Selecting candidates: the value of non-uniform tests and self-reports

Carl Heese *

June 29, 2022

This paper provides and studies a model for the optimal design of candidate selection processes in hiring, promotion, and task allocation decisions. A principal has the capacity to evaluate a finite number of candidates with coarse quality tests. We show that it is never optimal to use tests of the same difficulty uniformly across candidates—though using uniform tests is a common practice in many industries. Further, asking the candidates to self-report their qualities leads to first-best allocation decisions when combined appropriately with subsequent testing. This holds even when the firm is restricted to test candidates simultaneously. This result suggests that data-driven decision processes relying on tests should be viewed as a complement to traditional approaches like interviews rather than as a replacement.

1 Introduction

Being able to select good employees is fundamental to the success of any firm. Hiring, promotion, and task allocation decisions are just some examples of numerous managerial decisions where candidate selection is important. However, candidate selection is typically difficult since there is incomplete information about candidate quality. Available signals about candidate quality— educational degrees, certificates, previous work performance and evaluations—are useful parameters; however, they are oftentimes considerably limited in their informativeness. For example, educational degrees may be incomparable across countries or schools. Similarly, previous work performance evaluations may be incomparable across evaluators, branches, and offices, or may be simple unreliable. Firms can also not simply ask candidates to reveal their qualities since they may have incentives to be untruthful about their weaknesses, past problems in the work place, et cetera. For these reasons, some practitioners go so far as to describe traditional hiring tactics such as interviews and screening of resumes as "notoriously unreliable".[1] To get more reliable information on candidates, more and more companies begin to use a data-driven approach that relies on tests. In line with this, the empirical literature documents that test-based worker selection results in more productive hires (Autor and Scarborough, 2008), and that employees hired with testing have about 15% longer tenures than those hired without testing (Hoffman, Kahn and Li, 2018).[2]

*University of Vienna, Department of Economics, Oskar-Morgenstern-Platz 1, 1020 Vienna, Austria; carl.heese@univie.ac.at

[1]See, for example, this practitioner's blog by a data science startup providing assistance to companies with test-based hiring methods (https://blog.criteriacorp.com/the-limitations-of-simulation-tests/) or this practitioner's commentary in the Harvard Business Review (https://hbr.org/2020/06/how-to-design-a-better-hiring-process).

[2]A study conducted by Gallup concluded that companies that selected the top 20 percent of applicants based on talent assessments increased productivity by 10 percent and decreased turnover by 10 percent.

Theory ought to anticipate these developments in order to accompany firms in designing selection processes with tests efficiently and in order to inform regulatory decision-making. However, the previous literature in personnel economics provides few theoretical models of test-based hiring.[3] None of the existing work has analyzed how to optimally design the testing of multiple candidates and the interaction with self-reports of the candidates. In this paper, we provide a stylized mechanism design framework to characterize qualitative features of such optimal selection processes.

In the model, a firm tries to allocate a task or a job to one candidate from a finite pool of candidates. The quality of each candidate is uncertain and is drawn from a commonly known distribution. The firm may perform a pass-fail test with each of the candidates and it is assumed that candidates with a higher ability are more likely to pass any given test (*monotonicity*). We think of the test primarily as a computerized evaluation, speaking to the data-driven evaluation approaches that are by now prevalent in many industries.[4][5] Our model is general enough to accomodate testing strategies that are non-uniform in the sense that tests may vary across candidates. This way, we will be able to evaluate the efficiency of the common practice to use the same "standardized" test for all candidates. Further, we consider several model variants, including one in which tests may be chosen contingent on a message by the candidate about her quality. We may think of the message stage as a formalization of selection tools in which candidates are asked to report about themselves, like in traditional interviews. This model variant allows to study how to efficiently combine testing with asking for self-reports.

The first main insight is that optimal tests are *non-uniform*. That is, for the firm, it is optimal to confront candidates with tests of different difficult levels. This result is not driven by any heterogeneity of the candidates since our model deliberately considers ex-ante symmetric candidates. The result is robust across settings: It holds when tests can be chosen contingent on a cheap-talk message and when they cannot; it holds when the firm is constrained to perform all tests simultaneously and when it can evaluate candidates sequentially. This finding suggests that the common practice of using uniform, standardized test procedures comes at an efficiency loss.

The second main insight is that asking candidates for a self-report about their abilities is a very powerful tool of evaluation—contrary to the intuition that self-reports may not be informative since all candidates would simply pretend to be very qualified. Formally, we show that the firm can combine self-reports with quality tests in a way so that the first-best candidate is selected with probability one. In contrast, self-reports alone do not lead to any information transmission about candidate quality. The intuition for the result is that tests can be used to incentivize truth-telling. If a candidate knows that the statements and announcements he makes may be tested later on, this gives incentives not to lie. Incentivizing truth-telling is an additional argument for using non-uniform tests for each candidate. This result highlights that a data-driven hiring process relying on tests should be viewed as a complement to traditional

[3]Chapter 2 of Lazear and Gibbs (2014) provides an excellent discussion of test-based screening in hiring and its relevance. Autor and Scarborough (2008) study the effect of biased hiring tests on minority discrimination. Carroll and Egorov (2019) consider a setting complementary to ours with one candidate that has multiple skill dimensions; we discuss their work momentarily in more detail. More distantly related, there is a literature of matching and search models in which information about worker productivity arrives exogenously after the hire as the worker's tenure increases (see e.g., Jovanovic, 1979); workers are layed off if their productivity reveals to be low, similar to a probationary test period.

[4]According to a survey by Psychology Today, about 80% of Fortune 500 companies use pre-employment testing as a recruiting strategy.

[5]Alternatively, one may also think of the test in our model as a practical test or a probation period in which the firm observes the worker's performance before a final hiring decision.

interviews rather than as a replacement. We describe in more detail how tests and self-reports may be combined efficiently in Section 3.3.2.

We show that optimal mechanisms have further desirable properties. First, they have a high level of comparability. This qualifies testing over other quality signals (e.g., school diploma or previous work evaluation) that are typically available, but often lack this degree of comparability. Specifically, we show that optimal pass-fail tests are of a cutoff type: A candidate passes only if the quality level exceeds a given cutoff. Cutoff tests are ordered completely according to their difficulty through the cutoff. This way, whenever several candidates pass their test, the test results are easily comparable. Comparability does not necessitate using the same test for all candidates, but only a system of tests that has a consistent order. Second, we show that there is no trade-off between efficiency and discrimination: Optimal test mechanisms can be designed in a non-discriminatory way—although they combine different tests across candidates. Namely, they can be designed so that, for a given candidate quality, each candidate faces any given test with the same likelihood, independent of the candidate's label.[6]

The paper contributes to the literature on asymmetric information in personnel economics, in particular on hiring decisions. There is a long-standing and large theoretical literature that has studied methods how firms and society might solve the informational problem. Notably, the literature on labor market signaling pioneered by Spence (1978) analyzes how candidates may costly signal their types, for example, through engaging in educational activities. Jobs may be designed in a way that induces employees to self-select into the "right" jobs, that is, jobs *indirectly* "screen" the applicants (see e.g., Salop and Salop, 1976; Stiglitz, 1975). This paper is concerned with quality tests, that is, mechanisms that can *directly* screen quality types of candidates.

This paper contributes to the literature on mechanism design with *partial verification*, as pioneered by Green and Laffont (1986). In Green and Laffont (1986), different types of agents face different restrictions regarding the set of reports that they can send to the principal. The interpretation is that there is an *exogenous structure* of tests and certain reports are detected as false with certainty. In this paper, we analyze optimal tests, that is, the testing environment is not exogenous. We follow Caragiannis, Elkind, Szegedy and Yu (2012) who generalize the framework by Green and Laffont (1986) and allow for probabilistic verification. The firm in our model can use probabilistic tests, that is, a given type may both pass and fail the test with positive probability. However, we show that the monotonicity of quality tests implies that optimal tests are non-probabilistic. They are simple cutoff tests that check if the quality exceeds a given cutoff level. Complementary to our paper, Carroll and Egorov (2019) consider a firm that can verify one of multiple skill dimensions of a single candidate. They characterize conditions under which the firm can achieve first-best outcomes by using a random verification strategy contingent on the candidate's report. In contrast, we consider a setting with multiple candidates and in which the firm only needs to use coarse, binary tests instead of having to verify exact types. We compare the efficiency of uniform and non-uniform testing. In particular, we establish an efficiency result similar to Carroll and Egorov (2019), showing that an appropriate non-uniform testing strategy can incentivize candidates to reveal all the privately held information about their quality and lead to first-best outcomes.

In the main text, we consider a model of a selection process that makes stark assumptions. We believe, however, that the economic messages that the paper's results convey are of a more

[6]Typically, there is also no legal reason for firms to adhere to uniform testing. For example, in the US, in the guidelines of the Equal Employment Opportunity Commission (EEOC), the single most important property of a testing procedure to be legally compliant is job relatedness. Tests must measure job-related skills and abilities.

general nature. In Section 3.5, we discuss several generalizations (noisy tests, costly tests) and how the results extend to those generalized settings.

2 Model

There is a principal (he) and multiple candidates (she) $i \in \{1, \ldots, n\} = N$ with $n > 1$. The principal has one task that he can allocate to one of the candidates. A candidate receives a utility $v = 1$ when the task is allocated to her. The candidates have quality types $\theta_i \in \mathbb{R}$ that are drawn independently from a log-concave distribution F with continuous density $f > 0$.[7] The principal receives utility θ_i when allocating the ask to a candidate with quality θ. When the task is not allocated to any candidate, the principal receives a utility of zero from an outside option.

The following sections study several tools of candidate selection. Section 3.1 and 3.2 analyze quality tests. Section 3.3.1 studies self-reports. Finally, in Section 3.3.2, we analyze how tests and self-reports can be combined efficiently.

3 Mechanisms of candidate selection

A *(quality) test* for candidate i is a weakly increasing function $T_i : \mathbb{R} \to [0, 1]$ from types θ to the likelihood $T_i(\theta)$ that candidate i passes the test if her type is θ. The monotonicity of T_i captures that candidates of better quality are more likely to pass a quality test. The principal can run one test on each candidate $i \in \{1, \ldots, n\}$. We think of n as a capacity constraints on the number of tests. We assume $\mathrm{E}(\theta_i) = 0$, so that the outside option of utility zero has a simple interpretation: It can be understood as allocating the good to a random candidate $n+1$ (which cannot be tested given the capacity constraint). he set of outcomes is $N_0 = \{0, \ldots, N\}$ where the outcome 0 means that the principal chooses the outside option.

3.1 Simultaneous tests

First, we consider the scenario in which the principal is constrained to run the quality tests on all of the candidates *simultaneously*. This captures real-life scenarios in which candidates literally have to be evaluated simultaneously, e.g., due to time constraints. The simultaneity may also capture situations in which the principal delegates the tests to several subordinates or departments that are constrained in their communication.

After observing all test results, the principal has to allocate the task to one of the candidates or choose the outside option. Formally, a *test mechanism* σ is a profile of tests and a mapping $a : \{0, 1\}^N \to N_0$ from profiles of test results to outcomes.

Theorem 1 characterizes the simultaneous test mechanisms that maximize the principals payoff. It shows that the optimal simultaneous test mechanisms are characterized by two properties. First, optimal test mechanisms utilize *cutoff tests* T. These are tests that pass the candidate only if a candidate's quality exceeds a cutoff level $\tau > 0$. Formally,

$$T(\theta) = \begin{cases} 0 & \text{if} \quad \theta < \tau \\ 1 & \text{if} \quad \theta \geq \tau. \end{cases}$$

[7]Many standard distributions are log-concave. For example, the normal distribution, the exponential distribution, and the gamma distribution. See Bagnoli and Bergstrom (2005) for a larger list of commonly used distributions that are log-concave.

Second, optimal tests mechanisms are non-uniform. That is, heterogeneous test levels are applied to the n candidates.

Theorem 1. *Any optimal simultaneous test mechanism uses a cutoff test with cutoff level τ_i for each candidate $i = 1, \ldots, n$. There is a permutation p of the candidates such that the cutoff levels are strictly ordered, $\tau_{p(1)} < \tau_{p(2)} < \ldots < \tau_{p(n)}$. The mechanism allocates the task to the candidate i with the most difficult test (the one with the highest cutoff level τ_i) among those who pass their test.*

The proof of Theorem 1 is in the Appendix. Here, we sketch the argument.

Take a test profile $(T'_1, \ldots, T'_n)$ and note that for any test T'_i that is not a cutoff test, there is a cutoff test T''_i that is passed with the same probability by a random candidate. Since the distribution of θ is absolutely continuous,

$$\mathrm{E}(\theta | T''_i(\theta) = 1) > \mathrm{E}(\theta | T'_i(\theta) = 1) \text{ and } \mathrm{E}(\theta | T''_i(\theta) = 0) < \mathrm{E}(\theta | T'_i(\theta) = 0). \tag{1}$$

The payoff of the principal from allocating the task optimally given test results $(t_1, \ldots, t_n) \in \{0, 1\}^n$ is $u(t_1, \ldots, t_j) = \max\left(\max_{j \in N} \mathrm{E}(\theta_j | T_j = t_j), \mathrm{E}(\theta)\right)$. Using (1), we show in the Appendix that the distribution of u given the test profile $(T'_1, \ldots T'_i, T'_n)$ is strictly first-order stochastically dominated by the distribution of u given the test profile $(T'_1, \ldots T''_i, T'_n)$ in which the cutoff test T''_i replaces the original test for candidate i. This implies that the principal's expected payoff is higher. This shows that any optimal test mechanism only involves cutoff tests.

Algorithm. Next, we provide a recursive algorithm for the optimal cutoffs of the tests. For this, denote V_k denote the principal's expected payoff from an optimal simultaneous test mechanism when there are $k \geq 0$ candidates, where $V_0 = \mathrm{E}(\theta)$ is the value of the outside option.[8] Take any optimal profile of cutoff tests given by cutoff levels $\tau_1, \ldots, \tau_n$. In any optimal mechanism, the task is given to the highest candidate passing her test and the principal chooses the outside option if all candidates fail their test since $\mathrm{E}(\theta | \theta < \tau) < \mathrm{E}(\theta)$ for all $\tau > 0$. Further, since the candidate's quality is ex-ante symmetric, we can relabel the candidates so that $\tau_1 \leq \ldots \leq \tau_n$. With this labeling, the principal gives the task to candidate $i = 1$ only if all candidates $j > i$ fail their tests. Conditional on this event, the task is allocated to $i = 1$ only if she passes her test and otherwise the principal's expected payoff is given by V_{i-1}. Hence, conditional on the candidates $j > i$ failing their tests, the principal's expected payoff is

$$V_i = \Pr(\theta \geq \tau_i)\ \mathrm{E}(\theta_1 | \theta \geq \tau_i) + \Pr(\theta \leq \tau_i)\ V_{i-1}, \tag{2}$$

for $i \in N$. Note that V_i is defined recursively.

The first-order condition with respect to τ_i is $f(\tau_1)(\tau_i - V_{i-1}) = 0$, where $f > 0$ is the continuous density of the type distribution. Hence,

$$\tau_i = V_{i-1}. \tag{3}$$

for all $i \in N$. Combining (2) and 3, we see that V_i are strictly increasing in i. Hence, the cut-offs τ_i are strictly increasing in i and the optimal test is non-uniform.

[8] Optimal simultaneous test mechanisms exist by an application of Kakutani's fixed point theorem since they are characterized by finitely many cutoff levels $\tau_1, \ldots, \tau_k \in \mathbb{R}$ and since the expected payoff of the principal is continuous in the cutoff levels.

3.2 Sequential tests

We consider the scenario in which the principal can test the candidates sequentially. At each point of time $t = 1, \dots, n$, the principal runs a test on candidate $i = t$. The set of quality tests is denoted $\mathcal{T}$. A history h_t at time t is a collection of the tests run on the candidates $i < t$ together with their pass-fail results, and $\mathcal{H}_t$ is the set of histories h_t. We generalize the notion of a test mechanism. A *(sequential) test mechanism* σ is a collection of functions $\psi_t : \mathcal{H}_t \to \mathcal{T}$ for $t = 1, \dots, n$, mapping histories before time t to quality tests and a mapping $a : \{0,1\}^N \to N_0$ from profiles of test results to outcomes. A test mechanism is *incentive-compatible* if the principal makes an optimal choice at each history h_t, that is, if the test mechanism corresponds to a perfect Bayesian equilibrium (Fudenberg and Tirole, 1991). In the following, we use the terms incentive-compatible test mechanism and equilibrium test mechanism interchangeably.

Theorem 2 shows that the qualitative results of Theorem 1 extend to sequential test mechanisms. Equilibrium test mechanisms use cutoff tests and are non-uniform. The proof is in Appendix 5.

Theorem 2. *Take any equilibrium test mechanism. Then, for any history h_t, the firm uses a cutoff test. At any history h_t, the test applied to candidate t has a strictly higher cutoff than the test of any candidate $j < t$ that passed her test, given h_t. The task is allocated to the last candidate that passes her test. If all candidates fail their tests, the firm chooses the outside option.*

Note that Theorem 2 implies that later candidates typically face more difficult tests.

3.3 Are self-reports useful?

We analyze the usefulness of asking the candidates for self-reports. 3.3.1 considers the scenario in which the principal can ask for reports, but has no access to quality tests. Section 3.3.2 then considers the scenario in which the principal has both access to self-reports and to quality tests.

3.3.1 Self-Reports.

We consider the scenario in which the principal asks the candidates to send in a report, e.g. about their type, and then, based on the reports, chooses how to allocate the task. Formally, the principal chooses a message set M. A strategy of a candidate i is a mapping from types to distributions of messages, $\eta_i : [0,1] \to \Delta(M)$. Further, the principal chooses an allocation mechanism $\phi : M^n \to \mathbb{R}^{n+1}$ where $\phi_i(m_1, \dots, m_n)$ specifies the likelihood of outcome $i \in N_0$ given any profile of reports $(m_1, \dots, m_n)$ sent by the candidates. Any mechanism (M, ϕ) defines a Bayesian game of the candidates. Take any mechanism (M, ϕ) and take any equilibrium $\eta = (\eta_1, \dots, \eta_n)$ of the corresponding Bayesian game. Then, given η, any type θ_i of candidate i chooses a message that maximizes the likelihood of her being allocated the task. Thus, the likelihood of allocating the task to candidate i is independent of her type η_i and $\mathrm{E}(\theta_i | \phi_i(\theta_i) = 1) = \mathrm{E}(\theta_i)$. Since the argument applies to all candidates $i = 1, \dots, n$,

$$\sum_{i=1\dots,n} \Pr(\theta_i = 1)\ \mathrm{E}(\theta_i | \phi_i(\theta_i) = 1) = \mathrm{E}(\theta_1),$$

given that types are identically distributed across candidates. We conclude:

Theorem 3. *Self-reports alone do not allow the principal to improve over random candidate selection.*

3.3.2 Combining self-reports and tests.

The next result, Theorem 4, shows that combining self-reports together with quality tests allows to allocate the task *always* to the candidate with the highest quality. Moreover, to allocate the task optimally, it is sufficient if the principal has access to simultaneous tests. Comparing this result with the efficiency of self-reports and quality tests alone (Theorem 2 and Theorem 3) shows that self-reports and quality tests are *complementary* tools.

Formally, we consider the scenario in which the principal commits to a *joint allocation and test mechanism.* This is a message set M, a mapping $\psi : M^n \to \mathcal{T}^n$, specifying the profile of tests $(T_1, \ldots, T_n)$ to run on the candidates after seeing a report profile $(m_1, \ldots, m_n)$, and a mapping $\phi : M^n \times [0,1]^n \to N_0$ from report profiles and profiles of pass-fail test results to outcomes. Any joint mechanism defines a Bayesian game of the candidates. A joint mechanism is *direct* if the message set is identical to the type set, $M_i = \mathbb{R}$ for all $i \in N$, and an equilibrium σ of the Bayesian game induced by a direct mechanism is a *truth-telling equilibrium* if $\sigma(\theta_i) = \theta_i$ for all $i = 1, \ldots, n$ and $\theta_i \in \mathbb{R}$.

Theorem 4. *There is a direct joint allocation and test mechanism and a truth-telling equilibrium in which the task is assigned to the candidate with the highest quality with probability 1.*

Proof. Consider the following joint mechanism. Let $M = \mathbb{R}$, and for any report profile $m = (m_1, \ldots, m_n)$. Let $\psi(m) \in \mathcal{T}^n$ be the profile of cutoff tests evaluating if each candidate has at least the quality that she reports. That is, given m, the quality test for candidate i is

$$T_i(m)(\theta_i) = \begin{cases} 0 & \text{if} \quad \theta_i < m_i, \\ 1 & \text{if} \quad \theta_i \geq m_i. \end{cases}$$

Given the test results, the mechanism allocates the task to the candidate with the highest report among those who passed their test. The following shows that truth-telling is an equilibrium, thereby finishing the proof of Theorem 4. Suppose that all candidates $j \neq i$ report truthfully. Hence, all candidates $j \neq i$ pass their cutoff tests. If candidate i reports $m_i > \theta_i$, she fails the test and is not allocated the task for sure. This is not profitable. If candidate i reports $m_i < \theta_i$, she passes the test and is allocated the task whenever $\max_{j \neq i} \theta_j < m_i$. If she reports truthfully, $m_i = \theta_i$, she passes the test and is allocated the task whenever $\max_{j \neq i} \theta_i < \theta_i$, hence, with a weakly higher probability than when reporting $m_i < \theta_i$. We conclude that reporting $m_i < \theta_i$ is not profitable either. ■

3.4 Optimal mechanisms that are non-discriminatory

We argue that there is an optimal mechanism that does not discriminate against any candidate. By this, we mean that the mechanism is *anonymous*: permuting the labels $i = 1, \ldots, n$ of the candidates does not affect the distribution of tests that a given candidate faces or the distribution of outcomes. Note that the optimal joint test and allocation mechanism of Theorem 4 is anonymous since tests and outcomes do not depend on labels, but just reports of the candidates.

Further, consider any optimal profile of simultaneous tests as in Theorem 1. If the firm randomizes uniformly over the given profile and any permutation of it, this *random* test mechanism is anonymous. That is, it does not discriminate any candidate by its label. Similarly, uniformly randomizing the order of the candidates in any sequential mechanism yields an anonymous mechanism.

3.5 Extensions

Costly tests. In many settings, tests are costly. We do not model costs explicitly, but impose an exogenous restriction on the time or equivalently the number of candidates that can be evaluated. If there would be explicit cost $c > 0$ for running a test on a candidate, the principal would stop evaluating candidates at a history h_t as soon as the cost exceed the expected benefit of testing an additional candidate. This would endogenize the finite number of candidates that is evaluated.

Noisy tests. Consider the alternative setting in which, for each candidate, a random signal $s_i \in \mathbb{R}$ is drawn from an exogenous signal distribution that depends on the true ability of the candidate, and the principal can only run test mechanisms that evaluate the realized signal. Hence, tests are noisy in the sense that a candidate of a given quality may pass a given test, but may also fail the test, depending on the realization of the noisy signal. In this alternative setting, the type and signal distribution jointly induce a distribution of posterior quality $p_i = \mathrm{E}(\theta_i|s_i)$. Recall that the principal's objective is to maximize the average posterior quality of the candidate who is allocated the task. Therefore, for the purposes of analyzing optimal choices of the principal, this alternative setting is equivalent to the model in Section 3.2 if we let the type distribution equal the induced distribution of posteriors p_i. We conclude that it is without loss to consider a model in which tests are not "noisy", that is, in which tests can discriminate types and not just signals of types.

4 Conclusion

We have provided and analyzed a model for the optimal design of candidate selection processes. Applications include hiring, promotion, and task allocation decisions. Two main findings emerge: First, it is never optimal to use tests of the same difficulty uniformly across candidates. This observation is important because it is a common practice in many industries to rely on standardized test procedures that are uniform. Second, asking the candidates to self-report their skills and abilities leads to first-best allocation decisions when combined appropriately with subsequent testing. This result suggests that a data-driven hiring process relying on tests should be viewed as complementary to traditional interviews rather than as a replacement.

We provided further results to qualify the optimal non-uniform testing mechanism. We show that it is consistent with high levels of comparability and can be designed in a way that is non-discriminatory.

We conclude with an observation related to discrimination: Our results also show that optimal tests increase in their difficulty level over time if testing is sequential; compare to Theorem 2. Intuitively, this is because firms will try to find candidates that are better than the ones they already evaluated. If the timing of the tests correlates with non quality-related characteristics of the candidates, this may imply discrimination. While we are not aware of any instance where such correlation between candidate characteristics and timing is prevalent, this observation may inspire to re-think the classical dichotomy of the discrimination literature (for recent reviews of the discrimination literature, see Bertrand and Duflo, 2017; Charles and Guryan, 2011). Much of this literature categorizes discrimination as one of two types: Taste-based discrimination (Becker, 1957) and statistical discrimination (Arrow, Ashenfelter and Rees, 1973; Phelps, 1972). While this binary distinction is important, it does not include other types of discrimination that may be also relevant. Our observation suggests that some part of discrimination may be driven by natural properties of hiring procedures that arise from optimizing behaviour of firms and only *happen* to correlate with the candidates' characteristics. Similarly, the political science literature has started to analyze the empirical relevance of discrimination that arises out of strategic behaviour: e.g., Bateson (2020) provides evidence that individuals hesitates to support a candidate out of concern that others will object to the candidate's identity.

5 Appendix

5.1 An auxiliary result

We denote a generic profile of test results by $t = (t_1, \ldots, t_n) \in \{0,1\}^n$ and a generic profile of (quality) tests by $T = (T_1, \ldots, T_n)$. The set of all profiles of tests is $\mathcal{T}$. Any distribution $G \in \Delta(\mathcal{T})$ of test profiles induces a joint distribution $R(G)$ of profiles of tests and profiles of test results. For any G and any realization of tests and test profiles $T' = (T'_1, \ldots, T'_n)$ and $t' = (t'_1, \ldots, t'_n)$, denote the posterior profile by

$$p(T', t') = (\mathrm{E}(\theta_1 | T'_1(\theta_i) = t'_1, T_1 = T'_1), \ldots, \mathrm{E}(\theta_n | T'_n(\theta_n) = t'_n, T_n = T'_n)).$$

For any given profiles of tests and test results, the payoff of the principal when allocating the task optimally is

$$u(T', t') = \max_{j \in N} (\mathrm{E}(\theta_j | T'_j(\theta_j) = t'_j, T_j = T'_j), \mathrm{E}(\theta))$$

Fix the tests and test results for a subset $J \subset N$ of the candidates. That is, consider the event $E_0 = \{(t'_1, \ldots, t'_n), (T'_1, \ldots, T'_n) : T'_j(\theta_j) = t'_j, T_j = T'_j \text{ for all } j \in J\}$ for some given $(T'_j, t'_j)_{j \in J}$. Suppose that there are two distributions G_{-J} and G'_{-J} of test profiles and test results of the candidates $N \setminus J$ with the following properties. First, there is $i \in N \setminus J$, and tests $T'_i, T''_i \in T$ so that the realized test of i is T'_i with probability 1 given G and T''_i with probability 1 given G'. Second,

$$\begin{aligned}
\Pr(T'_i(\theta_i) = 1) &= \Pr(T''_i(\theta_i) = 1), & (4)\\
\mathrm{E}(\theta_i | T'_i(\theta_i) = 1) &> \mathrm{E}(\theta_i | T''_i(\theta_i) = 1), & (5)\\
\mathrm{E}(\theta_i | T''_i(\theta_i) = 1) &\geq \max_{j \in J} (\mathrm{E}(\theta_j | T'_j = t'_j, T_j = T'_j), \mathrm{E}(\theta)), & (6)
\end{aligned}$$

Third, the joint distribution of the tests and test results of the candidates $N \setminus J \cup \{i\}$ conditional on E_0 and conditional on any realization of i's test result is the same given G_{-J} and G'_{-J},

$$R(G | E_0 \cap \{(T, t) : (T_i, t_i) = (T'_i, t'_i))\}) = R(G'' | E_0 \cap \{(T, t) : (T_i, t_i) = (T''_i, t'_i))\}) \qquad (7)$$

for any $t'_i \in \{0, 1\}$.

Lemma 1.

$$\mathrm{E}(u(T', t') | E_0; G') > \mathrm{E}(u(T', t') | E_0; G). \qquad (8)$$

Proof. Consider any profile of tests and test results of the candidates $N \setminus J \cup \{i\}$, $(T'_k)_{k \in N \setminus J \cup \{i\}}$ and $(t'_k)_{k \in N \setminus J \cup \{i\}}$, and any test result $t'_i \in \{0, 1\}$ of candidate i. The unique profile of tests and test results consistent with E_0 and t'_i that can realize given G is

$$(\hat{T}, \hat{t}) = ((T'_1, \ldots, T'_i, \ldots, T'_n), (t'_1, \ldots, t'_n)). \qquad (9)$$

The unique profile of tests and test results consistent with E_0 and t'_i that can realize given G'

is

$$(\bar{T}, \bar{t}) = ((T_1', \ldots, T_i'', \ldots, T_n'), (t_1', \ldots, t_n')). \tag{10}$$

The profiles only differ in the test performed on i. Given (4) and (7) and since the types of the candidates are independent, the likelihood of $(\hat{T}, \hat{t})$ conditional on E_0 given G is the same as as the likelihood of $(\bar{T}, \bar{t})$ conditional on E_0 given G'. Further, (4) implies

$$\begin{aligned} u(\hat{T}, \hat{t},) &\geq u(\bar{T}, \bar{t}) \\ \max_{j \in N}(\mathrm{E}(\theta_j | \hat{T}_j(\theta_j) = \hat{t}_j, T_j = \hat{T}_j), \mathrm{E}(\theta)) &\geq \max_{j \in N}(\mathrm{E}(\theta_j | \bar{T}_j(\theta_j) = \bar{t}_j, T_j = \bar{T}_j), \mathrm{E}(\theta)) \end{aligned}$$

with a strict inequality if the profile of tests and test results of the candidates $N \setminus J \cup \{i\}$ is so that

$$\max_{j \in N}(\mathrm{E}(\theta_j | \hat{T}_j(\theta_j) = \hat{t}_j, T_j = \hat{T}_j), \mathrm{E}(\theta)) = \mathrm{E}(\theta_i | \hat{T}_i(\theta_i) = \hat{t}_i, T_i = \hat{T}_i), \text{ and} \tag{11}$$

$$\max_{j \in N}(\mathrm{E}(\theta_j | \bar{T}_j(\theta_j) = \bar{t}_j, T_j = \bar{T}_j, \mathrm{E}(\theta)) = \mathrm{E}(\theta_i | \bar{T}_i(\theta_i) = \hat{t}_i, T_i = \bar{T}_i). \tag{12}$$

To establish Lemma 1, it is sufficient to show that the likelihood of profiles of tests and test results of the candidates $N \setminus J \cup \{i\}$ so that (11) and (12) hold is not zero. For this, note that if all canidates $j \notin J \cup \{i\}$ fail, and candidate i passes, (11) holds. This is, because of (6) and because $\mathrm{E}(\theta_k | T_k' = 0) \leq \mathrm{E}(\theta)$ for all k and any test $T_k' \in \mathcal{T}$. ■

5.2 Proof of Theorem 1

Take any optimal simultaneous test $T' = (T_1', \ldots, T_n')$. It remains to show that for any $i \in N$, T_i' is a cutoff test. Suppose, on the contrary, that there is an candidate i for which T_i' is not a cutoff test. Denote by T_i'' the cutoff test that lets the candidate i pass with the same probability as T_i',

$$\Pr(T_i'(\theta) = 1) = \Pr(T_i'(\theta) = 1)$$

Note that

$$\begin{aligned} \mathrm{E}(\theta | T_i''(\theta) = 1) &> \mathrm{E}(\theta | T_i'(\theta) = 1), \text{ and} \\ \mathrm{E}(\theta | T_i''(\theta) = 1) &< \mathrm{E}(\theta | T_i'(\theta) = 1). \end{aligned}$$

Consider the test profile $T'' = (T_1', \ldots, T_i'', \ldots, T_n')$ that arises from T' by replacing T_i' with T_i''. Denote by G and G' the corresponding distributions of profiles of tests and test results that put probability one on T' and T'' respectively. Then, G and G' satisfy (4) - (7) for $E_0 = \emptyset$.[9] An application of Lemma 1 implies that the principal is *strictly* better off using T'' than T'. This contradicts with the optimality of T'. We conclude that any optimal test mechanism consists only of cutoff tests.

[9] Note that (6) holds since $J = \emptyset$ and since $\mathrm{E}(\theta | T_i(\theta) = 1) \geq \mathrm{E}(\theta)$ for any test $T_i \in T$ since all tests are monotone.

5.3 Proof of Theorem 2

First, we prove another auxiliary result.

Lemma 2. *For any optimal sequential test mechanism σ, any history h_t that occurs with positive probability given σ, and test T'_i performed on an candidate i at time t given σ and h_t,*

$$\mathrm{E}(\theta_i|T'_i(\theta_i) = 1) > \max_{j<t} (\mathrm{E}(\theta_j|T'_j(\theta_j) = 1), \mathrm{E}(\theta)) \tag{13}$$

Proof. Take an optimal test mechanism σ and any history h_t that arises with strictly positive probability. Suppose that following h_t, at time t, the principal uses a test T'_i on candidate $i = t$ and that (13) does not hold.

In the following, we construct another sequential mechanism σ' that yields a strictly higher expected payoff to the principal than σ. This sequential mechanism coincides with σ for all histories that are not consistent with h_t, but differs from σ in the continuation play after h_t.

Denote $M = \max_{j\in N} (\mathrm{E}(\theta_j|h_t), \mathrm{E}(\theta))$. Since there are only finitely many continuation histories after h_t that can occur with positive probability, we can find $\epsilon > 0$ so that

$$\bar{x} = \epsilon + \max (M, \max_{h_\infty=(h_t,h_{t>0}) \text{ with } \Pr(h_\infty)>0} \max_{j\in N\setminus\{i\}} \mathrm{E}(\theta_j|h_\infty)) < \infty \tag{14}$$

Here we use the notation h_∞ for final histories, and $h_{>t}$ for the continuation history after h_t, so that $h_\infty = (h_t, h_{>t})$.

Now, modify the continuation play after h_t as follows. Following h_t, perform the cutoff test $T_{\bar{x}}$ with cutoff $\bar{x}$ on i instead of the test T'_i. Hence, given the new strategy σ', the candidate i is less likely to pass her test following the history h_t since we increased the difficulty of i's test. Following h_t and a pass result of i, do not modify the continuation testing strategy. Following h_t and a fail result modify the continuation testing strategy so that the distribution of the sequence of tests performed on the candidates $j > t$ conditional on h_t is the same given σ and σ'. Finally, the principal allocates the task optimally given the test results of all candidates. Note that, given (14), this implies that the principal allocates the task to i whenever she passes her test.

Following the history h_t, averaging over final histories, the principal is strictly better off given σ' compared to σ. This is because the principal strictly benefits from running a more difficult test on i and allocating the task to i whenever she passes the test. In comparison, under σ, it was never optimal to allocate the task to i following h_t. We conclude that the principal is strictly better off using σ' compared to σ. This contradicts with the optimality of σ. This finishes the proof by contradiction, establishing Lemma 2. ■

Now, we provide the proof of Theorem 2. First, note that Lemma 2 implies that the condition (6) is satisfied by an optimal sequential test mechanism at any on-path history. Therefore, Lemma 1 and Lemma 2 together imply that an optimal sequential test mechanism only involves cutoff tests.

Second, recall that the distribution of the types is log-concave and absolutely continuous with a density $f > 0$. This implies that, for any $\tau, \tau' \in \mathbb{R}$,

$$\tau > \tau' \Leftrightarrow \mathrm{E}(\theta|\theta > \tau) > \mathrm{E}(\theta|\theta > \tau'); \tag{15}$$

see, for example, Burdett (1996). In the following, we combine Lemma 2 and (15) to prove the statement from Theorem 2 about the dynamics of the cutoff tests used. Take any history h_t and denote by $\tau(h_t)$ the maximal cutoff of a test that a candidate $j < t$ has passed, given

h_t (set $\tau(h_t) = -\infty$ for $t = 0$). Then, the principal evaluates the candidate t with a test that has a cutoff $\tau(t)$ so that $\mathrm{E}(\theta|\theta > \tau(t)) > \mathrm{E}(\theta|\theta > \tau(h_t)$, given Lemma 2. Then, (15) implies $\tau(t) > \tau(h_t)$.

References

Arrow, K. J., Ashenfelter, O. and Rees, A. (1973). *The theory of discrimination*. Princeton University Press.

Autor, D. H. and Scarborough, D. (2008). Does job testing harm minority workers? evidence from retail establishments. *The Quarterly Journal of Economics*, **123** (1), 219–277.

Bagnoli, M. and Bergstrom, T. (2005). Log-concave probability and its applications. *Economic theory*, **26** (2), 445–469.

Bateson, R. (2020). Strategic discrimination. *Perspectives on Politics*, **18** (4), 1068–1087.

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

Bertrand, M. and Duflo, E. (2017). Field experiments on discrimination. *Handbook of economic field experiments*, **1**, 309–393.

Burdett, K. (1996). Truncated means and variances. *Economics Letters*, **52** (3), 263–267.

Caragiannis, I., Elkind, E., Szegedy, M. and Yu, L. (2012). Mechanism design: from partial to probabilistic verification. In *Proceedings of the 13th acm conference on electronic commerce*, pp. 266–283.

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

Charles, K. K. and Guryan, J. (2011). Studying discrimination: Fundamental challenges and recent progress. *Annu. Rev. Econ.*, **3** (1), 479–511.

Fudenberg, D. and Tirole, J. (1991). *Game theory*. MIT press.

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

Hoffman, M., Kahn, L. B. and Li, D. (2018). Discretion in hiring. *The Quarterly Journal of Economics*, **133** (2), 765–800.

Jovanovic, B. (1979). Job matching and the theory of turnover. *Journal of political economy*, **87** (5, Part 1), 972–990.

Lazear, E. P. and Gibbs, M. (2014). *Personnel economics in practice*. John Wiley & Sons.

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

Salop, J. and Salop, S. (1976). Self-selection and turnover in the labor market. *The Quarterly Journal of Economics*, pp. 619–627.

Spence, M. (1978). Job market signaling. In *Uncertainty in economics*, Elsevier, pp. 281–306.

Stiglitz, J. E. (1975). The theory of" screening," education, and the distribution of income. *The American economic review*, **65** (3), 283–300.