informative the public and private signals are about quality: variance going to 0 would be perfectly informative, and variance going to infinity would be uninformative. Finally, σ_B^2 tells us the strength of the agent's biases.[13] The agent's

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

utility for an applicant is $Q + B$, so an agent with higher σ_B^2 is more biased in that her utility comparisons across applicants depend less on variation in Q and more on variation in B. The agent's type here consists of the two parameters (σ_S^2, σ_B^2), the extent of her information and bias.

2.4.2 Two-Factor Model

In the *two-factor model*, let quality Q be decomposed as $Q = Q_1 + Q_2$. The two quality factors Q_1 and Q_2 follow some joint distribution F_{Q_1,Q_2}. The test result and private signal follow conditional distributions $F_{T|Q_1,Q_2}$ and $F_{S|Q_1,Q_2,T}$. Assume further that $\mathbb{E}[Q_1|S,T] = \mathbb{E}[Q_1|T]$ – in other words, the agent gets private information about the second quality factor but not the first. Think of the public test as measuring a job candidate's *t*echnical ability, while the private interview with a hiring manager yields information about the candidate's *s*ocial skills. Or, in college admissions, an applicant's public SAT score and GPA reveal his or her academic skills, whereas the admissions officer subjectively assesses other "holistic" aspects. In the same vein, the agent might be an expert brought in to evaluate applicants only on features related to her specialty: a writing instructor reads and scores the application essays.

Next, let the agent's objective be given by $V_A = \mathbb{E}[Q_1 + \lambda Q_2|\text{Hired} = 1]$ for some $\lambda > 0$. Compared to the principal's objective of $V_P = \mathbb{E}[Q_1 + Q_2|\text{Hired} = 1]$, there is a bias when $\lambda \neq 1$. We might expect that a hiring manager would overemphasize the importance of social skills, or that a writing instructor would overemphasize writing ability. This "advocate" agent, who values the factor that she evaluates more highly than does the principal, corresponds to $\lambda > 1$. The agent may also be a "cynic" with $\lambda < 1$: an interviewer who thinks that social skills don't matter much, or a writing instructor who thinks that writing ability is overrated.[14]

We can now rewrite the agent's payoff in the notation of Section 2.1, in which the agent's utility was given as $Q+B$. The agent's utility here is $Q_1+\lambda Q_2 = Q+(\lambda-1)Q_2$, and so the agent maximizes $\mathbb{E}[Q + B|\text{Hired} = 1]$ for $B = (\lambda - 1)\mathbb{E}[Q_2|T, S]$.[15] We see that the "systematic" bias manifests itself as a correlation between bias B and signals

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

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

T, S. An advocate with $\lambda > 1$, for instance, will be biased in favor of applicants for whom the signals reveal positive news on Q_2.

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

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

3 Common knowledge of agent type

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

It must hold that this acceptance rule χ selects k applicants, and that χ is monotonic in the agent's expected utility. Writing out these two necessary conditions:

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

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

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

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

3.1 Solving a relaxed problem

One upper bound on the principal's payoff would result from maximizing the objective (3) subject to the budget constraint (4), without imposing the monotonicity constraint (5). Denote this *upper bound acceptance rule*, or UBAR, as χ^{UBAR}. If UBAR satisfies (5) then it can be implemented as an optimal contract.

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

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

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

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

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

Proposition 1. *Under common knowledge of the agent's type, suppose that the upper bound acceptance rule is monotonic. Then the deterministic contract requiring the agent to select an applicant set satisfying the acceptance rate function α^{UBAR} implements the upper bound acceptance rule. This contract is an optimal contract.*

The contract of Proposition 1 separately specifies an acceptance rate of $\alpha^{\text{UBAR}}(t)$ at each possible test result $t \in \mathcal{T}$. There is another implementation – a contract yielding an identical outcome – that may in some cases be simpler to express. This alternative implementation first defines a "score function" mapping every realization of the (arbitrary-dimensional) hard information into a real number. The contract then asks the agent to select k applicants subject to the single restriction that the average score of those selected is sufficiently high.[16]

As a first step to defining such a score function, recall that under a monotonic and deterministic acceptance rule, an applicant is accepted if and only if U_A is above some test-result-specific cutoff. Let $u_A^c(t)$ be this cutoff at $T = t$ under UBAR: take $u_A^c(t) \in \mathbb{R} \cup \{-\infty, \infty\}$ such that for $t \in \mathcal{T}$ and u_A in the support of $U_A | T = t$, it holds that $\chi^{\text{UBAR}}(t, u_A) = 1$ if $u_A > u_A^c(t)$ and $\chi^{\text{UBAR}}(t, u_A) = 0$ if $u_A < u_A^c(t)$.[17] Heuristically, the function $u_A^c(t)$ can be thought of as a principal's indifference curve in (T, U_A)-space for which $U_P(t, u_A^c(t))$ is identically equal to u_P^c, with $U_P(t, u) \geq u_P^c$ for $u > u_A^c(t)$ and $U_P(t, u) \leq u_P^c$ for $u < u_A^c(t)$. Now let a *score function* $C : \mathcal{T} \to \mathbb{R} \cup \{-\infty, \infty\}$ be any negative affine transformation of u_A^c, i.e., $C(t) = a_0 - a_1 u_A^c(t)$ for some $a_1 > 0$. Given score function C and acceptance rule χ, the average score of hired applicants, $\mathbb{E}[C(T) | \text{Hired} = 1]$, can be expanded out as $\frac{1}{k}\mathbb{E}[\chi(T, U_A) C(T)]$.

Proposition 2. *Under common knowledge of the agent's type, suppose that the upper bound acceptance rule is monotonic. Let C be a score function, and suppose further that under the upper bound acceptance rule, the average score of hired applicants is finite and equal to $\kappa \in \mathbb{R}$. Then the contract requiring the agent to deterministically select any k applicants with an average score at or above κ implements the upper bound acceptance rule (in which the average score is exactly κ). This contract is an optimal contract.*

Consider the following Lagrangian intuition. Let λ_0 be the multiplier representing the shadow cost on the agent of hiring more than k applicants, and let λ_1 be the shadow cost of decreasing the average score of hired applicants below κ. The agent will hire an applicant if $U_A \geq \lambda_0 - \lambda_1 C(T)$. Now plug in score function $C(T) = a_0 - a_1 u_A^c(T)$ along with multipliers $\lambda_0 = a_0/a_1$ and $\lambda_1 = 1/a_1$. We see that the agent

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

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

hires an applicant if $U_A \geq u_A^c(T)$, which is the condition defining UBAR.

The role of taking the score function to be a *negative* affine transformation of u_A^c is to ensure that, on the margin, the agent prefers to *decrease* the average score below the floor κ. (With a positive affine transformation, the agent would prefer to increase the score, so the contract would impose a ceiling instead of a floor.) In the Lagrangian intuition, the fact that the agent prefers to decrease the average score below κ follows from the shadow cost $\lambda_1 = 1/a_1$ being positive.

In Section 5.1 I discuss how contracts analogous to those in Propositions 1 and 2 – an acceptance rate function and a minimum average score – could be implemented if there were finitely applicants rather than a continuum. I argue that in a finite economy, the minimum average score contract might be preferable.

Sufficient conditions for monotonicity of UBAR.

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

Loosely speaking, utilities are aligned up to distinguishability if at each test result, applicants who are more preferred by the agent are of higher average quality. Under alignment, at every test result the ordering of applicants by agent utility U_A is the same as by principal utility U_P.

Observation 2. If utilities are aligned up to distinguishability then the upper bound acceptance rule is monotonic.[18]

I confirm below that alignment up to distinguishability holds for the normal specification (Section 3.3), the two-factor model (Appendix D.1), and for a combination of the two (Appendix E). I can therefore apply Propositions 1 and 2 to characterize optimal contracts for each of these environments.

Note that there are joint distributions of Q, T, S, and B for which alignment is violated even when the bias B is independent of the soft information S: some high realization of U_A may indicate a very high bias combined with negative information on quality. That said, the following log-concavity condition on the bias distribution guarantees alignment up to distinguishability regardless of the quality and signal distributions.

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

Lemma 1. *Suppose that for all $t \in \mathcal{T}$, the bias distribution $F_{B|T=t,S}$ is independent of S and is log-concave.*[19] *Then utilities are aligned up to distinguishability.*

For instance, if the bias is independent of the private signal and is normally distributed – as it is in the normal specification – then utilities must be aligned up to distinguishability.

3.2 The general solution

In Appendix A, I solve for the optimal contract – maximizing (3) subject to (4) and (5) – in the general case where the upper bound acceptance need not be monotonic. The solution involves "ironing" at each test result. For instance, at some test result the agent may be required to treat all applicants from the 70th through 80th percentiles of U_A identically. If the agent is to accept less than 20% of the applicants at this test result, she deterministically accepts her favorites. If she is to accept $(20+x)\%$ for $0 < x < 10$, she accepts her favorite 20% deterministically, and gives each of the 10% of applicants in the pooling range an $x/10$ chance of acceptance. If she is to accept 30% of applicants or more, she once again deterministically accepts her favorites. After ironing applicants in an appropriate manner, the problem can be solved using an approach similar to that in Section 3.1.

To see the potential benefit of randomization, consider the special case in which there is no meaningful hard information: all applicants share the same test result. Under alignment up to distinguishability, where higher agent utility U_A implies higher principal utility U_P, it would be optimal to give the agent full discretion. Under anti-alignment, with higher U_A corresponding to *lower* U_P, the principal should ignore the agent and randomly accept applicants. If U_P were nonmonotonically increasing, decreasing, then increasing in U_A, though, then the principal might let the agent accept some applicants deterministically and others probabilistically.

One conclusion from Appendix A is that, despite the potential benefit of randomized acceptances, randomization need only be used on at most a single test result. So when test results are continuously distributed, and thus behavior at any single test result is irrelevant, we still get a deterministic optimal contract that can be implemented by specifying an acceptance rate function. This optimal contract may not

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

achieve the upper bound of principal payoff of UBAR, though.

3.3 The normal specification

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

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

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

The agent's expected utility is $U_A = \mathbb{E}[Q|T,S] + B$. One can derive the principal's expected utility $U_P(T,U_A)$ as below; see details of the calculations in Appendix G.1.

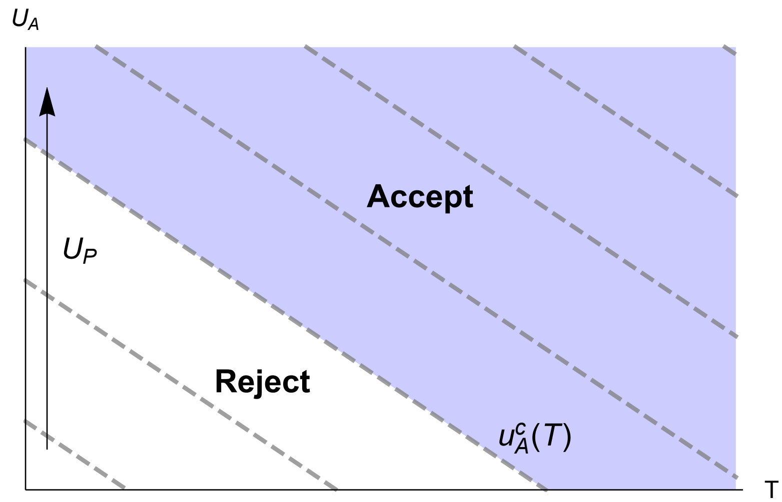
$$\begin{aligned} U_P(T,U_A) =& \beta_T \cdot T + \beta_{U_A} \cdot U_A, \text{ for} && (8)\\ \beta_T =& \frac{\sigma_B^2\sigma_Q^2}{(\sigma_Q^2+\sigma_T^2)(\eta+\sigma_B^2)}, && (9)\\ \beta_{U_A} =& \frac{\eta}{\eta+\sigma_B^2}, && (10)\\ \text{and } \eta \equiv& \frac{\sigma_Q^4\sigma_T^4}{(\sigma_Q^2+\sigma_T^2)(\sigma_Q^2\sigma_T^2+\sigma_Q^2\sigma_S^2+\sigma_T^2\sigma_S^2)}.^{20} && (11) \end{aligned}$$

$U_P(T,U_A)$ is linear in both T and U_A, with coefficients $\beta_T > 0$ and $\beta_{U_A} \in (0,1)$. The fact that β_{U_A} is positive confirms that utilities are aligned up to distinguishability, as implied by Lemma 1: (one unit) higher quality plus bias in U_A translates into (less than one unit) higher expected quality in U_P. Linearity of U_P means that the principal indifference curves are linear in (T,U_A)-space, with slope $-\frac{\beta_T}{\beta_{U_A}} < 0$. UBAR accepts all applicants "up and to the right" of a cutoff indifference curve. See Figure 1.

Following Proposition 1, one implementation of UBAR specifies an acceptance rate at each test result. Applicants' test results and agent utilities (T,U_A) are distributed joint normally and are positively correlated. So the share of applicants above the cutoff indifference curve – a downward sloping line – is increasing and follows a normal CDF function. Indicating the CDF of a standard normal distribution by Φ, the

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

Figure 1: Upper Bound Acceptance Rule for the Normal Specification



The dashed curves are principal indifference curves, each a line with slope $-\beta_T/\beta_{U_A}$. The arrow indicates the direction of higher principal utilities U_P, i.e., that utilities are aligned up to distinguishability. Under UBAR, applicants are accepted if U_A is above the cutoff indifference curve $u_A^c(T)$ for which $U_P = u_P^c$, inducing a mass of k applicants accepted in total.

share of applicants accepted under UBAR at test result T works out to $\alpha^{\text{UBAR}}(T) = \Phi(\gamma_T^* T - \gamma_0^*)$, with $\gamma_T^* > 0$ given by

$$\gamma_T^* \equiv \frac{\sigma_Q^2\sqrt{\eta + \sigma_B^2}}{\eta(\sigma_Q^2 + \sigma_T^2)}. \tag{12}$$

The value of γ_0^* – for which I do not provide an explicit formula – is then set so that a total of k applicants are accepted.

We can equivalently express this contract in the manner of Proposition 2: the agent can pick any k applicants subject to $\mathbb{E}[C(T)|\text{Hired} = 1] \geq \kappa$, for some score function C and average score κ, and the inequality constraint binds. The score function can be chosen as any negative affine transformation of the cutoff indifference curve of Figure 1. Indifference curves are downward sloping lines, so C can be any increasing linear function. For the normal specification I take the convention of setting $C(T) = T$, i.e., a score equal to the test result. (In the context of the normal specification, I hereafter sometimes call T a *test score*.) That is, the agent can choose any k applicants with a sufficiently high average test score. Denote the optimal choice of this minimum average test score – the average test score induced by UBAR – by κ^*.[21]

The following Proposition formalizes these results.

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

1. *An acceptance rate function of $\alpha(T) = \Phi(\gamma_T^* T - \gamma_0^*)$; or,*
2. *An average test score of accepted applicants, $\mathbb{E}[T|\text{Hired} = 1]$, at or above κ^*. In this case the agent will choose applicants so that $\mathbb{E}[T|\text{Hired} = 1] = \kappa^*$.*

Let us focus on the first implementation, the acceptance rate function. The optimal contract induces a normal CDF acceptance rate – an S curve – of the form $\alpha(t) = \Phi(\gamma_T t - \gamma_0)$. More applicants are accepted at higher test results, with the share of applicants accepted approaching 0 as $t \to -\infty$ and approaching 1 as $t \to \infty$. A steeper contract, with a higher γ_T, would correspond to a higher average test score in the second implementation. See Figure 2 for an illustration of such contracts.

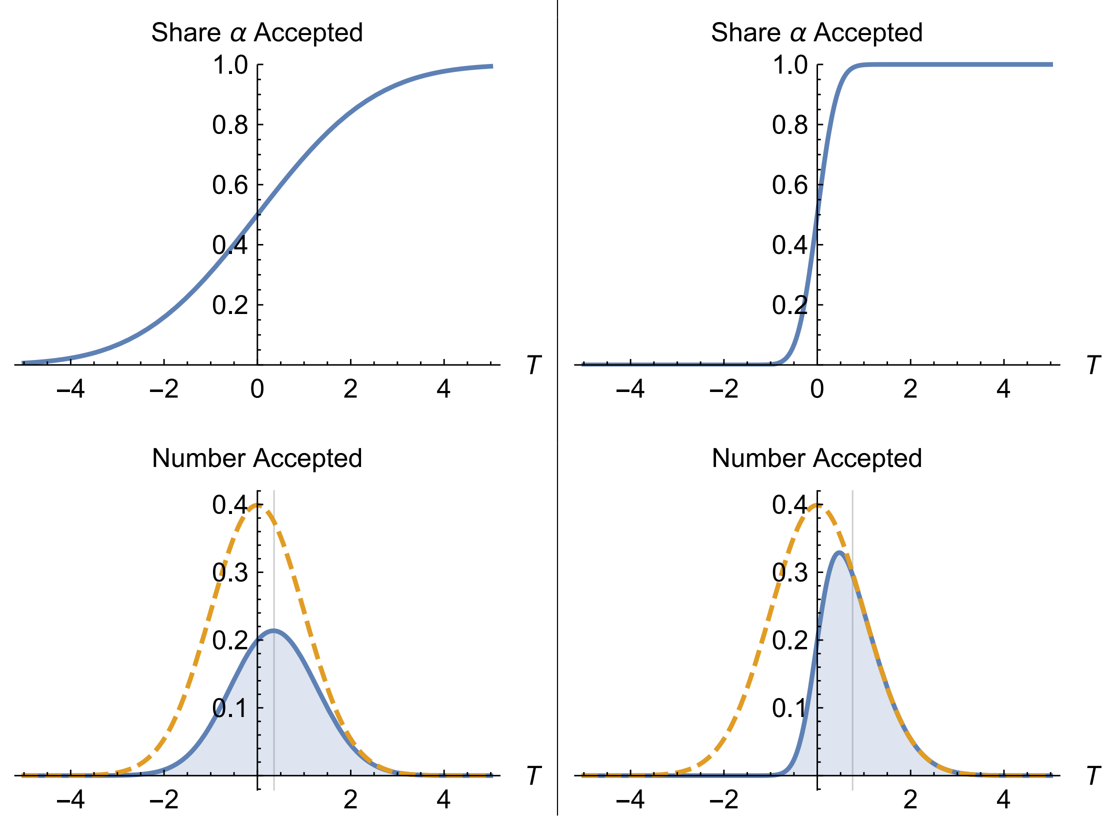
Heuristically, a steeper contract means that hiring depends more on the test and less on the agent's input: steeper contracts give the agent "less discretion." Being more precise, in Appendix B.2 I show that the Full Discretion contract, in which the agent selects her favorite applicants, would induce a normal CDF acceptance rate with steepness γ_T^{FD} satisfying $0 < \gamma_T^{\text{FD}} < \gamma_T^*$. The agent prefers a flatter acceptance rate than does the principal ($\gamma_T^{\text{FD}} < \gamma_T^*$) because she cares about idiosyncratic factors in addition to quality. So in the range of contracts that an agent may face ($\alpha(T) = \Phi(\gamma_T T - \gamma_0)$ with $\gamma_T > \gamma_T^{\text{FD}}$), a steeper contract requires the agent to pick fewer applicants with low test scores, whom she prefers on the margin, and more with high scores. Taking γ_T to infinity would yield the No Discretion contract in which an applicant is accepted if and only if their test score is above a cutoff.

We can now explore comparative statics on the optimal steepness γ_T^* with respect to the agent's information (decreasing in σ_S^2) and bias (increasing in σ_B^2). Additional comparative statics – on k, σ_T^2, and σ_Q^2 – can be found in Appendix B.1.

[21] A formula for κ^* in terms of the primitive parameters is given by Equation (54) in Appendix G.3, plugging in for $\sigma_{U_P}^2$ from (18): $\kappa^* = \dfrac{\sigma_Q^2 \sqrt{\sigma_Q^2 + \sigma_T^2} \cdot R(k)}{\sqrt{\sigma_Q^2 + \frac{\eta^2}{\eta + \sigma_B^2}}}$, with $R(k)$ as defined below in (22).

Figure 2: Contracts with $\alpha(T) = \Phi(\gamma_T T - \gamma_0)$ for different γ_T.

Low γ_T: Flat contract, Low avg test score	High γ_T: Steep contract, High avg test score



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

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

1. *γ_T^* increases in σ_S^2, with $\lim_{\sigma_S^2 \to 0} \gamma_T^* \in (0, \infty)$ and $\lim_{\sigma_S^2 \to \infty} \gamma_T^* = \infty$.*
2. *γ_T^* increases in σ_B^2, with $\lim_{\sigma_B^2 \to 0} \gamma_T^* \in (0, \infty)$ and $\lim_{\sigma_B^2 \to \infty} \gamma_T^* = \infty$.*

Part 1 finds that as the agent becomes better informed, the contract gets flatter, giving the agent more discretion. Part 2 finds that as the agent becomes more biased, the contract gets steeper, giving the agent less discretion. In the limits where the agent is fully uninformed or where her preferences are entirely unrelated to quality, we approach No Discretion. Recall that Hoffman et al. (2018) compares the two extreme contracts of Full versus No Discretion. They argue that Full Discretion is preferred when the agent has weak biases and high private information, and No Discretion when the agent has strong biases or low information. Parts 1 and 2 point to similar tradeoffs in the optimal contracts. The firm should use a flatter contract, yielding more discretion, when biases are weak or when information is high.

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

4 Unknown agent type in the normal specification

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

4.1 Overview and connection to one-dimensional delegation

Before moving to the analysis, it will be helpful to recall the one-dimensional delegation problem, introduced by Holmström (1977, 1984). In that problem, the agent observes a one-dimensional state, which determines principal and agent preferences over a one-dimensional action. The contract takes the form of a "delegation set" specifying the actions that the agent may choose. A common result in this literature is that, under some functional form and distributional assumptions, an interval delegation set is optimal. For instance, an agent who is biased towards lower actions than the principal wants would be given an action floor. Some of these papers also consider the possibility of "money burning," an auxiliary action that reduces the payoffs of both players.[22] Those papers study conditions under which money burning is or is not used in conjunction with optimal delegation sets.

The problem of selecting applicants is posed as a higher-dimensional one. The action effectively corresponds to the entire function mapping test results to acceptance rates. (In fact, that action space would only describe deterministic contracts; stochastic contracts could be more general.) The players' preferences over this action are determined by the agent's two-dimensional type. When the agent's type was known and the principal faced no aggregate uncertainty, Section 3 found that it was optimal to specify a single acceptance rate function, i.e., a single action. When the agent's type is unknown, however, the principal may offer a menu of acceptance rates to screen across agent types.[23]

I will be able solve for this optimal menu under the normal specification. I do so below by formally transforming the problem of selecting applicants into a one-

[22]Some version of a result that interval delegation sets are optimal appears in Melumad and Shibano (1991), Martimort and Semenov (2006), Alonso and Matouschek (2008), Goltsman et al. (2009), Kováč and Mylovanov (2009), Amador and Bagwell (2013), Ambrus and Egorov (2015), Amador and Bagwell (2016), and Amador et al. (2018). The latter six of these papers consider money burning of some form. Kováč and Mylovanov (2009) highlights that when players are risk averse, taking a random action rather than its expectation can be one source of money burning.

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

dimensional delegation problem with money burning.[24] The agent's behavior in any contract is determined by a one-dimensional projection of her two-dimensional type. Moreover, there is a one-dimensional set of "frontier" actions – the normal CDF acceptance rates, parametrized by steepness γ_T. Any other acceptance rate function gives the players the payoffs of some normal CDF acceptance rate, minus money burning that harms both players. Because the agent puts less weight on test scores than does the principal (and more weight on idiosyncratic factors), the agent has a corresponding "bias" towards flatter acceptance rates that follow test scores less closely. Under this transformation, I can apply conditions from the one-dimensional delegation analysis of Amador and Bagwell (2013) and Amador et al. (2018) to characterize the optimal contract as one that sets a floor on actions – a floor on the steepness of the acceptance rate – and does not burn money.[25] This contract can also be implemented by allowing the agent to select applicants subject to a floor on their average test score.

4.2 Rewriting payoffs

As a preliminary step, let us calculate the distribution of principal and agent utilities at each test result for a fixed agent type θ. The conditional distribution of agent utilities U_A given test score T is derived in the proof of Proposition 3 part 1 as

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

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

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

The mean $\mu_{U_A}(T)$ is linear in the test result T but does not depend on the agent's type θ. The variance $\sigma^2_{U_A}(\theta)$ is constant in T but depends on the type θ. Recall that $\eta(\theta)$, defined in (11), depends on the agent's type through σ_S^2 but not σ_B^2.[26]

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

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

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

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

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

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

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

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

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

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

The z-score captures how much the agent likes an applicant, controlling for the public information. Since U_P is increasing in U_A, it also describes how much the principal likes the applicant (up to distinguishability) relative to others at that test score .

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

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

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

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

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

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

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

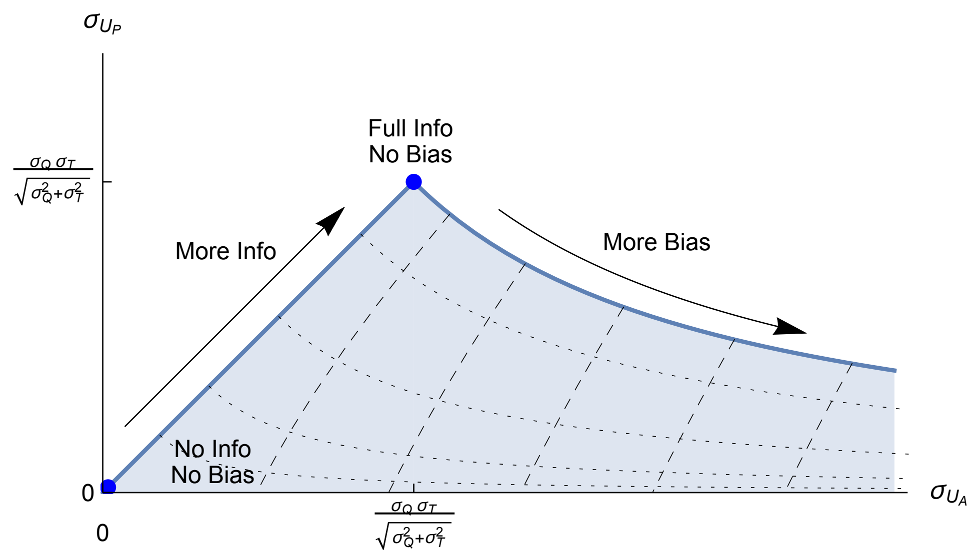
Lemma 2.

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

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

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

Figure 3: 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.

4.3 Rewriting the contracting space

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

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

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

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

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

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

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

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

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

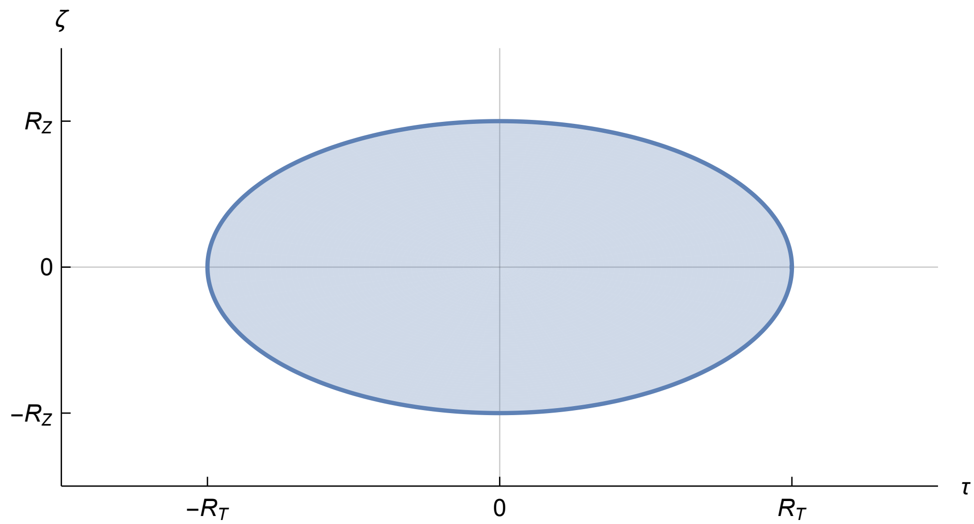
That is, the set $\overline{W}$ of possible (τ, ζ) is an ellipse centered at $(0, 0)$ with principal axes R_T and R_Z; see Figure 4. For intuition, recall that the unconditional distribution of test scores is normal with mean 0 and variance $\sigma_Q^2 + \sigma_T^2$, while the distribution of z-scores at every test result is normal with mean 0 and variance 1. The highest possible average test score over sets of k applicants comes from selecting the k applicants with the highest test scores (a step function acceptance rate in T), yielding $\tau = R_T$ and $\zeta = 0$. The highest possible average z-score comes from selecting the k applicants with the highest z-scores (a constant acceptance rate function), yielding $\tau = 0$ and $\zeta = R_Z$. We get the opposite points if we select applicants with the lowest test results or the lowest z-scores. The boundary connecting these four extreme points is an ellipse.

As we have seen, if a contract asks an agent to select any set of applicants subject only to a fixed average test score, that contract induces a normal CDF acceptance rate; see Proposition 3. Such a contract can now be interpreted as one which restricts τ but not ζ. Given that the agent's payoff is increasing in ζ, the agent must be choosing ζ on the upper frontier of $\overline{W}$. In other words, the (τ, ζ) points on the upper boundary of the ellipse in Figure 4 are induced by deterministic contracts with normal CDF acceptance rates of the form $\alpha(T) = \Phi(\gamma_T T - \gamma_0)$. Applicant pools with $\tau > 0$ correspond to $\gamma_T > 0$.

Any contract can be characterized as some subset $W \subseteq \overline{W}$ of feasible (τ, ζ) pairs. The contract can specify any (measurable) subset of possible τ in $[-R_T, R_T]$, since test scores are observable. Then, for each allowed τ in W, there is some range of possible ζ.[27] Monotonicity – Observation 1 part 2 – implies that the highest possible

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

Figure 4: The space of feasible (τ, ζ) values, $\overline{W}$.



ζ at each allowed τ must be weakly positive. (Otherwise the agent could "permute" her report, listing her less preferred applicants as more preferred, to flip the sign of ζ.) And for any chosen τ the agent would always pick this highest ζ. For the purposes of contracting, then, we can restrict attention to the subset of $\overline{W}$ with $\zeta \geq 0$; graphically, the upper half of the ellipse in Figure 4.

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

4.4 Projecting the type space to one dimension

Equations (20) and (21) show that the principal and agent preferences over (τ, ζ) depend on θ only through the standard deviation terms $\sigma_{U_A}(\theta)$ and $\sigma_{U_P}(\theta)$. Indifference curves are downward sloping and linear in (τ, ζ)-space. A higher value of the respective standard deviation leads to a higher weight on ζ relative to τ, implying flatter indifference curves. Because $\sigma_{U_P}(\theta) < \sigma_{U_A}(\theta)$, the agent has flatter indifference curves than does the principal. The ideal point for each player is on the upper-right frontier of the ellipse $\overline{W}$, defined in Lemma 3 and illustrated in Figure 4. Due to her flatter indifference curves, the agent's ideal point has a higher average z-score ζ and a lower average test score τ than the principal's.

The agent's utility over (τ, ζ) depends on θ only through the one-dimensional statistic $\sigma_{U_A}(\theta)$; we can equivalently write $V_A(\tau, \zeta; \theta)$ as $V_A(\tau, \zeta; \sigma_{U_A}(\theta))$. Hence, the principal can never separate any two agent types θ with the same $\sigma_{U_A}(\theta)$. They choose the same (τ, ζ) given any contract.[28] So in solving for the optimal contract, it is without loss of generality to average the principal's payoffs across types θ with the same $\sigma_{U_A}(\theta)$. Formally, I now replace the principal's objective function $V_P(\tau, \zeta; \theta)$ from (21), the payoff given an agent of type θ, with $\hat{V}_P(\tau, \zeta; \sigma_{U_A}(\theta))$, the principal's subjective expectation of $V_P(\tau, \zeta; \theta)$ conditional on $\sigma_{U_A}(\theta)$:

$$
\begin{aligned}
V_A(\tau, \zeta; \sigma_{U_A}(\theta)) &= \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} \cdot \tau + \sigma_{U_A}(\theta) \cdot \zeta && (26) \\
\hat{V}_P(\tau, \zeta; \sigma_{U_A}(\theta)) &\equiv \mathbb{E}_{\theta \sim G}[V_P(\tau, \zeta; \theta) | \sigma_{U_A}(\theta)] \\
&= \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} \cdot \tau + \mathbb{E}_{\theta \sim G}[\sigma_{U_P}(\theta) | \sigma_{U_A}(\theta)] \cdot \zeta \\
&= \frac{\sigma_Q^2}{\sigma_Q^2 + \sigma_T^2} \cdot \tau + \hat{\sigma}_{U_P}(\sigma_{U_A}(\theta)) \cdot \zeta, && (27) \\
\text{for } \hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A}) &\equiv \mathbb{E}_{\theta \sim G}[\sigma_{U_P}(\theta) | \sigma_{U_A}(\theta) = \tilde{\sigma}_{U_A}].
\end{aligned}
$$

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

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

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

4.5 Projecting the action space to one dimension plus money burning

Focus on the top half of the ellipse $\overline{W}$ of Figure 4, the values of (τ, ζ) that may be induced by an agent's choices in a contract. Any pool of applicants with (τ, ζ) off of the upper-right frontier of the ellipse – an acceptance rate that is not of the form $\Phi(\gamma_T T - \gamma_0)$, for $\gamma_T \geq 0$ – is dominated. There is another pool of k applicants with strictly higher τ at the same ζ that improves the payoff of both players. It is as if there is a one-dimensional action space along this upper-right frontier, plus the possibility of joint "money burning" that hurts both players. Let us make that formal.

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

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

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

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

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

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

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

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

Importantly, the money burning cost δ is the same for both players, and does not depend on the agent's type θ. This will mean that it fits the framework of one-dimensional delegation with money burning developed in Amador and Bagwell (2013). Instead of thinking about an applicant pool as having payoff-relevant moments τ and ζ, we can equivalently think about it as having payoff-relevant moments ζ and δ. An outcome of a contract corresponds to an average z-score $\zeta \in [0, R_Z]$ along with a level of money burning $\delta \geq 0$.[29]

4.6 Optimal contracts

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

More meaningfully, a *ceiling* on the unobservable ζ is equivalent to a *floor* on the observable average test score τ of accepted applicants. In either case, the agent picks (τ, ζ) from an interval on the upper-right frontier of the ellipse $\overline{W}$. See Figure 5.

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.

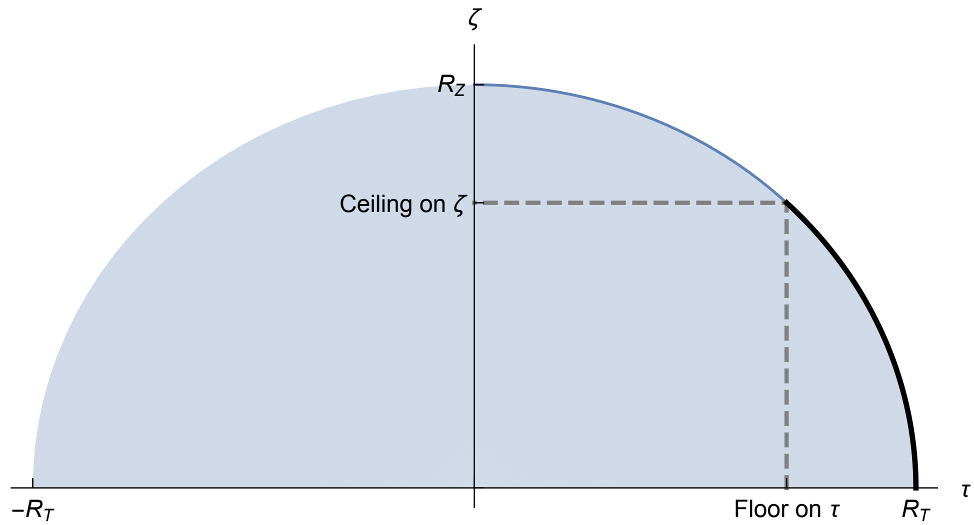
Proposition 5 formalizes the implementation of this optimal contract as either a floor on the steepness of a normal CDF acceptance rate function, or as a floor on the average test score.[30]

Proposition 5. *Under the normal specification, let the distribution H have continuous pdf h over its support, with the support a bounded interval in $\mathbb{R}_+$, and let* $\hat{\sigma}_{U_P}(\cdot)$

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

[30]Proposition 10 in Appendix E.3 derives these same contract forms as optimal in a model combining the idiosyncratic biases of the normal specification with the systematic biases of the two-factor model.

Figure 5: The contract as a floor on the average test score.



Here I highlight on the upper-half of the ellipse $\overline{W}$ from Figure 4. An agent of any type values both τ and ζ positively. So if the agent is given a ceiling on ζ, she chooses some (τ, ζ) point on the thick black curve on the upper-right frontier, and has full flexibility among these points. The agent acts identically if given a corresponding floor on τ.

be continuous over the support. Suppose $H(\tilde{\sigma}_{U_A}) + (\tilde{\sigma}_{U_A} - \hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A}))h(\tilde{\sigma}_{U_A})$ is nondecreasing in $\tilde{\sigma}_{U_A}$. Then the optimal contract is deterministic and can be characterized in either of the following ways. The agent is allowed to hire any set of k applicants, subject to:

1. *An acceptance rate function $\alpha(T)$ of the form $\Phi(\gamma_T T - \gamma_0)$, with γ_T at or above some specified level $\Gamma > 0$; or,*
2. *An average test score of hired applicants τ at or above some specified level $\kappa > 0$.*

As in Amador and Bagwell (2013) and other papers on one-dimensional delegation, the floor (Γ or κ) is set to a level that is correct, on average, for the agents who are bound by the floor.[31] It is possible that the floor always binds – that all agent types

[31] By itself, the first-order condition on how the floor is set only implies that there is no benefit from making the minimum acceptance rate steeper or shallower in the class of normal CDFs; it does not imply that there is no benefit from pointwise increases or decreases in the acceptance rate at individual test results that take us outside the normal CDF class. However, this latter claim also holds. To see why, consider applying an arbitrary newly proposed acceptance rate function (which still accepts k applicants) to the set of agents who are currently bound at the floor. This new acceptance rate function induces some average test score and average z-score (τ, ζ) in the ellipse $\overline{W}$

choose the same average test score of κ. Or the principal might "screen" agent types by setting the floor to be low enough that agents sometimes exceed it.[32]

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

Commonly known information, uncertain bias.

Proposition 6. *Under the normal specification, suppose that the agent's information level σ_S^2 is commonly known. Then the optimal contract takes the same form as in Proposition 5. Under either implementation of the optimal contract, the agent will choose applicants so that the floor constraint ($\gamma_T \geq \Gamma$ or $\tau \geq \kappa$) binds with equality.*

Proposition 6 states that when the agent's bias is uncertain but her information is commonly known, the principal doesn't screen across agent types. As in Proposition 3, he simply fixes the acceptance rate function, or the average test score, in advance. Additional distributional assumptions such as those in Proposition 5 are not needed.

In fact, Proposition 6 follows from the slightly stronger result of Proposition 11 in Appendix G. Proposition 11 shows that there will likewise be no screening if the

(see Lemma 3) – the same pair (τ, ζ) for each of the agents. Any (τ, ζ) is weakly dominated for the principal by some point on the frontier, induced by a normal CDF acceptance rate; and the current normal CDF acceptance rate has been established to be preferred to any other.

[32] Call the principal's ex ante preferred test score the one he would set if restricted to giving the agent a (no-screening) contract that fixed the average test score. By Lemma 2 of Amador et al. (2018), the floor in the contract of Proposition 5 will be always-binding if and only if the agent's highest possible ideal average test score over the type support is below the principal's ex ante preferred test score. In this case the principal would set the floor at his ex ante preferred level and the agent would always want a lower score. One sufficient condition for a non-binding floor would be that the support of $\sigma_{U_A}(\theta)$ under H extends to 0, in which case any floor would be nonbinding for agents with $\sigma_{U_A}(\theta)$ small enough.

principal's minimum ideal τ over all possible types is above the agent's maximum ideal τ – that is, if the maximum $\hat{\sigma}_{U_P}(\sigma_{U_A}(\theta))$ in the support is below the minimum $\sigma_{U_A}(\theta)$. For instance, the conclusions of Proposition 6 apply any time $\hat{\sigma}_{U_P}(\tilde{\sigma}_{U_A})$ is always decreasing in $\tilde{\sigma}_{U_A}$. Commonly known information is sufficient to imply that $\hat{\sigma}_{U_P}$ is decreasing (as seen in Lemma 2 part 2 and in Figure 3) but is not necessary.

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

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

5 Discussion and extensions

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

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

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

5.1 Implementation in finite economies

One important simplification of this paper is to assume that there is a "large number" of applicants, modeled as a continuum. Under this simplification, I derive optimal contracts that can be implemented in two alternative ways, through restrictions on acceptance rate functions or through minimum average scores. With a finite number of applicants instead of a continuum, of course, the firm could implement approximations of each of these contract forms. But these approximated contracts would no longer necessarily be exactly optimal – nor would approximations of the two different implementations remain equivalent.

Let the agent's type be commonly known, in which case Propositions 1 and 2 formalize the acceptance rate and minimum average score implementations. Suppose further that hard information takes the form of a continuous-valued test result. Even with a very large finite number of applicants, then, no two applicants will be exactly identical on observables. How should we think about putting into practice a contract that specifies a distinct acceptance rate at each realization of hard information?

One natural finite approximation of the acceptance rate implementation might be to divide realizations of hard information into bins and then specify an acceptance rate at each bin. For instance, with test results ranging from 0 to 100, the firm may divide test results into the four bins of 0-25, 25-50, 50-75, and 75-100.[33] The firm then requires the manager to accept zero applicants in the bottom bin, say, 10% of applicants in the next higher bin, 25% of applicants in the third bin, and 50% of applicants in the top bin. The firm here faces a tradeoff over the number versus the size of these bins. Combining multiple small bins into a larger one means that instead of separately choosing the top applicants in each small bin, the manager can select the top set of applicants across the bins. But having fewer bins restricts the firm's ability to force the manager to treat observably distinct applicants differently from

[33] One might want to condition the cutoffs for these bins on the realized distribution of test results; for example, each of the four bins could be a test result quartile. Similarly, for the average score implementation below, one might want to adjust the average score cutoff in response to the realized distribution of test results.

one another.

In contrast, the minimum average score contract can be approximated for finite economies without any need for binning. Take the score function as specified in the continuum contract, and require that the average score of selected applicants must be at or above a floor. (With aggregate uncertainty, the floor may not exactly bind.) The score function effectively puts all applicants into the "same bin" and allows them all to be compared to one another. For the example above, perhaps the firm specifies a minimum average test result of 70. If the manager happens to see unexpectedly strong private signals about some applicants with test results below 50, she has the flexibility to accept more of these low-scoring applicants as long as she also accepts more high-scoring ones (and fewer in the middle).

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

5.2 Statistical discrimination and commitment

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

In this paper, I begin with the assumption that an agent is in fact biased and I search for an optimal contract restricting her behavior. There is, however, a connection to the bias test. When the agent's bias and information structure are known, the "upper bound acceptance rule" of Propositions 1 and 2 equalizes the marginal quality across all observables (realizations of hard information). In other words, it "de-biases" the agent by inducing her to select applicants in a manner that passes the bias test.

Interestingly, agents are *not* necessarily fully de-biased by the optimal contract when their types are unknown. The screening contract of Proposition 5 proposes a floor on the steepness of the acceptance rate as a function of the test score. First consider the agents who find this floor binding: a mix of those with good information and/or a strong bias. Those with a stronger bias will pick worse marginal applicants at low test scores, and those with better information will pick better marginal applicants at low test scores. Averaged across all agents at the floor, the quality of the marginal applicant is indeed equalized across scores – the agents collectively pass the bias test.[34] Now consider those agents who choose steeper acceptance rates than required: those with poor information and/or a weak bias. They hire their first-best pool of applicants. The marginal applicants they hire at high test scores are of higher quality, on average, than those at low test scores. These agents fail the bias test.

The bias test has a nice connection to the role of the principal's commitment power. If an agent's choices pass the bias test, the principal does not want to adjust acceptance rates ex post. When the bias test is failed, though, the principal is tempted to intervene. He wants the agent to accept more applicants at test results with high marginal quality, and fewer at test results with low marginal quality. In particular, if an agent facing the contract of Proposition 5 chooses an acceptance rate steeper than what is required, the principal wants to force this agent to go back and choose an even steeper acceptance rate. Of course, the agent would alter her initial choice of applicants if she did not trust the principal to honor the contract; the principal is ex ante better off by committing not to change the rules after the fact.

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

[34]Recall from footnote 31 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.

5.3 Hidden test results and multiple agents

If the principal can hide the test results from the agent, he can potentially improve on the contracts that I consider. As a simple illustration, consider the normal specification with known agent type. There, when test results can be hidden, the principal can actually achieve his first-best outcome. One such contract would be as follows. The principal hides the test results T and asks the agent to report her privately observed signal S for every applicant. The principal then calculates the variance of the reported values of S as well as the covariance of S with T. If the agent were in fact to report each S truthfully, the variance of S reports would be $\sigma_S^2 + \sigma_Q^2$ and the covariance between S and T would be σ_Q^2. Any misreport that maintained the variance of S would reduce the covariance. So, if the variance and covariance match the predictions, the principal infers that S has been truthfully reported and the contract implements the principal's first best. If not, the contract chooses applicants uniformly at random. The threat of random selection incentivizes the agent to report truthfully.

Now return to public test results, but suppose that multiple agents evaluate each applicant and these agents cannot communicate. From the perspective of one agent, another agent's soft reports are exactly like hidden test results. We see that the existence of other agents' reports can give the principal levers to extract additional information from the agents.

5.4 Inference from performance data

This paper has been studying the principal's problem of choosing a contract given some specified beliefs about an agent's type. Now suppose that the principal is looking to set the contract for an agent of unknown type by using data from past hires. In particular, the principal has access to ex post performance data – quality realizations – of previously hired applicants. How might this performance data be used to determine contracts going forward? [35]

[35] To remain consistent with the previous analysis, assume that the newly introduced performance data will only used to design new contracts: there is still no way to directly reward agents for hiring applicants who end up performing better. Also maintain the assumption that "principal fundamentals" – the distribution of applicant quality in the population and the informativeness of the test – are known. It is an interesting question in its own right to consider how the principal would best learn about these fundamentals from the data. See the discussion "Inferring σ_Q^2 and σ_T^2" in Appendix F.

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

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

There are two obvious objections to the above exercise. First, the agent might not act myopically – she might alter her hiring behavior at early periods if she knows that her behavior will affect the contract she is offered in later periods. Second, performance data for one's hires might only be available after a long delay. A more reasonable exercise may be to suppose that the principal gathers performance data from a pool of agents and then uses the aggregated results to determine other agents' contracts.

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

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Hoffman, M., L. B. Kahn, and D. Li (2018): "Discretion in Hiring," *Quarterly Journal of Economics*, 133, 765–800.

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

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

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

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

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

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

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

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

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

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

Prendergast, C. and R. H. Topel (1996): "Favoritism in organizations," *Journal of Political Economy*, 104, 958–978.

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

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

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