Bidimensional screening with substitutable attributes*

Preliminary. Click to access latest version.

Orestis Vravosinos†

October 7, 2024

Abstract

An agent (she) has a bidimensional type consisting of her (hard) evidence and talent, both valued by the principal (he), who decides whether to reward the agent by asking her for evidence and a cheap talk message about her talent and then possibly testing her at a cost. The test score is increasing in evidence and talent. When the test score is less sensitive to talent than talent is valuable to the principal, the agent has incentives to hide evidence to influence how the principal interprets her test score. The optimal mechanism makes two types of errors, both favoring high- over low-evidence agents: (i) it rewards some unworthy (i.e., whom the principal would prefer not to reward) high-evidence agents without testing them, only asking them for evidence, and (ii) among agents who do not have enough evidence to get rewarded without a test, it rewards (after testing) some unworthy high-evidence agents while rejecting some worthy low-evidence ones.

Keywords: evidence game, signal jamming, manipulation, scoring, testing, under-disclosure, multidimensional screening
JEL classification codes: D82, D83

*I am grateful to Erik Madsen for his continued guidance and terrifyingly insightful comments. The paper has also significantly benefited from discussions with Debraj Ray. I also thank Agata Farina, Alessandro Bonatti, Alexander Bloedel, Alexander Jakobsen, Annie Liang, Ayça Kaya, Basil Williams, Bobby Pakzad-Hurson, Dana Foarta, David Pearce, Dilip Abreu, Gonzalo Cisternas, Guillaume Haeringer, Idione Meneghel, Inga Deimen, Laura Doval, Leeat Yariv, Marina Halac, Moritz Meyer-ter-Vehn, Sevgi Yuksel, Yeon-Koo Che, and participants in the 35th Stony Brook International Conference on Game Theory and seminars at NYU for helpful comments and discussions.

†New York University; e-mail: orestis.vravosinos@nyu.edu.

1 Introduction

In many environments, a principal (he) needs to evaluate an agent (she) by using a test (or signal) that measures a combination of the agent's valuable attributes—without revealing each attribute separately. The agent can present (favorable) evidence to prove that some of her attributes are high but has no evidence on other attributes. Ideally, the principal would like to evaluate the agent in two ways: (i) by testing her and (ii) by having her present evidence. Both the test and evidence can provide valuable information to the principal.

In principle, it is in the agent's best interest to present all her evidence to show that her corresponding attributes—which are valued by the principal—are high. However, by presenting evidence, the agent may affect how the test result—which measures a combination of the agent's attributes—is interpreted by the principal. Particularly, evidence on one attribute can affect the informational content of the test result on another attribute, which the agent cannot provide evidence on. Thus, testing by the principal may interfere with the agent's incentives to present evidence. This indicates a potential conflict between two ways of evaluating agent. When does the principal need to worry about this conflict? When the conflict arises, how should the principal resolve it? How should he test and evaluate the agent taking the conflict into account? These are the questions that this paper aims to answer.

This problem arises in various settings. For example, a college applicant may downplay her privileged background and overstate the struggles that she has gone through to paint her academic performance and standardized test scores as results of her brilliance rather than high-quality education. A job candidate may hide her privileged background to make the employer attribute her achievements to talent and hire her. A micro theorist on the academic job market may not present some results which she has already derived in order to use them to answer the audience's questions and appear to think fast. An employee may understate how long it took her to complete a task to make her ability (i.e., the rate at which her work hours translate into value to the firm) appear higher than it actually is. This strategy can pay off if promotion decisions rely mostly on the employer's beliefs over the employee's ability, because in the higher position, ability is more important than working long hours.

This way of thinking is so fundamental that kids also seem to follow it. Students often eagerly proclaim that they have not studied hard for an exam—not only if they are informed of their substandard performance but also when they have performed exceptionally well. By stressing their low effort (or even understating it) they may be trying to have their score attributed to their (overstated) brilliance. Effortless perfection (i.e., the need to seem perfect without apparent effort) and hiding one's effort have been documented among university students (Travers et al., 2015; Casale et al., 2016).

Despite how fundamental this way of thinking is, to the best of my knowledge, no previous work has studied the problem of evaluating people when—in order to affect how a combined signal of their various virtues is interpreted—they can hide evidence that both (i) is in principle favorable to them and (ii) contains useful information to the evaluator. I study this problem in the following setting. An agent (she) has a bidimensional type. The first dimension is her *evidence* (e.g., a college or job applicant's training, an employee's effort, a researcher's knowledge) and the second is her *talent* (e.g., a college or job applicant's innate ability, an employee's talent, a researcher's ability to think fast). The agent can verifiably disclose any part of her evidence but cannot prove that she is not withholding evidence. She cannot unilaterally prove anything about her talent, although she privately observes it.

The value of the agent to the principal (he) is non-decreasing in both her evidence and her talent. The principal ultimately wants to make a binary choice: whether to reward the agent or not. He does so by committing to a direct mechanism that conditional on (i) the evidence presented and (ii) the cheap talk statement made by the agent about her talent, (possibly) tests the agent at a cost and then decides whether to reward her. If performed, the test returns a one-dimensional (deterministic) signal (i.e., the test score) of the agent's bidimensional type. The test score is increasing in both the agent's evidence and talent. The agent wants to get rewarded independently of her type.

If the test score measures exactly what the principal values in an agent (and, thus, the principal wants to reward the agent if and only if her test score is high enough), then the test's usefulness is apparent. But what happens if he values talent to a different degree (relative to evidence) than the test measures talent (relative to evidence)? Or, in economics jargon, what if his marginal rate of substitution between talent and evidence is different from the marginal rate of substitution between the two in the test (i.e., holding fixed the test score)?

The main result concerns the optimal screening mechanism when the test (score) is less sensitive to talent than talent is valuable to the principal. The optimal mechanism features double penalization of low-evidence agents, favoring high-evidence agents in two ways: (i) it rewards some high-evidence agents—including ones that are not deserving (i.e., they give the principal a negative payoff when rewarded)—without testing them but rather only by asking them for a certain threshold level of evidence, and (ii) among agents that do not meet that threshold level of evidence, it rewards (after testing) some undeserving agents with high evidence but low talent while rejecting some deserving agents with high talent but low evidence. Remarkably, this is the structure of the optimal mechanism in the extreme case where the principal *only* values talent (i.e., his payoff for rewarding the agent is increasing in talent and constant in evidence).[1] The principal still optimally

[1] Notice that when the principal's payoff is only sensitive to talent, then automatically the test is less sensitive to talent than talent is valuable to the principal.

favors high-evidence agents by rewarding evidence even though it is worthless to him. He does so (i) to save in testing costs by rewarding high-evidence agents without testing them and (ii) to reward (after testing) some worthy low-evidence agents by also testing and rewarding some unworthy high-evidence agents, who can hide their high evidence to imitate the more talented worthy agents.

I now discuss the results in more detail. A screening mechanism is incentive compatible if and only if three conditions are satisfied. First, among agents with the same level of evidence, an agent with higher talent should be rewarded with (weakly) higher probability (than a less talented one), since she can successfully imitate an agent with (the same amount of evidence but) lower talent given that the test score is increasing in talent. Second, among agents with the same level of evidence, in order to reward a talented agent with (strictly) higher probability (than a less talented one), the principal needs to test the talented agent with high enough probability to prevent the less talented one from posing as the more talented agent—betting on the possibility that she will be rewarded without a test. Third, among agents with the same (potential) test score (i.e., the test score that they will achieve if tested), an agent with higher evidence should be rewarded with (weakly) higher probability, since she can hide (part of) her evidence and over-report her talent to imitate an agent with the same test score, lower talent, and higher evidence.

Given this characterization of incentive compatible mechanisms, I study the principal's problem under free testing. If the test is more sensitive to talent than talent is valuable to the principal, then testing does not create incentives for the agents to hide evidence, so the principal's problem is easy. By observing the agent's test score *and* evidence—which the agent presents fully—the principal learns the agent's type, thereby achieving the first-best. On the other hand, if the test is less sensitive to talent than talent is valuable to the principal, then testing *does* create incentives for the agents to hide evidence. The optimal mechanism makes two types of errors: (i) a Type I error of rejecting some valuable agents with high talent but low evidence to avoid rewarding unworthy agents with low talent but high evidence, and (ii) a Type II error of rewarding some unworthy agents with high evidence in order to also reward worthy ones with low evidence. The less sensitive the test is to talent, larger the errors are.

The results capture a stark contrast in the difficulty of hiring different types of agents employees. When skills and knowledge that can be proven through hard evidence are most valuable, the hiring process is easy. On the other hand, when talent—which is assessed by tests that are also sensitive to the candidate's training and knowledge, which the candidate can hide—is most valuable, the hiring process is flawed, favoring candidates with high-quality training and education at the expense of equally or more valuable candidates with limited training.

Under costly testing, the optimal mechanism is as follows. If the test is less sensitive to talent than talent is valuable to the principal, the principal gives the agent two paths to getting rewarded: either (i) provide enough evidence to meet a certain threshold or (ii) score high enough in the test without providing evidence.[2] As in the face of free testing, the optimal mechanism makes a Type I and a Type II error. The test score threshold balances these two errors, while the evidence threshold captures the trade-off between the benefit of testing (i.e., rejecting some unworthy high-evidence agents) to its cost. If the test is more sensitive to talent than talent is valuable to the principal, then every agent with at least a certain threshold of evidence is rewarded without a test, and among agents that do not meet that threshold, an agent is rewarded (after testing) if and only if her value to the principal is high enough to cover the testing cost. High-evidence agents are favored only to the extent that their evidence is high enough to get them rewarded without a test. Among agents who do not have such high levels of evidence, the mechanism rewards every worthy agent—without favoring high- (or low-) evidence agents.

The results have implications for hiring by prestigious employers (or, in general, hiring for highly desirable positions), promotions, and college admissions. Consider, first, hiring by a prestigious employer. Interpret (i) the candidate's CV quality (e.g., high school quality, undergraduate institution quality and GPA, awards, distinctions, reference letters) as her evidence and (ii) her ability and drive not captured by the evidence as her talent. Interpret testing as letting a less prestigious employer hire the candidate—with the option to poach the candidate later at a cost (after observing her performance in that employer). In the optimal mechanism, Ivy-Leaguers are immediately hired by prestigious employers, while worthy candidates with less impressive credentials have to go through less prestigious employers to prove their worth before they land a prestigious position. If the candidates' performance in the less prestigious position is less sensitive to talent than talent is valuable in the more prestigious position, then worthy candidates with low credentials are at a disadvantage—not only in the first stage of hiring by the prestigious employer but—also in the poaching stage.

In the context of promotions, evidence can be understood as the employee's effort and talent as her efficiency (i.e., the rate at which effort translates into productivity or value to the firm) or managerial skills.[3] Testing amounts to monitoring the employee's productivity. Then, the payoff to the principal from rewarding (i.e., promoting) the employee is the difference between her productivity in the new position (if promoted) and her productivity in her current position.The payoff is non-decreasing in effort and efficiency if both effort and efficiency have a (weakly) higher marginal productivity in the higher position. This can be naturally interpreted to mean that the higher position comes with increased responsibilities that allow the employee's effort and talent to have a larger

[2]The first option need not always provided (e.g., when the testing cost is low enough, the first option need not be offered).

[3]The employee has chosen effort in a previous stage (see section 5.3 for a discussion of endogenous evidence production), and it is assumed that the employee can show or hide how much effort she has exerted.

impact. It is also natural to think that talent (e.g., managerial skills) is (relative to effort) more important in the higher position than in the employee's current one. Then, the test (i.e., current productivity) is less sensitive to talent than talent is valuable to the employer, which means that some hard-working employees are (optimally) promoted—either with or without their productivity monitored—although their promotion destroys firm value. At the same time, some talented but not hard-working employees are not promoted to managerial positions, although their promotion would generate value for the firm.

Last, consider college admissions and standardized testing.[4] Assume that standardized tests are sensitive to the students' nurture (e.g., quality education, tutoring, extracurricular activities, opportunities to participate in competitions) and that students can pose as coming from a more modest background than they actually do (e.g., by overstating their struggles and not mentioning the tutoring that they have received in their college applications). If colleges value diversity by looking for talent and potential (i.e., trying to control for the applicants' unequal backgrounds), then (optimal) admission decisions are imperfect at the expense of students from disadvantaged backgrounds, favoring students from privileged backgrounds.

After a discussion of related literature, section 2 presents the model. Section 3 characterizes incentive-compatible mechanisms and then solves the principal's problem. Section 4 discusses applications. Section 5 presents extensions of the model, and section 6 concludes. Proofs are gathered in Appendix A.

Related literature. In my setting, the comparison between two marginal rates of substitution (MRS; between the agent's attributes) lies at the heard of the problem, combined with the fact that the agent can provide evidence on one but not her other attribute. The comparison is between the MRS in the principal's preferences and the test's MRS. While, to the best of my knowledge, no other work has studied a similar problem, the paper has links to a few strands of the literature.

First, it connects to the literature of persuasion games, where an sender discloses verifiable information (i.e., her type) to a decision maker to influence his actions.[5] When the sender (e.g., seller) perfectly knows her type and can costlessly and verifiably disclose it to the decision maker (e.g., buyer), whose payoff is increasing in the type, full unraveling emerges in equilibrium: every sender type (except possibly the lowest one) discloses her quality (Viscusi, 1978; Grossman, 1981; Milgrom, 1981). In my setting, this happens—without the need for commitment power, as in verifiable disclosure games—when the principal only values the agent's evidence.

Second, it has links to the literature on evidence games, where an agent chooses what

[4]In this setting, the college may not condition the need to take a test on the candidate's presentation of evidence (i.e., application package).

[5]See Milgrom (2008) for a review.

part of her verifiable evidence to disclose to the principal without being able to prove whether she has disclosed everything or not (e.g., see Shin, 1994; Dziuda, 2011; Hart et al., 2017). This literature several differences to mine. Most importantly, the principal cannot obtain a signal of the agent's type after she provides evidence.[6]

The ability of the principal to perform an experiment after the agent transmits information is, for example, considered in Glazer and Rubinstein (2004), Carroll and Egorov (2019), Bizzotto et al. (2020), Li (2020, 2021), Kattwinkel and Knoepfle (2023). However, in Glazer and Rubinstein (2004), Carroll and Egorov (2019), and Li (2021), while the agent's type is multi-dimensional, testing by the principal does not produce a mixed signal of the two dimensions but rather reveals one of the dimensions. Therefore, the interpretation of the information that the principal obtains from testing is not influenced by the agent's initial disclosure as in my model, where the substitutability between the two dimensions is key.[7] This force is absent also from Bizzotto et al. (2020), Li (2020), and Kattwinkel and Knoepfle (2023), where the agent's type is one-dimensional.

Third, the mixed signal that the testing generates is reminiscent of the signal jamming problem in career concern models (e.g., see Holmström, 1999). However, in these models the main force is the agent's incentives to exert effort in order to influence the principal's learning (though costless observation of the agent's productivity) of the agent's talent. Here, I focus on information transmission and testing. Namely, when evidence is about effort, the model shows how the agent's incentives to reveal or hide her effort are affected by the sensitivity of the test to effort and talent. Another difference from career concerns models is that the principal chooses whether to test the agent at a cost.

Fourth, the paper is related to the literature on mechanisms where the agent can affect the informational content of the designer's signal. In Perez-Richet and Skreta (2022), Frankel and Kartik (2019, 2022), and Ball (2024), the agent can manipulate the signal at a cost.[8] In my setting, withholding evidence does not manipulate the signal itself but it can affect the informational content of the signal. The amount of evidence that the agent possesses can then be interpreted as the ability of the agent to manipulate the signal. This gaming ability is exogenous and fixed in Perez-Richet and Skreta (2022) but stochastic and privately observed by the agent in Frankel and Kartik (2019, 2022) and Ball (2024), like evidence is in my model. However, in my setting, evidence controls the set of messages that the agent can send rather than how costly it is to send a certain

[6]There are more differences. For example, in Hart et al. (2017) the principal does not always prefer more evidence to less and the agent can choose to withhold damaging evidence. Damaging evidence also exists in Dziuda (2011), where the existence of a behavioral type is central, while in my model all agents are strategic.

[7]Also, in Glazer and Rubinstein (2004) rather than providing verifiable evidence, the agent only sends a cheap talk message. In Carroll and Egorov (2019), the principal has partial commitment power: she can commit on testing but not on action (after testing) decisions. It is assumed that the principal can commit to an action only to severely punish the agent.

[8]In Frankel and Kartik (2022), the receiver has no commitment power.

message. Also, unlike in models of costly signal manipulation, there is no pecuniary or direct cost of hiding evidence in my model. Still, there is an *endogenous* cost of hiding evidence given that the principal values evidence—unlike in the other models, where gaming ability is not valuable to the principal—and chooses how to reward it through the mechanism. These differences also set apart my setting from models of costly lying (e.g., Kartik, 2009; Sobel, 2020).

Last, my analysis contributes to the literature of multidimensional screening (e.g., see Armstrong, 1996; Rochet and Choné, 1998; Rochet and Stole, 2003). While duality approaches have proven useful in verifying a mechanism's optimality (Rochet and Choné, 1998; Carroll, 2017; Daskalakis et al., 2017; Cai et al., 2019), full characterizations of multidimensional screening problems remain challenging. Partial characterizations have, for example, been obtained (i) for the case where the principal can use costly instruments in screening (Yang, 2022) or (ii) that show when offering only the grand bundle of all products is optimal for a multi-product monopolist (Haghpanah and Hartline, 2021). I contribute to this literature by proposing a novel bidimensional screening problem and deriving a full characterization under general assumptions. I assume that the agent's type admits a full support density but impose no other restrictions on the type distribution. Also, I make no parametric assumptions on the principal's preferences or testing technology; they are only assumed to satisfy a single-crossing condition.[9] My analysis does not rely on ironing procedures (e.g., see Mussa and Rosen, 1978; Myerson, 1981; Rochet and Choné, 1998) or the duality approach. Instead, I show that the principal's problem can be reduced to a maximization problem where the objective is a linear (and thus convex) and continuous functional and the domain is a (convex and compact) space of monotone functions.[10] Bauer's maximum principle then implies that an extreme point solves the problem.[11] The proof then proceeds using well-known properties of extreme points of spaces of monotone functions. In that sense, my paper is also related to recent papers that characterize extreme points of spaces of monotone functions (e.g., see Kleiner et al., 2021; Yang and Zentefis, 2024) and then use Bauer's Maximum Principle.

2 A model of bidimensional screening with substitutable attributes

There are an agent (she) and a principal (he). The agent is privately informed of her bidimensional type (e,t), which has a full-support density $f : [0,1]^2 \to \mathbb{R}_{++}$. e is the

[9]They are also assumed to be monotone, but this assumption is not made for tractability. Rather, it captures the main force under study.

[10]For this result, it is important that the principal's (final) choice is binary.

[11]Manelli and Vincent (2007) also use Bauer's maximum principle in studying a multi-dimensional screening problem.

agent's *evidence*. That is, an agent of type (e,t) can prove to the principal that her e is at least r for any $r \in [0,e]$ by presenting evidence $r \in [0,e]$. If she reveals $r < e$, we say that she hides/partially reveals evidence. However, for no $r \in [0,1)$ can she prove that her e is not higher than r; in other words, she cannot prove that she is not withholding evidence. t is the agent's *talent*, which she cannot unilaterally prove anything about. The principal can test the agent by paying a cost $c \geq 0$.

The testing technology. The test is imperfect and works as follows.[12] Testing the agent amounts to observing a mixed, deterministic signal $\sigma(e,t) \in [0,1]$ of the agent's type (e,t). $\sigma : [0,1]^2 \to [0,1]$ is increasing and continuous in both e and t. The assumption of a deterministic increasing signal is not uncommon. In fact, it is more general than the assumption that the test reveals one of the dimensions of the agent's type, which is for example made in Glazer and Rubinstein (2004), Carroll and Egorov (2019), and Kattwinkel and Knoepfle (2023).

Payoffs. Ultimately, the principal wants to choose whether to reward the agent or not. He receives Bernoulli (gross of testing costs) payoff $u(e,t)$ from rewarding an agent of type (e,t), where $u : [0,1]^2 \to \mathbb{R}$ is non-decreasing and continuous in both e and t. If he does not reward the agent, he receives payoff normalized to 0. An isocurve of the principal's (gross) payoff is given by $I_u(\overline{u}) \coloneqq \{(e,t) \in [0,1]^2 : u(e,t) = \overline{u}\}$.[13] Define also the (strict) upper and lower contour sets $I_u^{\uparrow}(\overline{u}) \coloneqq \{(e,t) \in [0,1]^2 : u(e,t) > \overline{u}\}$ and $I_u^{\downarrow}(\overline{u}) \coloneqq \{(e,t) \in [0,1]^2 : u(e,t) < \overline{u}\}$, respectively. The agent's Bernoulli payoff is equal to 1 if she gets rewarded and 0 if not.

Canonical examples. In a linear example, $u(e,t) \coloneqq \gamma_u e + (1-\gamma_u)t - \underline{q}$, where $\gamma_u \in [0,1]$ measures how much the principal values e versus t, and $\underline{q} \in (0,1)$ measures the threshold quality that the agent needs to have to be of (positive) value to the principal. Similarly, $\sigma(e,t) \coloneqq \gamma_s e + (1-\gamma_s)t$, where $\gamma_s \in (0,1)$ measures how sensitive the test is to e relative to t. In a Cobb-Douglas specification, $u(e,t) \coloneqq e^{\gamma_u} t^{1-\gamma_u} - \underline{q}$ and $\sigma(e,t) \coloneqq e^{\gamma_s} t^{1-\gamma_s}$ with $\gamma_u \in [0,1]$ and $\gamma_s, \underline{q} \in (0,1)$.

The principal's problem. To decide whether to reward the agent, the principal designs (with commitment) a direct mechanism $M \equiv \langle T,P \rangle$ that specifies (i) the probability $T(e,t) \in [0,1]$ with which the principal will test the agent if she presents evidence e and sends cheap talk message t and (ii) the probability $P(e,t,s)$, which should be non-decreasing in $s \in [0,1]$, with which the principal will reward the agent after the agent

[12] Although called a test, this need not be a written test. For example, it can also be the agent's performance in an interview or her productivity as an employee.

[13] $I_u(\overline{u})$ is indeed assumed to be a curve for any $\overline{u}$. This is the case if, for example, $u(e,t)$ is increasing in either e or t.

has presented evidence e, sent cheap talk message t, and the test has returned result $s \in [0,1]$.[14] When no test is performed, $s = \emptyset$ and the agent is rewarded with probability $P(e,t,\emptyset)$. Notice that (e,t) refers to the message sent by the agent. When necessary to avoid confusion, we will denote by (e',t') the agent's message (i.e., evidence e' presented and cheap talk message t' sent) to differentiate it from the agent's type, which in those cases will be denoted by (e,t). Overall, the principal chooses a mechanism $M \equiv \langle T,P \rangle$, where $T : [0,1]^2 \to [0,1]$ and $P : [0,1]^2 \times ([0,1] \cup \{\emptyset\}) \to [0,1]$ with $P(e,t,s)$ non-decreasing in $s \in [0,1]$, and an agent response rule $\phi : [0,1]^2 \to [0,1]^2$ to maximize

$$\int_0^1 \int_0^1 \left\{ \begin{bmatrix} T(\phi(e,t))P(\phi(e,t),\sigma(e,t)) \\ +[1-T(\phi(e,t))]P(\phi(e,t),\emptyset) \end{bmatrix} u(e,t) - cT(\phi(e,t)) \right\} f(e,t)dtde$$

subject to $\phi(e,t) \in \arg\max_{(e',t') \in [0,e] \times [0,1]} \{T(e',t')P(e',t',\sigma(e,t)) + (1 - T(e',t'))P(e',t',\emptyset)\}$.

3 Optimal bidimensional screening with substitutable attributes

This section characterizes incentive-compatible (IC) mechanisms and then solves the principal's problem.

3.1 Simplifying the class of mechanisms

Before characterizing IC mechanisms, we show that we can restrict—without loss—the class of mechanisms that we need to consider.

Truthful mechanisms are without loss. The first simplification comes from the fact that the principal can without loss of optimality restrict attention to truthful mechanisms (i.e., mechanisms that induce truth-telling). To see why, notice that the correspondence $(e,t) \mapsto \{(e',t') \in [0,1]^2 : e' \leq e\}$ (from the type space to the message space) that determines the admissible messages for each agent type (e,t) satisfies the Nested Range Condition (NRC) of Green and Laffont (1986), who show that under this condition, the set of implementable social choice functions coincides with the set of truthfully implementable social choice functions.[15] Therefore, from now on, we restrict attention to truthful mechanisms and define IC mechanisms as follows.

[14] The condition that $P(e,t,s)$ be non-decreasing in $s \in [0,1]$ can be understood as an incentive-compatibility condition in a model where $\sigma(e,t)$ gives the score that agent type (e,t) can achieve but the agent can intentionally make mistakes in the test to reduce her score.

[15] Notice that, essentially, the principal implements a social choice function $g : [0,1]^2 \to [0,1]^2 \times [0,1]^{[0,1]}$, where $g_1(e,t)$ the probability of testing, $g_2(e,t)$ the probability of rewarding conditional on not testing, and $g_3(e,t,\cdot)$ a self-map on $[0,1]$ that (conditional on testing) maps the test result s to the probability $g_3(e,t,s)$ of rewarding.

Definition 1. A mechanism $M \equiv \langle T,P \rangle$ is IC if for every $(e,t) \in [0,1]^2$

$$(e,t) \in \underset{(e',t') \in [0,e] \times [0,1]}{\arg\max} \{T(e',t')P(e',t',\sigma(e,t)) + (1 - T(e',t'))P(e',t',\emptyset)\}.$$

Pass-or-fail tests are without loss. Next, we can see that it is possible to constrain attention to mechanisms with threshold rewarding policies conditional on testing; that is, mechanisms such that

$$P(e,t,s) = \begin{cases} 0 & \text{if } s < \sigma(e,t) \\ P_{at}(e,t) & \text{if } s \geq \sigma(e,t). \end{cases} \tag{1}$$

for any (e,t) for some $P_{at} : [0,1]^2 \to [0,1]$, where at is a mnemonic for the probability of rewarding the agent *after testing* (given that the threshold test score is met). If type (e,t) reports her type truthfully and is tested, she is then rewarded with probability $P_{at}(e,t)$. Notice that the threshold is set exactly equal to the test score that a truthfully-reporting agent can achieve. To see why constraining attention to such mechanisms is without loss of optimality, observe that among all mechanisms that (conditional on testing) reward type (e,t) with probability $P_{at}(e,t)$, the one that satisfies equation (1) minimizes incentives of other types to imitate (e,t).[16]

Moreover, agents that meet the test score threshold are rewarded with certainty. To see this, notice that the total probability with which agent (e,t) is rewarded if she truthfully reports her type is equal to $\Pi(e,t) \coloneqq (1 - T(e,t))P(e,t,\emptyset) + T(e,t)P_{at}(e,t)$, and define outcome-equivalent mechanisms as follows.

Definition 2. A mechanism $M' \equiv \langle T',P' \rangle$ is outcome-equivalent to a mechanism $M \equiv \langle T,P \rangle$ if for every (e,t), $\Pi(e,t) = \Pi'(e,t)$, where $\Pi(e,t) \equiv (1-T(e,t))P(e,t,\emptyset)+T(e,t)P_{at}(e,t)$ and $\Pi'(e,t) \equiv (1 - T'(e,t))P'(e,t,\emptyset) + T'(e,t)P'_{at}(e,t)$.

Lemma 1 shows that when testing is costly, an agent that is tested and passes the test is rewarded with probability 1 in any optimal mechanism. When testing is free, it is still without loss to constrain attention to mechanisms that reward the agent with probability 1 when she passes the test.

Lemma 1. Given any IC mechanism M, there exists an IC mechanism $M' \equiv \langle T',P' \rangle$ with $P'_{at}(e,t) = 1$ for every (e,t) that is outcome-equivalent to M. Also, for $c > 0$, in any optimal mechanism $M \equiv \langle T,P \rangle$, $P_{at}(e,t) = 1$ for any (e,t) such that $T(e,t) > 0$.[17]

[16]Namely, rewarding the agent further for performing above $\sigma(e,t)$ will result in the same probability of rewarding type (e,t) after testing her and only give additional incentives to other agents to imitate (e,t). Similarly, there is no reason to reward the agent for test scores lower than $\sigma(e,t)$. Particularly, this argument holds when we compare all mechanisms that test (e,t) with the same probability, and thus have the same testing costs.

[17]Strictly put, $P_{at}(e,t)$ can be lower than 1 for a zero-measure set of (e,t) with $T(e,t) > 0$.

The intuition behind this result is the following. The only reason to test an agent before rewarding her—rather than reward her without a test—is to prevent others from imitating her. The total probability with which each agent is rewarded is the sum of (i) the probability $(1-T(e,t))P(e,t,\emptyset)$ of getting rewarded without getting tested and (ii) the probability $T(e,t)P_{at}(e,t)$ of getting rewarded after getting tested (and passing the test). Thus, simply put, if the principal pays the cost to test an agent, he may as well assign as large a part as possible of the total probability of rewarding her to the case where he rewards her after a test.

Specifically, if agent (e,t) is not rewarded with certainty after passing the test (i.e., $P_{at}(e,t) < 1$ and $T(e,t) > 0$), then we can (i) increase the probability $P_{at}(e,t)$ with which she is rewarded conditional on getting tested (and passing the test), (ii) decrease the probability $T(e,t)$ with which she is tested, and (iii) decrease (if positive) the probability $P(e,t,\emptyset)$ of rewarding her conditional on not testing her, keeping fixed both (a) the probability $(1 - T(e,t))P(e,t,\emptyset)$ of rewarding her without testing her and (b) the probability $T(e,t)P_{at}(e,t)$ of rewarding her with testing. By doing so, we (i) keep fixed the total probability $\Pi(e,t)$ of rewarding (e,t), (ii) do not change the incentives of other types to imitate (e,t), since any agent imitating (e,t) will be rewarded with probability $(1 - T(e,t))P(e,t,\emptyset)$ (if she only has at least as much evidence as (e,t) but cannot test as high as her) or $\Pi(e,t)$ (if she can also test as high as (e,t)), and (iii) reduce the probability of testing (e,t), thereby limiting testing costs. Thus, from now on, we constrain attention to mechanisms with $P_{at}(e,t) = 1$ for any (e,t).[18]

Untalented agents do not need to be tested. Lemma 2 shows that we can further simplify the analysis by constraining attention to mechanisms where agents with zero talent are never tested.

Lemma 2. Given any IC mechanism M, there exists an IC mechanism $M' \equiv \langle T',P' \rangle$ with $T'(e,0) = 0$ for every e that is outcome-equivalent to M and has the same (expected) testing cost as M.

Here is the intuition behind this result. The only reason to test an agent before rewarding her—rather than reward her without a test—is to prevent others from imitating her. But any agent (e',t') who has sufficient evidence (i.e., $e' \geq e$) to imitate an agent $(e,0)$ with no talent can also score (if tested) at least as high as the untalented agent $(e,0)$, since the test score is increasing in e and t. Therefore, there is no point in testing untalented agents, as doing so does not reduce incentives of others to imitate them.[19]

[18]For (e,t) with $T(e,t) = 0$ the value of $P_{at}(e,t)$ does not matter, so we can again set $P_{at}(e,t) = 1$ without loss.

[19]Of course, untalented agents are a zero-measure set. Thus, even if testing is costly, the principal could optimally test them. However, restricting attention to mechanisms with $T(e,0) = 0$ for every e helps simplify notation.

3.2 Incentive-compatible mechanisms

Given what we have seen, we constrain attention to truthful mechanisms with pass-or-fail tests where untalented agents are never tested. Let $\tau(e,s)$ be implicitly given by $\sigma(e,\tau(e,s)) = s$. $\tau(e,s)$ gives the level of talent that an agent with evidence e should have to achieve test score (exactly) s. $\tau(e,s)$ is well-defined for (e,s) such that $e \in [\underline{e}(s),\overline{e}(s)]$, where $\underline{e}(s) \coloneqq \min\{e \in [0,1] : \sigma(e,1) \geq s\}$ and $\overline{e}(s) \coloneqq \max\{e \in [0,1] : \sigma(e,0) \leq s\}$.[20] Proposition 1 then characterizes IC mechanisms.

Proposition 1. A mechanism $M \equiv \langle T,P \rangle$ is IC if and only if

(i) $\Pi(e,t)$ is non-decreasing in t for every e,

(ii) $\Pi(e, \tau(e,s))$ is non-decreasing in e over $e \in [\underline{e}(s),\overline{e}(s)]$ for every $s \in [0,1]$, and

(iii) $(1 - T(e,t))P(e,t,\emptyset) \leq \Pi(e,0)$ for every (e,t),

where $\Pi(e,t) \equiv (1 - T(e,t))P(e,t,\emptyset) + T(e,t)$.

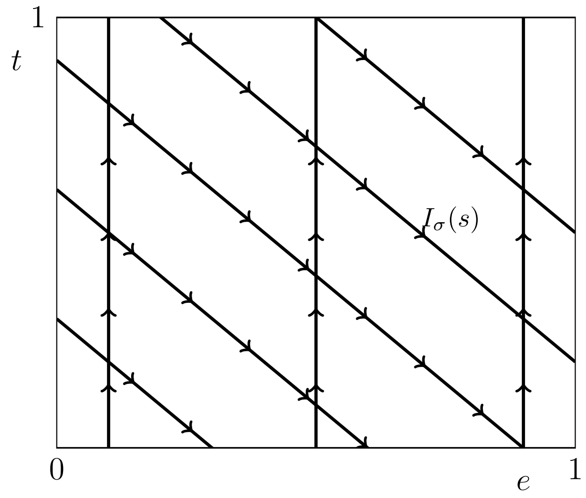
Figure 1 schematically summarizes IC conditions (i) and (ii) of Proposition 1.

Condition (i) is necessary and sufficient to ensure that an agent (e,t) does not want to reveal her evidence but under-report her talent to imitate agent (e,t') with $t' < t$, pass (e,t')'s test (since the test score is increasing in talent), and get rewarded with probability $\Pi(e,t')$. Condition (iii) is necessary and sufficient to ensure that no untalented agent $(e,0)$ has incentives to over-report her talent, trying to imitate an agent (e,t), whose test score she *cannot* achieve. Put differently, among agents with the same level of evidence e, in order to reward talented agents more frequently (than the untalented agent $(e,0)$), the principal needs to test them with high enough probability to prevent agent $(e,0)$ from imitating them. Last, condition (ii) is necessary and sufficient to ensure that agents do not want to hide some of their evidence in order to overstate their talent, thereby imitating agents whose test score they *can* achieve. Namely, an agent (e,t) does not want to imitate an agent (e',t') with less evidence $e' < e$, more talent $t' > t$, and equal test score $\sigma(e',t') = \sigma(e,t)$ in order to get rewarded with probability $\Pi(e',t')$ instead of $\Pi(e,t)$. Notice that for any possible level of evidence $e' < e$ that agent (e,t) may reveal, because of condition (i), she will want to overstate her talent as much as possible (making sure that she will be able to pass the test), up to the point where $\sigma(e',t') = \sigma(e,t)$.

We have so far seen that conditions (i), (ii), and (iii) are necessary and sufficient for the agent not to have incentives to deviate in any of the following three ways: (a) present all her evidence but under-report her talent, (b) present all her evidence but overstate

[20] $\underline{e}(s)$ (resp. $\overline{e}(s)$) is the minimum (resp. maximum) level of evidence that an agent can have while achieving test score at least (resp. at most) s. That is, agents with evidence lower than $\underline{e}(s)$ score less than s even if they have talent $t = 1$. Analogously, agents with evidence higher than $\overline{e}(s)$ score more than s even if they have talent $t = 0$.

Figure 1: Directions of (weak) increase in $\Pi(e,t)$ in IC mechanisms



her talent, or (c) hide some of her evidence and overstate her talent, imitating agents whose test score she *can* achieve. To see why they are necessary and sufficient for IC, it remains to observe that these conditions also rule out the fourth type of deviations by the agent: hiding evidence and overstating talent to imitate agents whose test score she *cannot* achieve. To see this, notice that conditions (i), (ii), and (iii) combined imply that $\Pi(e,t) \geq \Pi(e,0) \geq \Pi(e',0) \geq (1 - T(e',t'))P(e',t',\emptyset)$ for any $e' < e$,[21] ensuring that (e,t) does not want to overstate her talent so much (to a point where $\sigma(e',t') > \sigma(e,t)$) that she fails the test.

Condition (iii) of Proposition 1 is satisfied with equality. Lemma 3 shows that when testing is costly and some talented agents are (optimally) rewarded with higher probability than untalented ones with the same level of evidence, then the optimal mechanism satisfies condition (iii) of Proposition 1 with equality. Under free testing or when it is not optimal to reward talented agents with higher probability, it is still without loss to constrain attention to mechanisms that satisfy condition (iii) of Proposition 1 with equality.

Lemma 3. Given any IC mechanism $M \equiv \langle T,P \rangle$, there exists an IC mechanism $M' \equiv \langle T',P' \rangle$ with $(1 - T'(e,t))P'(e,t,\emptyset) = \Pi'(e,0)$ for every (e,t) that is outcome-equivalent to M and has at most as high testing costs as M. For $c > 0$, if also $\Pi(e,t) > \Pi(e,0)$ for a positive measure of agent types, then M' has lower testing costs than M.

Here is the intuition behind this result. When $\Pi(e,0) > (1 - T(e,t))P(e,t,\emptyset)$, it means that untalented agent $(e,0)$ strictly prefers to not overstate her talent. This strict preference is due to overtesting of talented agents. Namely, $T(e,t)$ can be reduced and $P(e,t,\emptyset)$ can be increased keeping $\Pi(e,t)$ fixed and making $\Pi(e,0) = (1 - T(e,t))P(e,t,\emptyset)$.

[21]The second inequality follows from conditions (i) and (ii) combined.

Then, talented agents are tested with just high enough probability to prevent untalented agents from imitating them.

From now on we constrain attentions to mechanisms with $(1-T(e,t))P(e,t,\emptyset) = \Pi(e,0)$, or equivalently, $\Pi(e,t) = \Pi(e,0) + T(e,t)$, for every (e,t). Given that untalented agents are never tested, condition (iii) being satisfied with equality means that $(1-T(e,0))P(e,0,\emptyset) = \Pi(e,0) = (1-T(e,t))P(e,t,\emptyset)$ for every (e,t); that is, the probability of rewarding an agent without a test depends only on presented evidence. The total probability of rewarding the agent has two components: (i) a base probability $\Pi(e,0)$ of rewarding the agent for her evidence, without a test and (ii) an additional probability $T(e,t)$ of rewarding the agent for her talent, which (through testing) allows her to differentiate herself from less talented agents with the same level of evidence.

3.3 Optimal screening under free testing

We are now ready to characterize the optimal mechanisms under free testing (i.e., $c = 0$). Condition (iii) of Proposition 1 is immaterial since testing is free. For example, it is automatically satisfied if $P(e,t,\emptyset) = 0$ for every (e,t) (i.e., nobody is ever rewarded without a test).[22] The principal's objective function is $\int_0^1 \int_0^1 \Pi(e,t)u(e,t)f(e,t)dtde$, which can be written as

$$\int_0^1 \int_{\underline{e}(s)}^{\overline{e}(s)} \Pi(e,\tau(e,s))u(e,\tau(e,s))f(e,\tau(e,s))deds, \tag{2}$$

where instead of integrating over e and t, we integrate over (test score level) s and e. The principal's problem amounts to choosing $\Pi(e,\tau(e,s))$—seen as a function of (e,s) instead of (e,t)—non-decreasing in e (condition (ii) of Proposition 1) and s (condition (i) of Proposition 1) to maximize (2), which is linear (and thus, convex) in Π. Bauer's maximum principle then implies that there exists an extreme Π (i.e., an extreme point of the space of non-decreasing functions from $\{(e,s) \in [0,1]^2 : e \in [\underline{e}(s),\overline{e}(s)]\}$ to $[0,1]$) that maximizes (2). It is a standard result that an extreme Π maps each agent to either 0 or 1 (e.g., see Börgers, 2015).

Lemma 4. There exists an optimal deterministic mechanism (i.e., an optimal mechanism where $\Pi(e,t) \in \{0,1\}$ for all (e,t)).

3.3.1 Testing technology biased in favor of talent

We are now ready to derive the optimal mechanism. The first case that we consider is when the testing technology is biased in favor of talent in the sense that the test is more

[22]Under zero testing costs, there are trivially infinitely many ways to implement any IC mechanism. For instance, it does not matter whether the principal tests an agent or not before rejecting her.

sensitive to talent than talent is valuable to the principal. Definition 3 describes pro-t biased testing for any testing cost c.[23]

Definition 3. σ is pro-t biased if for every test score $s \in [0,1]$ there exists e_s such that if $e > e_s$ (resp. $e < e_s$) and $\sigma(e,t) = s$, then $u(e,t) > c$ (resp. $u(e,t) < c$).

This is a single-crossing condition. It says that iso-test-score curves cross the principal's indifference curve from $I_u^{\downarrow}(c)$ into $I_u^{\uparrow}(c)$. Clearly, if the principal's payoff from rewarding the agent is increasing along iso-test-score curves, then σ is pro-t biased.

Claim 1. If $u(e,\tau(e,s))$ is increasing in e over $e \in [\underline{e}(s),\overline{e}(s)]$ for every $s \in [0,1]$, then σ is pro-t biased (for any c). The condition is satisfied if $\frac{\partial u(e,t)/\partial e}{\partial u(e,t)/\partial t} > \frac{\partial \sigma(e,t)/\partial e}{\partial \sigma(e,t)/\partial t}$ for every (e,t).

Proposition 2 shows that when testing is (i) free and (ii) pro-t biased, then the principal can achieve the full information benchmark.

Proposition 2. Let $c = 0$ and assume that σ is pro-t biased. Then, $\Pi(e,t) = \mathbf{I}\left(u(e,t) > 0\right)$ is IC. Thus, the principal achieves the full information first-best.

When the principal only values evidence, he can trivially achieve the first best—much like in the case where talent was absent from the model. Namely, rewarding every agent with sufficient evidence to be of positive value to the principal is IC, because it does not create incentives for agents to understate their evidence and/or overstate their talent as only evidence is rewarded. Similarly, if the principal also values talent but less strongly than the test score depends on talent, agents do not have incentives to hide evidence in order to overstate their talent when the principal rewards every agent of positive value. Figure 3(a) presents the optimal mechanism under pro-t biased testing.

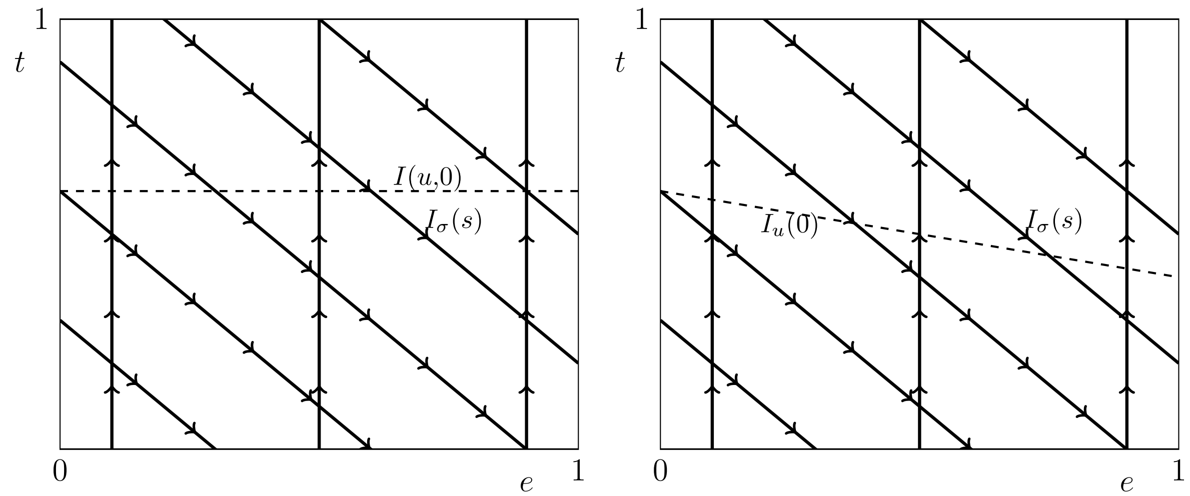
The principal can implement the first-best Π (and the one that we have restricted attention to given Lemma 3) is by setting $T(e,t) = \mathbf{I}(u(e,t) > 0 \wedge u(e,0) \leq 0)$ and $P(e,t,\emptyset) = \mathbf{I}(u(e,0) > 0)$. That is, agents that are not valuable to the principal truthfully report their type and are neither tested nor rewarded. Agents that are valuable but cannot prove so by presenting evidence e such that $u(e,0) > 0$ (which would prove that even if they have $t = 0$, they are valuable) are tested and then rewarded. Finally, agents that can prove that they are valuable by presenting evidence e such that $u(e,0) > 0$ do so and are rewarded without a test. Clearly, since testing is free, $T(e,t) = \mathbf{I}(u(e,t) > 0)$ and $P(e,t,\emptyset) = 0$ for every (e,t) is, for example, also optimal (and differs from the former implementation if there exists e such that $u(e,0) > 0$), as is testing every agent and rewarding only the valuable ones.

[23]The optimal mechanism under costly testing is studied in section 3.4.

Figure 2: *Not* achieving the first-best: testing technology biased in favor of evidence

(a) Only talent valuable to the principal

(b) Test less sensitive to talent than talent valuable to the principal



Note: the arrowed lines represent the directions of (weak) increase in $\Pi(e,t)$ in any IC mechanism. The dashed lines represent the principal's indifference curve $I_u(0)$.

3.3.2 Testing technology biased in favor of evidence

When some iso-test-score curves cross from $I_u^{\uparrow}(0)$ into $I_u^{\downarrow}(0)$, it is easy to see that the first-best is no longer achievable even if testing is free (i.e., $c = 0$). Indeed, Figure 2 shows that in that case, rewarding every agent in $I_u^{\uparrow}(0)$ and no agent in $I_u^{\downarrow}(0)$ is not IC, as it creates incentives for agents in $I_u^{\downarrow}(0)$ to hide evidence and imitate more talented agents.

But what *can* actually be achieved when the test is less sensitive to talent than talent is valuable to the principal? We now characterize the optimal mechanism when all iso-test-score curves cross the principal's indifference curve in the "wrong" direction (i.e., when σ is pro-e biased).

Definition 4. σ is pro-e biased if for every test score $s \in [0,1]$ there exists e_s such that if $e < e_s$ (resp. $e > e_s$) and $\sigma(e,t) = s$, then $u(e,t) > c$ (resp. $u(e,t) < c$).

Clearly, if the principal's payoff from rewarding the agent is decreasing along iso-test-score curves, then σ is pro-e biased.

Claim 2. If $u(e,\tau(e,s))$ is decreasing in e over $e \in [\underline{e}(s),\overline{e}(s)]$ for every $s \in [0,1]$, then σ is pro-e biased (for any c). The condition is satisfied if $\frac{\partial u(e,t)/\partial e}{\partial u(e,t)/\partial t} < \frac{\partial \sigma(e,t)/\partial e}{\partial \sigma(e,t)/\partial t}$ for every (e,t).

Proposition 3 describes the optimal mechanism when testing is (i) free and (ii) pro-e biased.

Proposition 3. Let $c = 0$. If σ is pro-e biased, then there exists an optimal mechanism with $\Pi(e,t) = \mathbf{I}(\sigma(e,t) \geq s^*)$.

Proposition 3 shows that in the optimal mechanism, agent (e,t) is rewarded if and only if $\sigma(e,t) \geq s^*$. An optimal way to implement this Π (and the one that we have restricted attention to given Lemma 3) is by setting $T(e,t) = \mathbf{I}(\sigma(e,t) \geq s^* \wedge e \leq \overline{e}(s^*))$ and $P(e,t,\emptyset) = \mathbf{I}(e > \overline{e}(s^*))$. That is, agents that cannot achieve test score (at least) s^* truthfully report their type and are neither tested nor rewarded. Agents that can achieve that test score and cannot prove this by presenting evidence $e > \overline{e}(s^*)$ (which would prove that even if they have $t = 0$, they can achieve test score s^*) are tested and then rewarded. Finally, agents that can prove that they can meet the test score threshold by presenting evidence $e \geq \overline{e}(s^*)$ do so and are rewarded without a test.[24]

Finding the optimal mechanism is remarkably simple. It amounts to solving a one-dimensional optimization problem on a closed interval with a continuous objective function. The principal needs to find $s^* \in \arg\max_{\widetilde{s}\in[0,1]} v(\widetilde{s})$, where

$$v(\widetilde{s}) := \int_{\widetilde{s}}^{1} \int_{\underline{e}(s)}^{\overline{e}(s)} u(e,\tau(e,s))f(e,\tau(e,s))deds$$

is continuous in $\widetilde{s}$.[25] When $s^* \in (0,1)$, it solves $\int_{\underline{e}(s^*)}^{\overline{e}(s^*)} u(e,\tau(e,s^*))f(e,\tau(e,s^*))de = 0$. The principal effectively chooses a threshold test score s^* and rewards every agent that can achieve this score. In choosing this threshold, he balances the Type I (i.e., rejecting agents in $I_u^{\uparrow}(0)$) and Type II (i.e., accepting agents in $I_u^{\downarrow}(0)$) errors. This trade-off can be seen in Figure 3(b).

Here is a sketch of the proof of Proposition 3. Because σ is pro-e biased, for any two types of zero value to the principal $(e,t),(e',t') \in I_u(0)$ with $e' > e$, $\sigma(e',t') \geq \sigma(e,t)$. But then, if $\sigma(e',t') \geq \sigma(e,t)$ and $e' > e$, then IC requires $\Pi(e',t') \geq \Pi(e,t)$. In other words, $\Pi(e,t)$ has to be non-decreasing as e increases along the $I_u(0)$ curve. Therefore, in any deterministic IC mechanism, there exists a threshold type on the $I_u(0)$ curve such that agents on the $I_u(0)$ curve with more (resp. less) evidence than the threshold type are rewarded (resp. not rewarded). Next, observe that IC requires that $\Pi(e,t)$ be non-decreasing along iso-test-score curves. Thus, having fixed $\Pi(e,t)$ along the $I_u(0)$ curve, keeping $\Pi(e,t)$ constant along iso-test-score curves maximizes the principal's payoff. That is because, on the part of an iso-test-score curve that lies below (resp. above) $I_u(0)$, the principal wants to make $\Pi(e,t)$ as low (resp. high) as possible but is constrained to set $\Pi(e,t)$ at least (resp. most) equal to its value on the curve $I_u(0)$ for that specific test score level. Condition (i) of Proposition 1 is automatically satisfied.

[24] Clearly, since testing is free, $T(e,t) = \mathbf{I}(\sigma(e,t) \geq s^*)$ and $P(e,t,\emptyset) = 0$ for every (e,t) is, for example, also optimal (and differs from the former implementation if $\overline{e}(s^*) < 1$), as is testing every agent and rewarding only those that pass the test score threshold s^*.

[25] The principal's problem reduces to this because all mechanisms with $\Pi(e,t) = \mathbf{I}(\sigma(e,t) \geq s^*)$ and appropriate T are IC.

Discussion. When seen against the results under pro-t biased testing (see Proposition 2), Proposition 3 reveals a stark contrast in the difficulty of, for example, hiring different types of employees. When skills and knowledge that can be proven through hard evidence are most valuable, the hiring process is easy. On the other hand, when talent—which is assessed by tests and interviews that are also sensitive to the candidate's training and knowledge—is most valuable, the hiring process is flawed, favoring some unworthy candidates with advanced training at the expense of agents with limited training who are, however, more valuable to the firm.

The revealed difference in the difficulty of hiring talented versus well-trained employees can be partly the reason behind the fact that firm survival rates increase with firm age and size (Evans, 1987; Dunne and Hughes, 1994; Farinas and Moreno, 2000; Agarwal and Gort, 2002; Bartelsman et al., 2005). To the extent that start-ups firms often face new challenges without the established procedures or clearly defined roles of older and larger firms, the success of a start-up will depend crucially on the ability of its employees to adapt and learn new tasks fast (i.e., $u(e,t)$ is very sensitive to t). On the other hand, the continued success of an established firm—where each employee's tasks are more clearly and narrowly defined—will depend (relatively) more on employee training, knowledge, and expertise (i.e., $u(e,t)$ is relatively more sensitive to e). Thus, hiring should be harder in start-ups than in established firms.

3.4 Optimal screening under costly testing

We now allow for a positive testing cost $c > 0$. The principal now needs compare the benefit of testing to its cost. The benefit of testing is that it increases rewarding accuracy: it allows the principal to reward talented agents with higher probability that untalented ones. The principal's objective function is $\int_0^1 \int_0^1 [\Pi(e,t)u(e,t) - cT(e,t)]\, f(e,t)dtde$. By Lemma 3, condition (iii) of Proposition 1 is satisfied with equality by the optimal mechanism, so in the objective function we can substitute $T(e,t) = \Pi(e,t) - \Pi(e,0)$. Then, the objective function reads

$$\int_0^1 \int_{\underline{e}(s)}^{\overline{e}(s)} [\Pi(e,\tau(e,s))(u(e,\tau(e,s)) - c) + c\Pi(e,0)]\, f(e,\tau(e,s))deds, \tag{3}$$

which is again linear in Π, so by Bauer's maximum principle, there exists an extreme Π—among all Π that are non-decreasing in e and s—that solves the principal's problem.

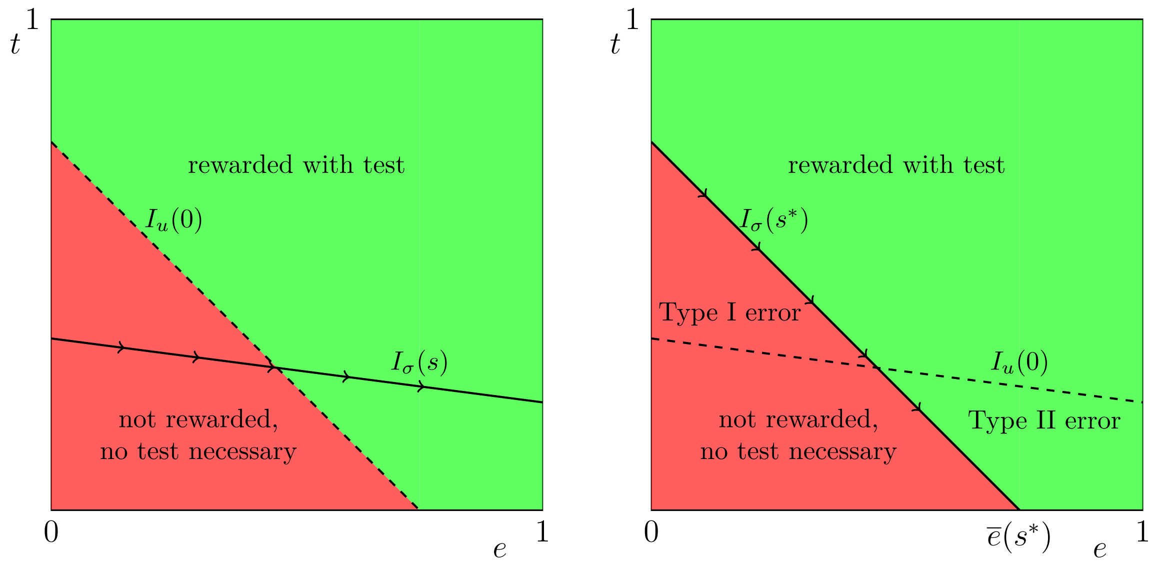
Lemma 5. There exists an optimal deterministic mechanism (i.e., an optimal mechanism where $\Pi(e,t) \in \{0,1\}$ for all (e,t)).

3.4.1 Testing technology biased in favor of talent

Proposition 4 characterizes the optimal mechanism under pro-t biased and costly testing.

Figure 3: The optimal mechanism under free testing

(a) Pro-t biased testing **(b)** Pro-e biased testing



Note: the dashed line represents the principal's indifference curve $I_u(0)$; the arrowed line represents an iso-test-score curve (at an arbitrary level s in the left panel and at level s^* in the right panel). The green (resp. red) area denotes the set of agents that are rewarded (resp. not rewarded) in the optimal mechanism. The Type I error corresponds to the part of the red area that lies above the dashed line. The Type II error corresponds to the part of the green area that lies below the dashed line.

Proposition 4. If σ is pro-t biased, then there exists an optimal mechanism with $\Pi(e,t) = \mathbf{I}(u(e,t) \geq c \text{ or } e \geq e^*)$ and $T(e,t) = \mathbf{I}(u(e,t) \geq c \text{ and } e < e^*)$ for some $e^* \in [0,1]$.

The principal's problem amounts to choosing a threshold level of evidence $e^* \in \arg\max_{\tilde{e}\in[0,1]} v(\tilde{e})$,[26] where

$$v(\tilde{e}) := \int_0^1 \int_0^{\tilde{e}} (u(e,t) - c)\mathbf{I}(u(e,t) \geq c) f(e,t) dedt + \int_0^1 \int_{\tilde{e}}^1 u(e,t) f(e,t) dedt.$$

Every agent with evidence $e \geq e^*$ evidence is rewarded without a test, while agents with evidence $e < e^*$ are tested and rewarded if their value $u(e,t)$ to the principal is higher than the cost c of testing. The remaining agents are neither tested nor rewarded. Figure 4(a) presents the structure of the optimal mechanism.

When $e^* \in (0,1)$, the first-order condition is

$$v'(e^*) = \int_0^1 (u(e^*,t) - c)\mathbf{I}(u(e^*,t) \geq c) f(e^*,t) dt - \int_0^1 u(e^*,t) f(e^*,t) dt = 0,$$

[26]The principal's problem reduces to this because all mechanisms with $\Pi(e,t) = \mathbf{I}(u(e,t) \geq c \text{ or } e \geq e^*)$ and $T(e,t) = \mathbf{I}(u(e,t) \geq c \text{ and } e < e^*)$ T for some $e^* \in [0,1]$ are IC.

or equivalently

$$v'(e^*) = \overbrace{-\int_0^1 u(e^*,t)\mathbf{I}(u(e^*,t) \leq 0)f(e^*,t)dt}^{>0\text{: gain from rejection of unworthy agents (ii)}} - \overbrace{\int_0^1 u(e^*,t)\mathbf{I}(0 < u(e^*,t) < c)f(e^*,t)dt}^{>0\text{: loss from rejection of worthy agents (iii)}}$$
$$- \underbrace{c\int_0^1 \mathbf{I}(u(e^*,t) \geq c)f(e^*,t)dt}_{>0\text{: loss from increase in testing costs (i)}} = 0.$$

An increase in the threshold e^* would lead to: (i) increased testing costs by making additional agents that lie above $I_u(c)$ get tested before being rewarded (who were rewarded without a test before the increase in e^*), (ii) the rejection without a test of additional agents that lie below $I_u(0)$ (who were rewarded without a test before the increase in e^*), but also (iii) the rejection without a test of additional agents that lie below $I_u(c)$ but above $I_u(0)$ (who were rewarded without a test before the increase in e^*). Channels (i) and (iii) negatively affect the principal's payoff, while channel (ii) tends to increase his payoff. In choosing the optimal threshold e^*, the principal trades off testing costs (i.e., effect (i)) with the increase in rewarding accuracy (i.e., the net effect of (ii) and (iii)).

Comparative statics. We now briefly discuss some comparative statics. For simplicity, assume that $e^* \in (0,1)$ is unique with the second-order condition of the principal's problem satisfied strictly and that some agents are optimally tested.[27] First, an increase in c causes the (combined) magnitude of channels (i) and (iii) to increase without affecting the magnitude of channel (ii).[28] Thus, e^* is decreasing in c; the more costly testing is, the more high-evidence agents are rewarded without a test. Particularly, $v'(e)$ is decreasing in c with $\partial v'(e)/\partial c = -\int_0^1 \mathbf{I}(u(e^*,t) \geq c)f(e^*,t)dt < 0$, and by the Implicit Function Theorem $de^*/dc = -\partial v'(e)/\partial c|_{e=e^*}/v''(e^*) < 0$. Second, the principal's optimal payoff is decreasing in c. Third, since the principal's objective function is independent of the testing technology σ, the optimal mechanism and payoff are the same under any two pro-t biased testing technologies with the same testing cost c.

3.4.2 Testing technology biased in favor of evidence

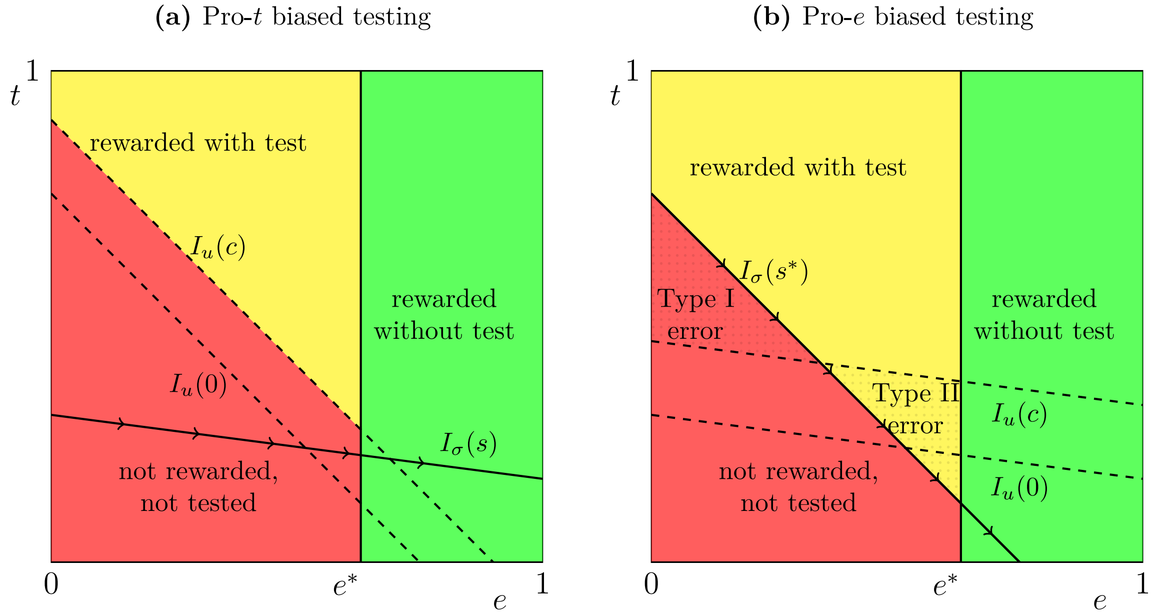
Proposition 5 characterizes the optimal mechanism under pro-e biased and costly testing.

Proposition 5. If σ is pro-e biased, then there exists an optimal mechanism with $\Pi(e,t) = \mathbf{I}(\sigma(e,t) \geq s^* \text{ or } e \geq e^*)$ and $T(e,t) = \mathbf{I}(\sigma(e,t) \geq s^* \text{ and } e < e^*)$ for some

[27] Namely, $e^* > 0$ and $u(e,t) > c$ for a positive measure of agents with $e < e^*$. This rules out the case $u(e,t) = e - \underline{q}$, where the principal only cares about evidence, in which case he does not test.

[28] In more detail, the partial derivative of $-c\int_0^1 \mathbf{I}(u(e^*,t) \geq c)f(e^*,t)dt$ with respect to c is $-\int_0^1 \mathbf{I}(u(e^*,t) \geq c)f(e^*,t)dt + cf(e^*,t')$ where t' is such that $u(e^*,t') = c$. The partial derivative of $\int_0^1 u(e^*,t)\mathbf{I}(0 < u(e^*,t) < c)f(e^*,t)dt$ with respect to c is $u(e^*,t')f(e^*,t') = cf(e^*,t') > 0$, which cancels out with the corresponding term in the derivative of $-c\int_0^1 \mathbf{I}(u(e^*,t) \geq c)f(e^*,t)dt$.

Figure 4: The optimal mechanism under costly testing



Note: the dashed line represents the principal's indifference curve $I_u(c)$: the principal is indifferent between rewarding with test and not rewarding without test agents on that curve. The arrowed line represents an iso-test-score curve (at an arbitrary level s in the left panel and at level s^* in the right panel). The green area denotes the set of agents that are rewarded without getting tested in the optimal mechanism. The red one denotes the set of agents that are neither tested nor rewarded in the optimal mechanism. The yellow area denotes the set of agents that are rewarded after getting tested in the optimal mechanism.

$(e^*,s^*) \in [0,1]^2$.

Finding an optimal mechanism is again remarkably simple. The principal's problem amounts to choosing threshold test score and evidence levels $(e^*,s^*) \in \arg\max_{(\widetilde{e},\widetilde{s})\in[0,1]^2} v(\widetilde{e},\widetilde{s})$, where

$$v(\widetilde{e},\widetilde{s}) := \int_{\widetilde{s}}^{1} \int_{\underline{e}(s)}^{\max\{\min\{\overline{e}(s),\widetilde{e}\},\underline{e}(s)\}} (\widetilde{u}(e,s) - c)\, \widetilde{f}(e,s)deds + \int_0^1 \int_{\max\{\underline{e}(s),\widetilde{e}\}}^{\max\{\overline{e}(s),\widetilde{e}\}} \widetilde{u}(e,s)\widetilde{f}(e,s)deds,$$

and $\widetilde{u}(e,s) := u(e,\tau(e,s))$ and $\widetilde{f}(e,s) := f(e,\tau(e,s))$.[29] Every agent with evidence $e \geq e^*$ is rewarded without a test, while agents with evidence $e < e^*$ are tested and rewarded if their (potential) test score is at least $\sigma(e,t) \geq s^*$. The remaining agents are neither tested nor rewarded. Figure 4(b) presents the structure of the optimal mechanism under pro-e biased testing.

When $\underline{e}(s^*) < e^*$ (i.e., some agents are rewarded after being tested) and $e^*,s^* \in (0,1)$,[30]

[29] The principal's problem reduces to this because all mechanisms with $\Pi(e,t) = \mathbf{I}(\sigma(e,t) \geq s^* \text{ or } e \geq e^*)$ and $T(e,t) = \mathbf{I}(\sigma(e,t) \geq s^* \text{ and } e < e^*)$ for some $(e^*,s^*) \in [0,1]^2$ are IC.

[30] Notice that $e^* \leq \overline{e}(s^*)$ (for if $e^* > \overline{e}(s^*)$ and $c > 0$, then reducing e^* would increase $v(e^*,s^*)$). For $c = 0$, $e^* = 1$ without loss.

the first-order conditions are

$$v_1(e^*,s^*) = \overbrace{-\int_0^{s^*} \widetilde{u}(e^*,s)\widetilde{f}(e^*,s)ds}^{\text{>0: gain from rejection of unworthy agents (i)}} - \overbrace{c\int_{s^*}^1 \widetilde{f}(e^*,s)ds}^{\text{>0: loss from increase in testing costs (ii)}} = 0,$$

$$v_2(e^*,s^*) = \underbrace{-\int_{\underline{e}(s^*)}^{e^*} \min\{\widetilde{u}(e,s^*)-c,0\}\widetilde{f}(e,s^*)de}_{\text{>0: gain from decrease in Type II error}} - \underbrace{\int_{\underline{e}(s^*)}^{e^*} \max\{\widetilde{u}(e,s^*)-c,0\}\widetilde{f}(e,s^*)de}_{\text{>0: loss from increase in Type I error}} = 0,$$

given that optimality requires $e^* \leq \overline{e}(s^*)$.[31] The principal chooses s^* considering the trade-off between Type I and Type II errors—conditional on the fact that only agents with evidence $e \geq e^*$ are rewarded without a test; for agents with evidence $e < e^*$, the principal chooses between rewarding after testing and rejecting without testing. The Type I error is due to the fact that the principal rejects without testing some agents whom he would prefer to reward after testing. The Type II error is due to the fact that the principal tests and rewards some agents that he would prefer to reject without testing. An increase in the threshold e^* would lead to: (i) the rejection of additional agents that lie below $I_u(0)$ (who were rewarded without a test before the increase in e^*) and (ii) increased testing costs by making additional agents that lie above $I_\sigma(s^*)$ get tested before being rewarded (who were rewarded without a test before the increase in e^*). Channel (ii) negatively affects the principal's payoff, while channel (i) tends to increase his payoff. In choosing the optimal threshold e^*, the principal trades off testing costs (i.e., effect (ii)) with the increase in rewarding accuracy (i.e., effect (i)).

Comparative statics. We now briefly discuss some comparative statics. For simplicity, assume that $s^*,e^* \in (0,1)$ are unique with the second-order condition of the principal's problem satisfied strictly and that some agents are optimally tested. Denote by $J(e^*,s^*)$ the Jacobian matrix of the first derivatives evaluated at (e^*,s^*), which is by assumption negative definite. Particularly, $v_{11}(e^*,s^*),v_{22}(e^*,s^*) < 0$ and $\det(J(e^*,s^*)) > 0$. Also, $v_{12}(e^*,s^*) = v_{21}(e^*,s^*) = -\left(\widetilde{u}(e^*,s^*) - c\right)\widetilde{f}(e^*,s^*) > 0$. First, the total derivatives of $e^*$and s^* with respect to c are:

$$\frac{de^*}{dc} \propto \overbrace{-v_{1c}(e^*,s^*)v_{22}(e^*,s^*)}^{\text{<0: direct effect of } c \text{ on } e^* \text{ due to increase in marginal testing costs}} + \overbrace{v_{2c}(e^*,s^*)v_{12}(e^*,s^*)}^{\text{>0: indirect effect of } c \text{ on } e^* \text{ through direct effect of } c \text{ on } s^*},$$

$$\frac{ds^*}{dc} \propto \underbrace{-v_{2c}(e^*,s^*)v_{11}(e^*,s^*)}_{\text{>0: direct effect of } c \text{ on } s^* \text{ due to increase in marginal testing costs}} + \underbrace{v_{1c}(e^*,s^*)v_{21}(e^*,s^*)}_{\text{<0: indirect effect of } c \text{ on } s^* \text{ through direct effect of } c \text{ on } e^*},$$

[31]If $e^* > \overline{e}(s^*)$, then decreasing e^* to make it equal to $\overline{e}(s^*)$ would decrease testing costs without changing the set of agents who are rewarded, thereby increasing the principal's payoff.

where $v_{1c}(e^*,s^*) = -\int_{s^*}^1 \widetilde{f}(e^*,s)ds < 0$ and $v_{2c}(e^*,s^*) = \int_{\underline{e}(s^*)}^{e^*} \widetilde{f}(e,s^*)de > 0$ are the partial derivatives of v_1 and v_2 with respect to c. An increase in the (marginal) cost c of testing tends to directly cause (i) e^* to decrease by magnifying the testing cost savings associated with a decrease in e^* and (ii) s^* to increase by magnifying the testing cost savings associated with an increase in e^*.[32] However, an increase in s^* tends to cause e^* to increase by reducing the marginal increase in testing costs associated with an increase in e^*. Conversely, an increase in e^* tends to cause s^* to increase by increasing the marginal (with respect to s^*) Type II error. Therefore, although an increase in c tends to directly cause e^* to fall and s^* to rise, the interaction between e^* and s^* works in the opposite direction making the net effect ambiguous. Still, we know that if s^* decreases in response to an increase in c, then e^* should also decrease and—the contrapositive—if e^* increases in response to an increase in c, then s^* should also increase. Second, the principal's optimal payoff is decreasing in c. Third, the optimal payoff is higher under less pro-e biased testing technologies. Namely, take any two pro-e biased testing technologies σ' and σ. If all iso-test-score curves of σ cross the iso-test-score curves of σ' from above (i.e., σ is less pro-e biased than σ'), then the principal's optimal payoff is higher under σ' than under σ.[33] Fourth, the principal's payoff should tend to increase with the correlation between evidence and talent. A strong (positive) correlation between e and t means that there are not many agents with high (resp. low) talent and low (resp. high) evidence, which implies that both Type I and Type II errors are small. As e and t become perfectly (positively) correlated, the principal achieves the first-best just by asking for evidence—regardless of his preferences and the testing technology.

Implementation of the optimal mechanism. We have so far restricted (without loss) attention to truth-telling mechanisms. However, the optimal mechanism under pro-e biased testing can be implemented in the following simple way. The principal gives the agent two paths to getting rewarded: either (i) provide evidence e^* and you will be rewarded without a test or (ii) take a test without providing any evidence, score at least s^*, and you will be rewarded. The first option is not always provided (e.g., when testing is free, the first option is not necessary in the optimal mechanism). Asking for evidence is useful to the principal (as long as $e^* < 1$). Last, notice that a similarly simple implementation of the optimal mechanism under pro-t biased testing is not possible. In

[32]Put differently, an increase in c can be seen to increase the marginal (with respect to s^*) Type II error and decrease the marginal Type II error, thereby tending to make s^* increase to equalize the magnitudes to the two errors.

[33]Comparative statics of s^* and e^* with respect to σ would have little value, since the optimal test score thresholds (which are determined simultaneously with the optinal evidence thresholds) under different testing technologies are not comparable, as they can only be interpreted with respect to their corresponding testing technologies.

that case, the principal needs to ask for evidence also from agents that are tested.[34]

4 Applications

In this section, I use the model to discuss hiring by prestigious employers, promotion decisions, college admissions, and academic job market hiring.

4.1 Hiring by prestigious employers

A job candidate's evidence e is her CV quality (e.g., high school quality, undergraduate institution quality and GPA, awards, distinctions, reference letters). t is her ability and drive not captured by e. A prestigious employer wants to decide whether to hire the candidate. Testing amounts to letting some other employer hire the candidate. Testing is costly because if the employer wants to then hire the candidate, he will have to poach her at a cost.

In the optimal mechanism, Ivy-Leaguers with high credentials get immediately hired by prestigious employers thanks to their evidence. On the other hand, talented candidates with education from lower-ranked institutions and lower grades have to go through less prestigious employers to prove their worth before they land a prestigious position. Also, if the candidates' performance in the less prestigious position is less sensitive to talent than talent is valuable in the more prestigious position—a natural assumption, then worthy candidates with low credentials are at a disadvantage also in the poaching stage.

4.2 Promotions

An employee of efficiency t has exerted effort e. $\sigma(e,t)$ is the employee's productivity, increasing in e and in t. Testing (by the employer/manager) amounts to verifying the employee's productivity $\sigma(e,t)$. The value to the principal of the agent that is not promoted (i.e., continues to work in her current position) is $\sigma(e,t)$. His value of the agent if promoted is $\widetilde{u}(e,t)$. Then, his problem is equivalent to the one in section 2 with $u(e,t) \coloneqq \widetilde{u}(e,t) - \sigma(e,t)$, as long as the difference $\widetilde{u}(e,t) - \sigma(e,t)$ is non-decreasing in both e and t. This condition on the difference can be interpreted to say that both effort and talent have a (weakly) higher marginal return in the higher position, which comes with increased responsibilities that allow the employee's talent and effort to have a larger impact.

Under differentiability and given Claims 1 and 2, the test is pro-t (resp. pro-e) biased if for every (e,t), $\partial u(e,t)/\partial e/(\partial u(e,t)/\partial t)$ is higher (resp. lower) than $\partial\sigma(e,t)/\partial e/(\partial\sigma(e,t)/\partial t)$,

[34]These observations on the implementation of optimal mechanisms also imply that under free testing (i.e., $c = 0$), if the principal (optimally) asks for evidence—which he does not need to do under pro-e biased testing, then he most likely values evidence (i.e., $u(e,t)$ is increasing in e).

or equivalently,

$$\frac{\partial\tilde{u}(e,t)/\partial e}{\partial\tilde{u}(e,t)/\partial t} \overset{\text{(resp. <)}}{>} \frac{\partial\sigma(e,t)/\partial e}{\partial\sigma(e,t)/\partial t},$$

that is, if the marginal rate of substitution of effort for talent is higher (resp. lower) in the production function of the new position than of the current one.

4.3 College admissions and standardized testing

A college applicant's evidence e is her high school quality, grades, private tutoring received, awards, and extracurricular activities. t is her "natural" ability or drive that is not captured by e. The college wants to decide whether to admit the applicant or not. Testing amounts to requiring the applicant to take the standardized test.[35]

In the optimal mechanism, if the standardized test is not sensitive enough to talent, then students can withhold evidence, which makes admission decisions imperfect at the expense of students with low evidence (e.g., those with limited access to quality education, tutoring, extracurricular activities, and opportunities to participate in competitions). Particularly, if colleges want diversity and only value talent (trying to control for the applicants' unequal backgrounds), then the above problem is necessarily present under standardized testing to the extent that applicants can pretend to be from a more modest background than they actually are. Students from a privileged background have an advantage over equally good—or even somewhat better—students from a more modest background.

Even if universities do not want to do diverse admissions by only valuing talent in applicants (but instead value the total ability of the candidate, part of which is due to nurture), then students from advantaged backgrounds are still favored over equally able students from disadvantaged backgrounds when the test is not sensitive enough to talent (compared to the total ability that colleges care about).

4.4 Academic job market talks

An academic job market candidate's research topic is comprised by a "mass" $b > 1$ of (uncountably infinitely many) problems.[36] $e \in [0,1]$ is the candidate's knowledge, the mass of problems which she has found answers to. t is her ability to think on her feet. More concretely, it is the probability with which she finds an answer on the spot to a problem that she has not already solved. After the candidate presents answers to a mass

[35]In this setting, the college does not condition the requirement to take a test on the candidate's report. However, when the college requires a test score, the optimal mechanism takes the same form as in the case $c = 0$.

[36]The analysis can apply to presentations more generally (e.g., by a start-up founder to a venture capital firm).

$e' \in [0,e]$ of problems and makes a claim about t, the hiring committee may test her. Testing amounts to posing to the candidate countably infinitely many problems randomly sampled from the mass of problems that the candidate has not already disclosed answers to.[37] Thus, if she presents answers to mass $e' \in [0,e]$ of problems and is tested, she will answer proportion

$$p(e,t,r) \coloneqq \frac{e - e' + (b-e)t}{b - e'}$$

of the problems posed to her. This is the sum of (i) the proportion $(e - e')/(b - e')$ of problems sampled from the set of problems that the candidate already has answers to (but has not disclosed them) and (ii) the proportion $(b - e)/(b - e')$ of problems sampled from the set of problems that the candidate does not already have answers to multiplied by the proportion t to which the candidate will find answers on the spot. $u(e,t)$ is the hiring committee's surplus from hiring the candidate (compared to the committee's outside option). Observing e' and $p(e,t,e')$ is equivalent to observing e' and $\sigma(e,t) \coloneqq e + (b-e)t$, so the committee's problem is equivalent to the problem that we have studied.

Different testing technologies can be interpreted as different values of b. An increase in the mass b of the universe of problems makes it less likely that the candidate will be asked a question that she already has an answer to (but has not presented), thereby making the test more sensitive to talent and increasing the principal's payoff.

5 Extensions and robustness

The section discusses two extensions of the model. First, it discusses the scenario where the principal has to pay a cost *before* the agent reports her type in order to design the test, which he will then (after the agent reports her type) choose whether to administer at an additional cost. Next, it discusses the case where evidence is not exogenous but rather endogenously produced by the agent before she interacts with the principal.

5.1 Optimal screening under alternative evidence structures

We now study optimal screening under two alternative scenarios: (i) the agent can present evidence also on talent or (ii) the agent cannot present evidence (on either dimension of her type).[38]

[37] This can be understood as there being a set of problems with cardinality equal to the cardinality of $\mathbb{R}$. There is no interdependence among the problems (e.g., the agent having the answer to a problem x carries no information with regard to whether she also has the answer to a problem y). Also, the agent is equally likely to have or find an answer to any of the problems. Thus, there is no need to identify problems with an index.

[38] The case where the agent can present evidence on t but not on e is a relabeling of the main model.

5.1.1 Optimal screening when the agent can present evidence also on talent

Consider the case where—apart from e—t also has an evidence structure. That is, agent (e,t) can report any $(e',t') \leq (e,t)$ but not $e' > e$ or $t' > t$. Then, the principal can achieve the full information first-best, inducing—without testing—every agent to present all her evidence on both e and t. The conclusion is the same if t is observed (at no cost) by the principal and e is evidence.

The comparison between this and the main model emphasizes (i) the difference in peoples' incentives to present evidence that is in principle (i.e., absent testing) favorable to them and (ii) how these incentives shape the principal's problem of evaluating them. The existence of an agent characteristic that is valuable to the principal but which (i) the agent cannot provide evidence on and (ii) the principal can only imperfectly test using a test that is overly (compared to the principal's preferences) sensitive to other (valuable) agent characteristics creates incentives for the agent to understate those other characteristics that she can actually provide favorable evidence on. This problem vanishes (i) if the agent can provide evidence on every characteristic (or if those that she cannot provide evidence on are observed by the principal) or (ii) if the test is sensitive only to the characteristic that the agent cannot provide evidence on.

These results are consistent with the finding that hiding one's effort is particularly prevalent among younger individuals. Effortless perfection (i.e., the need to seem perfect without apparent effort) and hiding one's effort have been documented among university students (Travers et al., 2015; Casale et al., 2016).The psychology literature has emphasized personality traits that may lie behind this finding. Namely, hiding effort has been identified as a unique expression of perfectionistic self-presentation (Flett et al., 2016). My model hints towards an alternative (or complementary) interpretation of this finding. If as an individual progresses in her career, her talent is revealed during all the evaluation stages that she goes through, then individuals that are further along in their career paths should have reduced incentives to hide their hard work.

5.1.2 Optimal screening when the agent cannot present evidence

Consider the case where the agent can present evidence on neither e nor t. That is, agent (e,t) can report any $(e',t') \in [0,1]^2$. We can still restrict attention to truthful mechanisms with pass-or-fail tests. Proposition 6 characterizes IC mechanisms.

Proposition 6. A mechanism $M \equiv \langle T,P \rangle$ is IC if and only if

(i) $\Pi(e,t)$ is non-decreasing in t for every e,

(ii) $\Pi(e,\tau(e,s))$ is constant in e over $e \in [\underline{e}(s),\bar{e}(s)]$ for every $s \in [0,1]$, and

(iii) $(1 - T(e,t))P(e,t,\emptyset) \leq \Pi(0,0)$ for every (e,t), where $\Pi(e,t) \equiv (1-T(e,t))P(e,t,\emptyset) + T(e,t)$.

Condition (i) is identical to the one in Proposition 6, where e is evidence. Condition (iii) is stronger (when combined with the other two conditions) than the corresponding condition (iii) of Proposition 6. It ensures that the *least* talented agent with the *least* evidence does not have incentives to over-report her talent and/or evidence to imitate an agent (e,t) whose test score she *cannot* achieve.[39] Put differently, in order to reward some agents with higher probability (than agent (0,0)), the principal needs to test those agents with high enough probability to prevent agent (0,0) from imitating them to get rewarded in case she is not tested. The condition is stricter than the one in Proposition 6 because now agents can also imitate agents with higher e to get rewarded in the case that they are not tested. Thus, that agents cannot present evidence on e enhances the need to test.

Last, condition (ii) makes sure that agents do not want to under- or over-report their e to imitate agents whose test score they *can* achieve. Namely, an agent (e,t) does not want to imitate an agent (e',t') with evidence $e' > e$ (resp. $e' < e$), talent $t' < t$ (resp. $t' > t$), and equal test score $\sigma(e',t') = \sigma(e,t)$ in order to get rewarded with probability $\Pi(e',t')$ instead of $\Pi(e,t)$. Notice that for any possible level of evidence e' that agent (e,t) may reveal, because of condition (i), she will want to report her talent to be as high as possible (making sure that she will be able to pass the test), up to the point where $\sigma(e',t') = \sigma(e,t)$. The condition is stricter than the one in Proposition 6 because now agents can not only understate but also overstate e. This nullifies the advantage that agents with high e have (relative to agents with the same test score but lower e) when they can present evidence.

The probability of getting rewarded without a test is the same for everyone. Lemma 6 shows that when testing is costly and some agents are (optimally) rewarded with higher probability than other ones, the optimal mechanism satisfies condition (iii) of Proposition 1 with equality. Under free testing or when it is not optimal to reward some agents with higher probability, it is still without loss to constrain attention to mechanisms that satisfy condition (iii) of Proposition 6 with equality.

Lemma 6. Given any IC mechanism $M \equiv \langle T,P \rangle$, there exists an IC mechanism $M' \equiv \langle T',P' \rangle$ with $(1-T'(e,t))P'(e,t,\emptyset) = \Pi'(0,0)$ for every (e,t) that is outcome-equivalent to M and has at most as high testing costs as M. For $c > 0$, if also $\Pi(e,t) > \Pi(0,0)$ for a positive measure of agent types, then M' has lower testing costs than M.

By Lemma 6 $\Pi(e,t) = \Pi(0,0) + T(e,t)$. Thus, the principal's objective function, $\int_0^1 \int_0^1 [\Pi(e,t)u(e,t) - cT(e,t)]\, f(e,t)dtde$, can be written as

$$\int_0^1 \int_{\underline{e}(s)}^{\overline{e}(s)} [\Pi(e,\tau(e,s))(u(e,\tau(e,s)) - c) + c\Pi(0,0)]\, f(e,\tau(e,s))deds, \tag{4}$$

[39] Combined with conditions (i) and (ii), this also means that no other agent has incentives to imitate an agent whose test score she cannot achieve.

which is linear in Π, so by Bauer's maximum principle, there exists an extreme Π (among $\Pi(e, \tau(e,s))$ that are constant in e and non-decreasing in s) that solves the principal's problem. Proposition 7 describes that extreme optimal mechanism.

Proposition 7. There exists an optimal mechanism with $\Pi(e,t) = \mathbf{I}(\sigma(e,t) \geq s^*)$ and $T(e,t) = \Pi(e,t) - \Pi(0,0)$ for some $s^* \in (0,1)$. That is, either

(i) $s^* = 0$, and every agent is rewarded without a test or

(ii) $s^* > 0$, and each agent (e,t) is (a) rewarded after getting tested if $\sigma(e,t) \geq s^*$ or (b) neither tested nor rewarded if $\sigma(e,t) < s^*$.

The inability of agents to present evidence on one of their attributes limits the set of IC mechanisms, thereby decreasing—in most cases—the principal's optimal payoff. Assume for simplicity that the optimal mechanism is unique. When the testing technology is pro-e biased, if some—but not all–agents are optimally rewarded without a test when e is actually evidence (i.e., the optimal evidence threshold for rewarding without a test lies strictly between 0 and 1), then the principal's payoff is lower if evidence is not available to the agents. When the testing technology is pro-t biased, if not all agents are optimally rewarded without a test when e is actually evidence, then the principal's payoff is lower if evidence is not available to the agents. Particularly, the principal now has to choose s^* trading-off Type I and Type II errors even when σ is pro-t biased. Pro-t biased tests are not inherently better than pro-e biased tests when the agents cannot present evidence on e.

The comparison between the baseline model and the case where the agent cannot present evidence implies the following about the "signal jamming" problem that arises in career concern models (e.g., see Holmström, 1999). Under—as in career concerns models—free monitoring of the employee's productivity,[40] if the employer can ask for hard evidence of effort, then the signal jamming problem is mitigated if productivity is sensitive enough to talent—compared to the employer's preferences for rewarding (e.g., promoting) the employee. However, when productivity is *not* sensitive enough to talent, then the signal jamming problem persists even if the employer can ask for evidence of effort. Particularly, agents have incentives to withhold evidence, which they should be paid information rents to reveal. Regardless of whether it is pro-t or -e biased, the more closely the test aligns with the principal's preferences, the higher the principal's optimal payoff is.

[40] The argument still holds as long as monitoring is not too costly. If it is, then no monitoring occurs, so there can be no signal jamming either.

5.2 Costly test design

Treating the testing technology σ as exogenous is reasonable in several applications. For example, in hiring by prestigious employers (section 4.1), the employee's production function in the less prestigious position is not chosen by the prestigious employer. In promotion decisions (section 4.2), the employee's production function in the current position depends on her current job description and responsibilities, which should mostly reflect the firm's regular operating needs rather than support the employer's promotion decisions.

However, in other cases (e.g., hiring decisions where testing amounts to actual tests and interviews), the principal may be able to choose the testing technology. How does his problem change in that case? Let there be a cost $C(\sigma)$ that the principal needs to pay before the interaction with the agent, so that she can use testing technology σ during the interaction with the agent. Indeed, it is reasonable that the principal needs to design a test (if she designs a test at all) *before* the interaction with the agent due to time constraints and the complexity of designing a test. During the interaction with the agent, the principal can only choose whether to administer the test at cost c. Then, the principal's problem can be solved in two steps: (i) finding the optimal mechanism for each possible testing technology $\sigma \in \Sigma$, and then (ii) choosing the optimal testing technology $\sigma^* \in \Sigma$ form the set Σ of conceivable testing technologies. The solution to the first step is the one we have already described.[41]

If tests that are more sensitive to talent are more expensive to devise, then our results imply that as long as the test is under-sensitive (compared to the principal's preferences) to talent, there are gains from increasing its sensitivity to it, which the principal will have to compare to the cost of making the test more sensitive to talent. The principal will want to make the tests at most as sensitive to talent as his preferences are, since tests that are overly sensitive to talent are as effective as those that are exactly aligned with the principal's preferences.[42]

However, when agents cannot present evidence on any of their attributes, the principal can always gain from finely calibrating the test's sensitivity (to the agent's attributes) to align it with his preferences. Regardless of whether it is pro-t or -e biased, the more closely the test aligns with the principal's preferences, the higher the principal's payoff is.

[41] That is, assuming that Σ contains only pro-t and pro-e biased testing technologies (and possibly a testing technology that exactly coincides with the principal's preferences). Also, it is easy to see that there are no gains from designing multiple tests to the extent that all tests have the same administration cost c.

[42] Still, if there is uncertainty over a test's properties, in a robust approach, the principal will not need to worry about making the test overly sensitive to talent.

5.3 Endogenous evidence production

If the agent produces evidence before the interaction with the principal, then in some applications, the principal may be able to affect the agent's evidence production by committing to a mechanism *before* the agent produces evidence. Indeed, in promotion decisions (section 4.2), the employer may use the prospect of promotion to incentivize the employee to exert effort.[43] Treating evidence as exogenous is more in line with other applications. For instance, in hiring decisions (section 4.1), a single employer has little labor market power to affect the candidate's (effort to obtain) credentials. Similarly, in college admissions (section 4.3), a single college cannot affect how hard high school students study.

Our characterization of the optimal mechanism then still applies—even if evidence is endogenous, as long as the principal cannot influence evidence production by committing ex ante to a mechanism. Let the agent's talent t follow a distribution with density g and support [0,1]. Taking as given the principal's mechanism, summarized by evidence and test score thresholds (e^*,s^*), the agent exerts costly effort $x \in \mathbb{R}_+$ to produce evidence.[44] Exerting effort x has cost $C_t(x)$, non-decreasing in x. Evidence is distributed, conditional on x, according to density function $h_x(e)$ with support [0,1]. Denote by $x^*(t)$ the equilibrium level of effort by type t. An equilibrium is a fixed point (x^*,e^*,s^*) where $x^* : [0,1] \to \mathbb{R}_+$ is a best-response to (e^*,s^*) and (e^*,s^*) is a best-response to x^* (i.e., the thresholds (e^*,s^*) that solve the principal's problem when the agent's type has density $f(e,t) = g(t)h_{x^*(t)}(e)$). (x^*,e^*,s^*) can be interpreted as a symmetric equilibrium where each of multiple "effort-taking" principals chooses thresholds (e^*,s^*).

While a detailed analysis of endogenous evidence production is beyond the scope of this paper, the following observation shows the importance of the fact that the optimal mechanism has been characterized under minimal assumptions on the agent's type distribution (i.e., that it admits a full-support density). In equilibrium, by exerting effort x, agent t will earn expected payoff $\int_{\min\{e^*,\varepsilon(t,s^*)\}}^1 h_x(e)de - C_t(x)$, where $\varepsilon(t,s)$ is implicitly given by $\sigma(\varepsilon(t,s),t) = s$. If, for example, $c = 0$, then $e^* = 1$ and so $x^*(t) = 0$ for every $t \geq \tau(0,s^*)$ or $t \leq \tau(1,s^*)$. That is, agents so talented that they are rewarded even without evidence and agents so untalented that they are not rewarded even if they present evidence $e = 1$ do not exert effort. More generally, the agent's incentives to exert effort—and thus effort itself—will often be non-monotone in t.[45] Thus, evidence and talent may be stochastically

[43] Still, if promotions are not the main motive for the employee to exert effort (e.g., a bonus could be the main motive), then effort can still be taken as approximately exogenous. For example, if the employee can obtain a higher position by changing employers, then the prospect of promotion in the current company may not significantly affect her effort.

[44] Notice that the optimal mechanism can always be summarized by these two thresholds. Under pro-t biased testing, there is only an evidence threshold.

[45] For example, let $x \in [0,1]$ with $C_t(x) := \xi(t)x^2/2$, where $\xi(t) > 0$ is decreasing in t, $u(e,t) := \gamma_u e + (1-\gamma_u)t - \underline{q}$, $\sigma(e,t) := \gamma_s e + (1-\gamma_s)t$, where $1 > \gamma_s > \gamma_u$, and $H_x(e) := -2xe(1-e) + e(2-e)$.

dependent in complicated ways.

6 Conclusion

This paper has proposed a model of bidimensional screening, where an agent (she) with two attributes—evidence and talent—presents evidence (i.e., verifiably discloses possibly part of her evidence) and is (possibly) tested at a cost by the principal (he), who then decides whether to reward the agent. The agent cannot unilaterally prove anything about her talent. The test delivers a mixed signal (i.e., the test score)—increasing in both evidence and talent—of the agent's type and the principal (weakly) values both evidence and talent in an agent. If the principal tests the agent, then the agent may have incentives to hide evidence—although the principal values evidence—to influence how the principal interprets the test result. Particularly, she may want to hide evidence to make the principal overestimate her talent.

This problem arises when the test (score) is less sensitive to talent than talent is valuable to the principal. In that case, the optimal mechanism features two types of inefficiencies, both of which favor high-evidence agents over low-evidence ones: (i) it rewards some undeserving agents without testing them but rather only by asking them to present a certain level of evidence, and (ii) even among agents that cannot meet that evidence threshold, it rewards (after testing) some undeserving agents with high evidence but low talent, while it rejects some deserving agents with high talent but low evidence. Remarkably, this is the structure of the optimal mechanism even when the principal *only* values talent. The principal still optimally rewards evidence even though it is worthless to him.

The results indicate how less worthy individuals with high credentials or effort to show are favored—by an optimal and objective evaluation mechanism—over more worthy ones, who have however lower credentials (or effort to show). Ivy-Leaguers are immediately hired by prestigious employers, while those from more modest backgrounds have to go through less prestigious employers to prove their worth before landing a prestigious position. Even controlling for the fact that they need to first take a less prestigious position, they may still be at a disadvantage when trying to transition to a more prestigious one. In college admissions, high school students from privileged backgrounds have an advantage over

Then,

$$x^*(t) = \begin{cases} 0 & \text{if } t \leq \dfrac{s^* - \gamma_s}{1-\gamma_s} \\ \dfrac{2[s^* - (1-\gamma_s)t][\gamma_s - s^* + (1-\gamma_s)t]}{\gamma_s^2 \xi(t)} & \text{if } t \in \left(\dfrac{s^* - \gamma_s}{1-\gamma_s}, \dfrac{s^*}{1-\gamma_s}\right) \\ 0 & \text{if } t \geq \dfrac{s^*}{1-\gamma_s}. \end{cases}$$

equally good (or even better) students from modest backgrounds—even if colleges value diversity and try to control for the applicants' unequal backgrounds but their evaluation mechanisms (e.g., standardized tests) are sensitive to the applicant's prior training. Last, hard-working employees with mediocre managerial skills are promoted to managerial positions over less hard-working ones who would, however, make better managers.

References

Agarwal, R. and Gort, M. (2002). Firm and product life cycles and firm survival. *American Economic Review*, 92(2):184–190.

Armstrong, M. (1996). Multiproduct nonlinear pricing. *Econometrica*, 64(1):51–75.

Ball, I. (2024). Scoring strategic agents. *arXiv preprint arXiv:1909.01888*.

Bartelsman, E., Scarpetta, S., and Schivardi, F. (2005). Comparative analysis of firm demographics and survival: evidence from micro-level sources in OECD countries. *Industrial and Corporate Change*, 14(3):365–391.

Bizzotto, J., Rüdiger, J., and Vigier, A. (2020). Testing, disclosure and approval. *Journal of Economic Theory*, 187:105002.

Börgers, T. (2015). *An introduction to the theory of mechanism design*. Oxford University Press, USA.

Cai, Y., Devanur, N. R., and Weinberg, S. M. (2019). A duality-based unified approach to bayesian mechanism design. *SIAM Journal on Computing*, 50(3):STOC16–160.

Carroll, G. (2017). Robustness and separation in multidimensional screening. *Econometrica*, 85(2):453–488.

Carroll, G. and Egorov, G. (2019). Strategic communication with minimal verification. *Econometrica*, 87(6):1867–1892.

Casale, S., Fioravanti, G., Rugai, L., Flett, G. L., and Hewitt, P. L. (2016). The interpersonal expression of perfectionism among grandiose and vulnerable narcissists: Perfectionistic self-presentation, effortless perfection, and the ability to seem perfect. *Personality and Individual Differences*, 99:320–324.

Daskalakis, C., Deckelbaum, A., and Tzamos, C. (2017). Strong duality for a multiple-good monopolist. *Econometrica*, 85(3):735–767.

Dunne, P. and Hughes, A. (1994). Age, size, growth and survival: Uk companies in the 1980s. *The Journal of Industrial Economics*, 42(2):115–140.

Dziuda, W. (2011). Strategic argumentation. *Journal of Economic Theory*, 146(4):1362–1397.

Evans, D. S. (1987). The relationship between firm growth, size, and age: Estimates for 100 manufacturing industries. *The Journal of Industrial Economics*, 35(4):567–581.

Farinas, J. C. and Moreno, L. (2000). Firms' growth, size and age: A nonparametric approach. *Review of Industrial organization*, 17:249–265.

Flett, G. L., Nepon, T., Hewitt, P. L., Molnar, D. S., and Zhao, W. (2016). Projecting perfection by hiding effort: supplementing the perfectionistic self-presentation scale with a brief self-presentation measure. *Self and Identity*, 15(3):245–261.

Frankel, A. and Kartik, N. (2019). Muddled information. *Journal of Political Economy*, 127(4):1739–1776.

Frankel, A. and Kartik, N. (2022). Improving information from manipulable data. *Journal of the European Economic Association*, 20(1):79–115.

Glazer, J. and Rubinstein, A. (2004). On optimal rules of persuasion. *Econometrica*, 72(6):1715–1736.

Green, J. R. and Laffont, J.-J. (1986). Partially verifiable information and mechanism design. *The Review of Economic Studies*, 53(3):447–456.

Grossman, S. J. (1981). The Informational Role of Warranties and Private Disclosure about Product Quality. *The Journal of Law and Economics*, 24(3):461–483.

Haghpanah, N. and Hartline, J. (2021). When is pure bundling optimal? *The Review of Economic Studies*, 88(3):1127–1156.

Hart, S., Kremer, I., and Perry, M. (2017). Evidence games: Truth and commitment. *American Economic Review*, 107(3):690–713.

Holmström, B. (1999). Managerial incentive problems: A dynamic perspective. *The review of Economic studies*, 66(1):169–182.

Kartik, N. (2009). Strategic communication with lying costs. *The Review of Economic Studies*, 76(4):1359–1395.

Kattwinkel, D. and Knoepfle, J. (2023). Costless information and costly verification: A case for transparency. *Journal of Political Economy*, 131(2):504–548.

Kleiner, A., Moldovanu, B., and Strack, P. (2021). Extreme points and majorization: Economic applications. *Econometrica*, 89(4):1557–1593.

Li, Y. (2020). Mechanism design with costly verification and limited punishments. *Journal of Economic Theory*, 186:105000.

Li, Y. (2021). Mechanism design with financially constrained agents and costly verification. *Theoretical Economics*, 16(3):1139–1194.

Manelli, A. M. and Vincent, D. R. (2007). Multidimensional mechanism design: Revenue maximization and the multiple-good monopoly. *Journal of Economic theory*, 137(1):153–185.

Milgrom, P. (2008). What the seller won't tell you: Persuasion and disclosure in markets. *Journal of Economic Perspectives*, 22(2):115–131.

Milgrom, P. R. (1981). Good News and Bad News: Representation Theorems and Applications. *The Bell Journal of Economics*, 12(2):380–391.

Mussa, M. and Rosen, S. (1978). Monopoly and product quality. *Journal of Economic theory*, 18(2):301–317.

Myerson, R. B. (1981). Optimal auction design. *Mathematics of operations research*, 6(1):58–73.

Perez-Richet, E. and Skreta, V. (2022). Test design under falsification. *Econometrica*, 90(3):1109–1142.

Rochet, J.-C. and Choné, P. (1998). Ironing, sweeping, and multidimensional screening. *Econometrica*, 66(4):783–826.

Rochet, J.-C. and Stole, L. A. (2003). The economics of multidimensional screening. *Econometric Society Monographs*, 35:150–197.

Shin, H. S. (1994). News Management and the Value of Firms. *The RAND Journal of Economics*, 25(1):58.

Sobel, J. (2020). Lying and deception in games. *Journal of Political Economy*, 128(3):907–947.

Travers, L. V., Randall, E. T., Bryant, F. B., Conley, C. S., and Bohnert, A. M. (2015). The cost of perfection with apparent ease: Theoretical foundations and development of the effortless perfectionism scale. *Psychological Assessment*, 27(4):1147.

Viscusi, W. K. (1978). A Note on "Lemons" Markets with Quality Certification. *The Bell Journal of Economics*, 9(1):277–279.

Yang, F. (2022). Costly multidimensional screening. *arXiv preprint arXiv:2109.00487*.

Yang, K. H. and Zentefis, A. K. (2024). Monotone function intervals: Theory and applications. *American Economic Review*, 114(8):2239–2270.

A Proofs

Proof of Lemma 1 Take an IC mechanism $M \equiv \langle T,P\rangle$. Construct the mechanism $M' \equiv \langle T',P'\rangle$ with (i) $P'_{at}(e,t) = 1$, (ii) $T'(e,t) = T(e,t)P_{at}(e,t) \leq T(e,t)$, and (iii) $P'(e,t,\emptyset) = (1-T(e,t))P(e,t,\emptyset)/(1-T'(e,t))$ for any (e,t).[46]

We have then that (a) $T'(e,t)P'_{at}(e,t) = T(e,t)P_{at}(e,t)$, (b) $(1-T'(e,t))P'(e,t,\emptyset) = (1-T(e,t))P(e,t,\emptyset)$ and (c) $\Pi'(e,t) = \Pi(e,t)$ for any (e,t). (a)-(c) combined imply that the problem of every agent type under M' is the same as it was under M'. (c) means that M' is outcome-equivalent to M.

Last, to see why the second part is true, notice that for $c > 0$, M' saves on testing costs compared to M if there exists (a positive measure of types) (e,t) with $T(e,t) > 0$ and $P_{at}(e,t) < 1$. **Q.E.D.**

Proof of Lemma 2 Let $M \equiv \langle T,P\rangle$ be an IC mechanism. Then, construct the mechanism $M' \coloneqq \langle T',P'\rangle$ with (i) $T'(e,0) = 0$ for every e and $T'(e,t) = T(e,t)$ for every (e,t) with $t > 0$, (ii) $P'(e,t,\emptyset) = P(e,t,\emptyset) + (T(e,t) - T'(e,t))\,(1 - P(e,t,\emptyset))$ for every (e,t), and (iii)

$$P'(e,t,s) \coloneqq \begin{cases} 0 & \text{if } s < \sigma(e,t) \\ P_{at}(e,t) & \text{if } s \geq \sigma(e,t) \end{cases}$$

for every (e,t) and $s \in [0,1]$.

If every type reports truthfully, M' rewards each agent type with the same probability that M does, so it remains to show that M' is IC.

By IC of M we have that for every (e,t)

$$(e,t) \in \arg\max_{(\widehat{e},\widehat{t}) \leq (e,1)} \left\{(1 - T(\widehat{e},\widehat{t}))P(\widehat{e},\widehat{t},\emptyset) + T(\widehat{e},\widehat{t})\mathbf{I}\left(\sigma(e,t) \geq \sigma(e',t')\right)\right\}. \tag{5}$$

By construction, we have that $T(\widehat{e},\widehat{t}) = T'(\widehat{e},\widehat{t})$ and $P(\widehat{e},\widehat{t},\emptyset) = P'(\widehat{e},\widehat{t},\emptyset)$ for every $(\widehat{e},\widehat{t})$ with $\widehat{t} > 0$, so no type (e,t) has incentives to imitate any type $(\widehat{e},\widehat{t})$ with $\widehat{t} > 0$ under mechanism M'. Also, for any (e,t) and any $(\widehat{e},\widehat{t})$ with $\widehat{t} = 0$ and $\widehat{e} \leq e$, $\sigma(e,t) \geq \sigma(\widehat{e},\widehat{t})$, which means that the payoff of (e,t) from reporting $(\widehat{e},\widehat{t})$ with $\widehat{t} = 0$ is equal to

$$(1 - T'(\widehat{e},\widehat{t}))P'(\widehat{e},\widehat{t},\emptyset) + T'(\widehat{e},\widehat{t})\mathbf{I}\left(\sigma(e,t) \geq \sigma(e',t')\right) =$$

[46] In $P'(e,t,\emptyset)$, if $T'(e,t) = 1$, cancel $(1-T(e,t))$ in the numerator and $(1-T'(e,t))$ in the denominator.

$$(1-0)P'(\widehat{e},\widehat{t},\emptyset) = P(\widehat{e},\widehat{t},\emptyset) + \left(T(\widehat{e},\widehat{t}) - T'(\widehat{e},\widehat{t})\right)\left(1 - P(\widehat{e},\widehat{t},\emptyset)\right) =$$
$$(1 - T(\widehat{e},\widehat{t}))P(\widehat{e},\widehat{t},\emptyset) + T(\widehat{e},\widehat{t}) = (1 - T(\widehat{e},\widehat{t}))P(\widehat{e},\widehat{t},\emptyset) + T(\widehat{e},\widehat{t})\mathbf{I}\left(\sigma(e,t) \geq \sigma(e',t')\right)$$
$$\leq (1 - T(e,t))P(e,t,\emptyset) + T(e,t),$$

where the inequality follows from (5). Thus, no type (e,t) has incentives to imitate any type $(\widehat{e},\widehat{t})$ with $\widehat{t} = 0$ under mechanism M'. We conclude that M' is IC. **Q.E.D.**

Proof of Proposition 1 Denote the total probability with which type (e,t) is rewarded if she reports $(\widehat{e},\widehat{t})$ (with $\widehat{e} \leq e$) by

$$\widetilde{P}(\widehat{e},\widehat{t}; e,t) := (1 - T(\widehat{e},\widehat{t}))P(\widehat{e},\widehat{t},\emptyset) + T(\widehat{e},\widehat{t})\mathbf{I}\left(\sigma(e,t) \geq \sigma(\widehat{e},\widehat{t})\right).$$

Also, define condition (iii') (a strengthening of condition (iii)) to say that $(1 - T(e,t))P(e,t,\emptyset) \leq \Pi(e',0)$ for every e,t,e' with $e \leq e'$.

Step 1: I first show that condition (i) is necessary for IC by showing the contrapositive. Assume that for some e,t_1,t_2 with $t_2 > t_1$, $\Pi(e,t_2) < \Pi(e,t_1)$. Then, IC of type (e,t_2) is violated, since $\widetilde{P}(e,t_1; e,t_2) = \Pi(e,t_1) > \Pi(e,t_2)$, that is, (e,t_2) can imitate (e,t_1) to (reach (e,t_1)'s test score threshold and) get rewarded with higher probability that she would if she truthfully reported her type.

Step 2: I now show that condition (iii') is necessary for IC by showing the contrapositive.[47] Assume that for some e,e',t with $e' \geq e$, $(1 - T(e,t))P(e,t,\emptyset) > \Pi(e',0)$. Then, IC of type $(e',0)$ is violated, since $\widetilde{P}(e,t; e',0) \geq (1 - T(e,t))P(e,t,\emptyset) > \Pi(e',0)$, that is, $(e',0)$ can imitate (e,t) to get rewarded with higher probability that she would if she truthfully reported her type (even if she cannot achieve (e,t)'s test score).

Step 3: I now show that provided that (i) and (iii') are satisfied, $\Pi(r, \tau(r,\sigma(e,t)))$ being non-decreasing in r over $r \in [\underline{e}(\sigma(e,t)),e]$ for every (e,t) is necessary and sufficient for IC.

IC of type (e,t) is satisfied if and only if

$$\max_{(\widehat{e},\widehat{t}) \leq (e,1)} \left[(1 - T(\widehat{e},\widehat{t}))P(\widehat{e},\widehat{t}; \emptyset) + T(\widehat{e},\widehat{t})\mathbf{I}\left(\sigma(e,t) \geq \sigma(\widehat{e},\widehat{t})\right)\right] = \Pi(e,t). \tag{6}$$

Assume that conditions (i) and (iii') are satisfied. Then, $\Pi(e,t) \geq \Pi(e,0) \geq (1 - T(\widehat{e},\widehat{t}))P(\widehat{e},\widehat{t},\emptyset)$ for any $(\widehat{e},\widehat{t})$ with $\widehat{e} \leq e$. Therefore, (6) is equivalent to

$$\max_{(\widehat{e},\widehat{t}) \in \{(x,y) \in [0,1]^2 : x \leq e \text{ and } \sigma(e,t) \geq \sigma(x,y)\}} \left[(1 - T(\widehat{e},\widehat{t}))P(\widehat{e},\widehat{t}; \emptyset) + T(\widehat{e},\widehat{t})\right] = \Pi(e,t). \tag{7}$$

Given that $\Pi(e,t)$ is non-decreasing in t (condition (i)), (7) can equivalently be written as

$$\max_{r \in [\underline{e}(\sigma(e,t)),e]} \{[1 - T(r, \tau(r,\sigma(e,t)))]P(r,\tau(r,\sigma(e,t)),\emptyset) + T(r,\tau(r,\sigma(e,t)))\} = \Pi(e,t)$$

[47]That $P(e,0,\emptyset) = \Pi(e,0)$ follows from $T(e,0) = 0$.

or equivalently,

$$e \in \underset{r\in[\underline{e}(\sigma(e,t)),e]}{\arg\max}\ \Pi(r,\tau(r,\sigma(e,t))). \tag{8}$$

Thus, IC is satisfied for every type if and only if for every (e,t), (8) is satisfied. This is true if and only if $\Pi(r,\tau(r,\sigma(e,t)))$ is non-decreasing in r for $r \in [\underline{e}(\sigma(e,t)),e]$ for every (e,t).

That the latter is sufficient for (8) to hold for every (e,t) is immediate. I show necessity by showing the contrapositive. Assume that for some (e,t), $\Pi(r,\tau(r,\sigma(e,t)))$ is *not* non-decreasing in r for $r \in [\underline{e}(\sigma(e,t)),e]$. That is, for some (e,t) there exist r_1, r_2 with $\underline{e}(\sigma(e,t)) \leq r_1 < r_2 \leq e$ such that $\Pi(r_2,\tau(r_2,\sigma(e,t))) < \Pi(r_1,\tau(r_1,\sigma(e,t)))$. Then,

$$r_2 \notin \underset{x\in[\underline{e}(\sigma(e,t)),r_2]}{\arg\max}\ \Pi(x,\tau(x,\sigma(e,t))).$$

Namely, IC of type $(r_2,\tau(r_2,\sigma(e,t)))$ is violated, as she prefers to imitate type $(r_1,\tau(r_1,\sigma(e,t)))$.

Step 4: It is easy to see that $\Pi(r,\tau(r,\sigma(e,t)))$ being non-decreasing in r over $r \in [\underline{e}(\sigma(e,t)),e]$ for every (e,t) is equivalent to condition (ii).

Step 5: Finally, notice that provided that conditions (i) and (ii) hold, conditions (iii) and (iii') are equivalent. That (iii') implies (iii) is immediate. We will show that the opposite direction also holds. Assume that conditions (i), (ii), and (iii) hold. Then, for any e,e',t with $e' \geq e$

$$\Pi(e',0) \geq \Pi(e,\tau(e,\sigma(e',0))) \geq \Pi(e,0) \geq (1-T(e,t))P(e,t,\emptyset),$$

where the first inequality follows from condition (ii),[48] the second from condition (i), and the third from condition (iii). **Q.E.D.**

Proof of Lemma 3 Take any IC mechanism $M \equiv \langle T,P\rangle$. Condition (iii) of Proposition 1 says that $\Pi(e,0) \geq (1-T(e,t))P(e,t,\emptyset)$ for any (e,t). Then, construct the mechanism $M' \coloneqq \langle T',P'\rangle$ with[49]

$$\begin{aligned} T'(e,t) &\coloneqq \Pi(e,t) - \Pi(e,0) = (1-T(e,t))P(e,t,\emptyset) + T(e,t) - \Pi(e,0) \\ &\leq \Pi(e,0) + T(e,t) - \Pi(e,0) = T(e,t), \quad \text{and} \\ P'(e,t,\emptyset) &\coloneqq \frac{\Pi(e,0)}{1-\Pi(e,t)+\Pi(e,0)} \geq \frac{(1-T(e,t))P(e,t,\emptyset)}{1-\Pi(e,t)+(1-T(e,t))P(e,t,\emptyset)} = P(e,t,\emptyset) \end{aligned}$$

for every (e,t), where the inequalities follow from $\Pi(e,0) \geq (1-T(e,t))P(e,t,\emptyset)$.

[48]The first inequality assumes that $e \geq \underline{e}(\sigma(e',0))$. If this is not the case, using conditions (i) and (ii) iteratively, we can still show that $\Pi(e',0) \geq \Pi(e,0)$.

[49]For (e,t) such that $\Pi(e,t) = 1$ and $\Pi(e,0) = 0$, set $P'(e,t,\emptyset) = 0$.

By construction we have that $\Pi'(e,t) = \Pi(e,t)$ for every (e,t), so M' satisfies conditions (i) and (ii) of Proposition 1. By construction, we also have that for every (e,t)

$$\Pi'(e,0) = \Pi(e,0) = (1 - T'(e,t))P'(e,t,\emptyset),$$

so M' also satisfies condition (iii) of Proposition 1. Therefore, M' is IC.

Last, to see why the second part is true, notice that for $c > 0$, M' saves on testing costs compared to M if there exists (a positive measure of) (e,t) with $P(e,t,\emptyset)(1 - T(e,t)) < \Pi(e,0)$, since $T'(e,t) < T(e,t)$ for such (e,t). **Q.E.D.**

Proof of Lemma 5 It is useful to look at the principal's choice as a function $\Pi(e,\tau(e,s))$ of (e,s). The objective function (2) is continuous and linear (and thus, convex) in Π. By Bauer's maximum principle, it follows that there exists a maximizing function $(e,s) \to \Pi(e,\tau(e,s))$ that is an extreme point of the set of non-decreasing functions from $\{(e,s) \in [0,1]^2 : e \in [\underline{e}(s),\bar{e}(s)]\}$ to $[0,1]$.[50] Last, a function $(e,s) \to \Pi(e,\tau(e,s))$ is an extreme point of that set if and only if $\Pi(e,\tau(e,s)) \in \{0,1\}$ for all (e,s) is its domain.[51] The proof of this part is analogous to the one of Lemma 2.7 in Börgers (2015).

If direction: consider any non-decreasing (in e and s) Π with $\Pi(e,\tau(e,s)) \in \{0,1\}$ for all (e,s), and take any function $g : \{(e,s) \in [0,1]^2 : e \in [\underline{e}(s),\bar{e}(s)]\} \to \mathbb{R}$ such that $g(e^*,s^*) \neq 0$ for some (e^*,s^*). If $g(e^*,s^*) > 0$ and $\Pi(e^*,\tau(e^*,s^*)) = 0$, then $\Pi(e^*,\tau(e^*,s^*)) - g(e^*,s^*) < 0$. If $g(e^*,s^*) > 0$ and $\Pi(e^*,\tau(e^*,s^*)) = 1$, then $\Pi(e^*,\tau(e^*,s^*)) + g(e^*,s^*) > 1$. Similarly, if $g(e^*,s^*) < 0$ and $\Pi(e^*,\tau(e^*,s^*)) = 0$, then $\Pi(e^*,\tau(e^*,s^*)) + g(e^*,s^*) < 0$. If $g(e^*,s^*) < 0$ and $\Pi(e^*,\tau(e^*,s^*)) = 1$, then $\Pi(e^*,\tau(e^*,s^*)) - g(e^*,s^*) > 1$. Thus, Π is an extreme point.

Only if direction: now consider any non-decreasing (in e and s) Π with $\Pi(e^*,\tau(e^*,s^*)) \notin (0,1)$ for some (e^*,s^*). Construct function g as follows. $g(e,s) \coloneqq \Pi(e,\tau(e,s))$ for every (e,s) such that $\Pi(e,\tau(e,s)) \leq 1/2$ and $g(e,s) \coloneqq 1 - \Pi(e,\tau(e,s))$ for every (e,s) such that $\Pi(e,\tau(e,s)) > 1/2$. $g(e^*,s^*) \in (0,1)$, so $g \neq 0$. Consider the function $(e,s) \mapsto \Pi(e,\tau(e,s)) + g(e,s)$. Take any e_1,e_2,s with $e_2 \geq e_1$ and observe that if $\Pi(e_2,\tau(e_2,s)) > 1/2$, then

$$\Pi(e_2,\tau(e_2,s)) + g(e_2,s) = 1 \geq \Pi(e_1,\tau(e_1,s)) + g(e_1,s)$$

since by construction $\Pi(e,\tau(e,s)) + g(e,s) \leq 1$ for every (e,s), while if $\Pi(e_2,\tau(e_2,s)) \leq 1/2$, then (since Π is non-decreasing) $\Pi(e_1,\tau(e_1,s)) \leq \Pi(e_2,\tau(e_2,s)) \leq 1/2$, and so

$$\Pi(e_2,\tau(e_2,s)) + g(e_2,s) = 2\Pi(e_2,\tau(e_2,s)) \geq 2\Pi(e_1,\tau(e_1,s)) = \Pi(e_1,\tau(e_1,s)) + g(e_1,s).$$

Similarly, it can be seen that $\Pi(e,\tau(e,s_2)) + g(e,s_2) \geq \Pi(e,\tau(e,s_1)) + g(e,s_1)$ for any

[50] Observe that this set of functions is convex and compact in the norm topology.

[51] More precisely, this should hold for almost all (e,s) is its domain (see Börgers, 2015).

s_1,s_2,e with $s_2 \geq s_1$. Also $\Pi(e,\tau(e,s)) + g(e,s) \in [0,1]$ for every (e,s). Therefore, the function $(e,s) \mapsto \Pi(e,\tau(e,s)) + g(e,s)$ lies in the set of non-decreasing functions from $\{(e,s) \in [0,1]^2 : e \in [\underline{e}(s),\overline{e}(s)]\}$ to $[0,1]$. Similarly, it can be seen that the function $(e,s) \mapsto \Pi(e,\tau(e,s)) - g(e,s)$ lies in the set of non-decreasing functions from $\{(e,s) \in [0,1]^2 : e \in [\underline{e}(s),\overline{e}(s)]\}$ to $[0,1]$. Thus, Π is not an extreme point. **Q.E.D.**

Proof of Claim 1 The first part is immediate. The total derivative of $u(e,\tau(e,s))$ with respect to e is equal to

$$\begin{aligned}\frac{du(e,\tau(e,s))}{de} &= \frac{\partial u(e,\tau(e,s))}{\partial e} + \frac{\partial \tau(e,s)}{\partial e} \left.\frac{\partial u(e,t)}{\partial t}\right|_{t=\tau(e,s)} \\ &= \frac{\partial u(e,t)}{\partial e} + \frac{\partial \sigma(e,t)/\partial e}{\partial \sigma(e,t)/\partial t} \left.\frac{\partial u(e,t)}{\partial t}\right|_{t=\tau(e,s)},\end{aligned}$$

and the second part follows. **Q.E.D.**

Proof of Proposition 2 We need to show that $\Pi(e,t) = \mathbf{I}\left(u(e,t) > 0\right)$ satisfies conditions (i) and (ii) of Proposition 1.

Condition (i): Since $\Pi(e,t) \in \{0,1\}$ for every (e,t), it suffices to show that for any (e,t), if $\Pi(e,t) = 1$, then $\Pi(e,t') = 1$ for every $t' \geq t$. Indeed, we have that for any (e,t)

$$\Pi(e,t) = 1 \implies u(e,t) > 0 \implies u(e,t') > 0 \text{ for every } t' \geq t,$$

where the second implication follows since $u(e,t)$ is non-decreasing in t.

Condition (ii): Similarly, it suffices to show that for any (r,s), if $\Pi(r,\tau(r,s)) = 1$, then $\Pi(r',\tau(r',s)) = 1$ for every $r' \in [r,\overline{e}(s)]$. Indeed, we have that for any (r,s), $\Pi(r,\tau(r,s)) = 1$ implies that $u(r,\tau(r,s)) > 0$, which in turn implies that $u(r',\tau(r',s)) > 0$ for every $r' \in [r,\overline{e}(s)]$.

To see why the last part follows, assume instead that $u(r',\tau(r',s)) \leq 0$ for some $r' \in [r,\overline{e}(s)]$. Particularly, it must be $r' > r$. Since σ is pro-t biased, there exists e_s such that if $e > e_s$ (resp. $e \leq e_s$) and $\sigma(e,t) = s$, then $u(e,t) > 0$ (resp. $u(e,t) \leq 0$). We have that $u(r',\tau(r',s)) \leq 0$, so σ being pro-t biased implies that $r' \leq e_s$. But $r' > r$, so $r < e_s$, and since $\sigma(r,\tau(r,s)) = s$, σ being pro-t biased implies that $u(r,\tau(r,s)) \leq 0$, a contradiction. **Q.E.D.**

Since u satisfies t-heavy single crossing for every test score s, there exists function $\overline{e} : [0,1] \to [0,1]$ such that for every $s \in [0,1]$, if $e < \overline{e}(s)$ (resp. $e > \overline{e}(s)$) and $(e,t) \in I(\hat{t},s)$, then $u(e,t) > 0$ (resp. $u(e,t) < 0$).

Proof of Proposition 3 *Step 1:* In definition 4 of pro-e biased testing, for s such that $u(e,t) > c = 0$ (resp. $u(e,t) \leq 0$) for every $(e,t) \in I_\sigma(s)$, e_s is not uniquely defined. In that case, for s such that $u(e,t) > 0$ (resp. $u(e,t) \leq 0$) for every $(e,t) \in I_\sigma(s)$, set $e_s = \overline{e}(s)$ (resp. $e_s = \underline{e}(s)$). We will show that (under pro-e biased testing) e_s is non-decreasing in s. Take any $\underline{s},\overline{s} \in [0,1]$ with $\overline{s} > \underline{s}$, and define $S := (e_{\overline{s}},e_{\underline{s}}) \cap [\underline{e}(\overline{s}),\overline{e}(\overline{s})] \cap [\underline{e}(\underline{s}),\overline{e}(\underline{s})]$.

Step 1, case 1: If $S = \emptyset$, then $e_{\underline{s}} \leq e_{\overline{s}}$. To see this, consider the following two subcases.

Step 1, case 1(a): if $\underline{e}(\overline{s}) \geq \overline{e}(\underline{s})$, then $e_{\underline{s}} \leq \overline{e}(\underline{s}) \leq \underline{e}(\overline{s}) \leq e_{\overline{s}}$, so $e_{\underline{s}} \leq e_{\overline{s}}$, a contradiction.

Step 1, case 1(b): if $\underline{e}(\overline{s}) < \overline{e}(\underline{s})$, then $S = (e_{\overline{s}},e_{\underline{s}}) \cap [\underline{e}(\overline{s}),\overline{e}(\underline{s})]$. Since $S = \emptyset$, either $\underline{e}(\overline{s}) \geq e_{\underline{s}}$ or $\overline{e}(\underline{s}) \leq e_{\overline{s}}$. If $\underline{e}(\overline{s}) \geq e_{\underline{s}}$, then $e_{\underline{s}} \leq \underline{e}(\overline{s}) \leq e_{\overline{s}}$, so $e_{\underline{s}} \leq e_{\overline{s}}$, a contradiction. Similarly, if $\overline{e}(\underline{s}) \leq e_{\overline{s}}$, then $e_{\underline{s}} \leq \overline{e}(\underline{s}) \leq e_{\overline{s}}$, so $e_{\underline{s}} \leq e_{\overline{s}}$, a contradiction.

Step 1, case 2: We now prove by contradiction that if $S \neq \emptyset$, then $e_{\underline{s}} \leq e_{\overline{s}}$. To this end, assume that $S \neq \emptyset$ and $e_{\underline{s}} > e_{\overline{s}}$. Given that $S \neq \emptyset$, we can take some $e^* \in S$. Since $e^* \in [\underline{e}(\underline{s}),\overline{e}(\underline{s})]$ and σ is continuous, there exists $t^* \in [0,1]$ such that $\sigma(e^*,t^*) = \underline{s}$. Since σ is pro-e biased and $e^* < e_{\underline{s}}$, it follows that $u(e^*,t^*) > 0$. Similarly, since σ is pro-e biased, $e^* > e_{\overline{s}}$, and $e^* \in [\underline{e}(\overline{s}),\overline{e}(\overline{s})]$, there exists $t^{**} \in [0,1]$ such that $\sigma(e^*,t^{**}) = \overline{s}$ and $u(e^*,t^{**}) \leq 0$. Also, because $\overline{s} > \underline{s}$ and $\sigma(e,t)$ is increasing in t, $t^{**} > t^*$. Overall, we have $t^{**} > t^*$ and $u(e^*,t^*) > 0 \geq u(e^*,t^{**})$, a contradiction to $u(e,t)$ being non-decreasing in t.

Step 2: Given e_s, define also t_s implicitly given by $\sigma(e_s,t_s) = s$. We have then that for every test score $s \in [0,1]$, (e_s,t_s) is the "threshold" agent that lies on the iso-test-score curve $I_\sigma(s)$. That is, any other agent (e,t) on that iso-test-score curve with $e < e_s$ (resp. $e > e_s$) gives—if rewarded—a positive (resp. negative) payoff to the principal.

We divide the problem of finding an optimal IC mechanism in three parts. First, we fix an arbitrary "partial" IC mechanism $s \mapsto \Pi(e_s,t_s)$ for every $s \in [0,1]$. Then, we complete that partial IC mechanism (i.e., we assign a value to $\Pi(e,t)$ for every (e,t) for which $\Pi(e,t)$ has no been assigned a value in the first step), so that the complete mechanism is IC and optimal given the fixed partial mechanism. Finally, we find an optimal partial mechanism.

Step 3: Fix the value of $\Pi(e_s,t_s)$ for every $s \in [0,1]$ such that these values are part of some IC mechanism.[52] Given that e_s is non-decreasing in s, by Proposition 1 the values of $\Pi(e_s,t_s)$ are part of some IC mechanism if and only if $\Pi(e_s,t_s)$ is non-decreasing in s. Therefore, by Proposition 5, there exists an optimal mechanism with $\Pi(e_s,t_s) = \mathbf{I}(s \geq \underline{s})$ for some $\underline{s} \in [0,1]$.

Step 4: It follows then that for IC to be satisfied by the complete mechanism, it must be that (i) $\Pi(e,t) = 1$ for every (e,t) such that $e > e_s$ and $\sigma(e,t) = s$ for some $s \geq \underline{s}$ and (ii) $\Pi(e,t) = 0$ for every (e,t) such that $e < e_s$ and $\sigma(e,t) = s$ for some $s < \underline{s}$. Also, since (e_s,t_s) is the "threshold" agent, the principal wants to make $\Pi(e,t)$ as high (resp. low) as possible for every (e,t) such that $e < e_s$ (resp. $e > e_s$). Thus, given the IC constraint, it is optimal to set (i) $\Pi(e,t) = 1$ for every (e,t) such that $e < e_s$ and $\sigma(e,t) = s$ for some $s \geq \underline{s}$ and (ii) $\Pi(e,t) = 0$ for every (e,t) such that $e > e_s$ and $\sigma(e,t) = s$ for some $s < \underline{s}$. **Q.E.D.**

[52]That is, fix the value of $\Pi(e_s,t_s)$ for every $s \in [0,1]$ to be such that there exists IC $\Pi : [0,1]^2 \to [0,1]$ that agrees with the values of $\Pi(e_s,t_s)$ for every $s \in [0,1]$.

Proof of Proposition 6 Denote the total probability with which type (e,t) is rewarded if she reports $(\widehat{e},\widehat{t})$ (with $\widehat{e} \leq e$) by

$$\widetilde{P}(\widehat{e},\widehat{t};e,t) := (1 - T(\widehat{e},\widehat{t}))P(\widehat{e},\widehat{t},\emptyset) + T(\widehat{e},\widehat{t})\mathbf{I}\left(\sigma(e,t) \geq \sigma(\widehat{e},\widehat{t})\right).$$

Also, define condition (iii') (a strengthening of condition (iii)) to say that $(1 - T(e,t))P(e,t,\emptyset) \leq \Pi(e',0)$ for every e,t,e'.

Step 1: I first show that condition (i) is necessary for IC by showing the contrapositive. Assume that for some e,t_1,t_2 with $t_2 > t_1$, $\Pi(e,t_2) < \Pi(e,t_1)$. Then, IC of type (e,t_2) is violated, since $\widetilde{P}(e,t_1;e,t_2) = \Pi(e,t_1) > \Pi(e,t_2)$.

Step 2: I now show that condition (iii') is necessary for IC by showing the contrapositive. Assume that for some e,e',t, $(1 - T(e,t))P(e,t,\emptyset) > \Pi(e',0)$. Then, IC of type $(e',0)$ is violated, since $\widetilde{P}(e,t;e',0) \geq (1 - T(e,t))P(e,t,\emptyset) > \Pi(e',0)$.

Step 3: I now show that provided that (i) and (iii') are satisfied, $\Pi(r,\tau(r,\sigma(e,t)))$ being constant in r over $r \in [\underline{e}(\sigma(e,t)),e]$ for every (e,t) is necessary and sufficient for IC.

IC of type (e,t) is satisfied if and only if

$$\max_{(\widehat{e},\widehat{t}) \leq (1,1)} \left[(1 - T(\widehat{e},\widehat{t}))P(\widehat{e},\widehat{t};\emptyset) + T(\widehat{e},\widehat{t})\mathbf{I}\left(\sigma(e,t) \geq \sigma(\widehat{e},\widehat{t})\right)\right] = \Pi(e,t). \tag{9}$$

Assume that conditions (i) and (iii') are satisfied. Then, $\Pi(e,t) \geq \Pi(e,0) \geq (1 - T(\widehat{e},\widehat{t}))P(\widehat{e},\widehat{t},\emptyset)$ for any $(\widehat{e},\widehat{t})$. Therefore, (9) is equivalent to

$$\max_{(\widehat{e},\widehat{t}) \in \{(x,y)\in[0,1]^2 : \sigma(e,t) \geq \sigma(x,y)\}} \left[(1 - T(\widehat{e},\widehat{t}))P(\widehat{e},\widehat{t};\emptyset) + T(\widehat{e},\widehat{t})\right] = \Pi(e,t). \tag{10}$$

Given that $\Pi(e,t)$ is non-decreasing in t (condition (i)), (10) can equivalently be written as

$$\max_{r\in[\underline{e}(\sigma(e,t)),1]} \{[1 - T(r,\tau(r,\sigma(e,t)))]P(r,\tau(r,\sigma(e,t)),\emptyset) + T(r,\tau(r,\sigma(e,t)))\} = \Pi(e,t)$$

or equivalently,

$$e \in \operatorname*{arg\,max}_{r\in[\underline{e}(\sigma(e,t)),\overline{e}(\sigma(e,t))]} \Pi(r,\tau(r,\sigma(e,t))). \tag{11}$$

Thus, IC is satisfied for every type if and only if for every (e,t), (11) is satisfied. This is true if and only if $\Pi(r,\tau(r,\sigma(e,t)))$ is constant in r for $r \in [\underline{e}(\sigma(e,t)),\overline{e}(\sigma(e,t))]$ for every (e,t).

That the latter is sufficient for (11) to hold for every (e,t) is immediate. I show necessity by showing the contrapositive. Assume that for some (e,t), $\Pi(r,\tau(r,\sigma(e,t)))$ is *not* constant in r for $r \in [\underline{e}(\sigma(e,t)),1]$. That is, for some (e,t) there exist r_1, r_2 with $\underline{e}(\sigma(e,t)) \leq r_1 < r_2 \leq \overline{e}(\sigma(e,t))$ such that $\Pi(r_2,\tau(r_2,\sigma(e,t))) \neq \Pi(r_1,\tau(r_1,\sigma(e,t)))$. If $\Pi(r_2,\tau(r_2,\sigma(e,t))) < \Pi(r_1,\tau(r_1,\sigma(e,t)))$, IC of type $(r_2,\tau(r_2,\sigma(e,t)))$ is violated, as she prefers to imitate type $(r_1, \tau(r_1,\sigma(e,t)))$. If, instead, $\Pi(r_2, \tau(r_2,\sigma(e,t))) > \Pi(r_1, \tau(r_1,\sigma(e,t)))$, IC of type $(r_1, \tau(r_1,\sigma(e,t)))$ is violated, as she prefers to imitate type $(r_2, \tau(r_2,\sigma(e,t)))$.

Step 4: It is easy to see that $\Pi(r, \tau(r,\sigma(e,t)))$ being constant in r over $r \in [\underline{e}(\sigma(e,t)),\overline{e}(\sigma(e,t))]$ for every (e,t) is equivalent to condition (ii).

Step 5: Finally, notice that provided that conditions (i) and (ii) hold, conditions (iii) and (iii') are equivalent. **Q.E.D.**

Proof of Lemma 6 Take any IC mechanism $M \equiv \langle T,P\rangle$. Condition (iii) of Proposition 1 says that $\Pi(0,0) \geq (1 - T(e,t))P(e,t,\emptyset)$ for any (e,t). Then, construct the mechanism $M' := \langle T',P'\rangle$ with[53]

$$\begin{aligned} T'(e,t) &:= \Pi(e,t) - \Pi(0,0) = (1 - T(e,t))P(e,t,\emptyset) + T(e,t) - \Pi(0,0) \\ &\leq \Pi(0,0) + T(e,t) - \Pi(0,0) = T(e,t), \quad \text{and} \\ P'(e,t,\emptyset) &:= \frac{\Pi(0,0)}{1 - \Pi(e,t) + \Pi(0,0)} \geq \frac{(1 - T(e,t))P(e,t,\emptyset)}{1 - \Pi(e,t) + (1 - T(e,t))P(e,t,\emptyset)} = P(e,t,\emptyset) \end{aligned}$$

for every (e,t), where the inequalities follow from $\Pi(0,0) \geq (1 - T(e,t))P(e,t,\emptyset)$.

By construction we have that $\Pi'(e,t) = \Pi(e,t)$ for every (e,t), so M' satisfies conditions (i) and (ii) of Proposition 6. By construction, we also have that for every (e,t)

$$\Pi'(0,0) = \Pi(0,0) = (1 - T'(e,t))P'(e,t,\emptyset),$$

so M' also satisfies condition (iii) of Proposition 6. Therefore, M' is IC.

Last, to see why the second part is true, notice that for $c > 0$, M' saves on testing costs compared to M if there exists (a positive measure of) (e,t) with $P(e,t,\emptyset)(1 - T(e,t)) < \Pi(0,0)$, since $T'(e,t) < T(e,t)$ for such (e,t). **Q.E.D.**

[53]If $\Pi(0,0) = 0$, then for (e,t) such that $\Pi(e,t) = 1$, set $P'(e,t,\emptyset) = 0$.