Probabilistic Verification in Mechanism Design*

Ian Ball[†] Deniz Kattwinkel[‡]

April 28, 2022

Abstract

We introduce a model of probabilistic verification in a mechanism design setting. The principal selects a statistical test to verify the agent's claim. The agent's true type determines the probability with which he can pass each test. We characterize whether each type has an associated test that best screens out all other types, no matter the social choice rule. In that case, the testing technology can be represented in a tractable reduced form. We use this reduced form to solve for profit-maximizing mechanisms with verification. As verification improves, the solution continuously interpolates from the no-verification solution to full surplus extraction.

Keywords: probabilistic verification, testing, ordering tests, evidence.
JEL Codes: D82, D86.

---

*We thank our former advisors, Dirk Bergemann and Stephan Lauermann, for continual guidance. For useful feedback, we thank audiences at Aalto, Arizona, Bonn, Brown, Columbia, PSE, Harvard/MIT, Humboldt, Rochester, SAET, and Yale. For helpful discussions, we thank Tilman Börgers, Marina Halac, Johannes Hörner, Navin Kartik, Andreas Kleiner, Daniel Krähmer, Bart Lipman, Benny Moldovanu, Jacopo Perego, Larry Samuelson, Sebastian Schweighofer-Kodritsch, Roberto Serrano, Philipp Strack, Roland Strausz, Juuso Välimäki, and Alex Wolitzky. We are grateful for funding from the German Research Foundation (DFG) through CRC TR 224 (Project B04). An extended abstract for this paper appeared in EC '19. Finally, we are grateful for the opportunity to attend the 28th Jerusalem Summer School in Economic Theory, which planted the first seeds of this project.

[†]Department of Economics, MIT, ianball@mit.edu.

[‡]Department of Economics, UCL, denizkattwinkel@gmail.com.

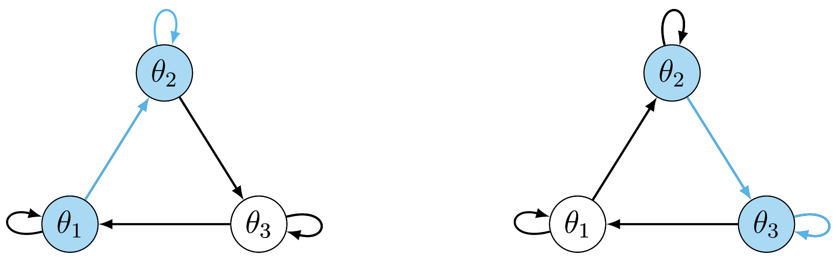

Figure 1. A challenge for the reduced-form model

1 Introduction

In the standard mechanism design paradigm, the principal cannot *verify* any agent's private information. Instead, the principal designs the game to encourage the agents to reveal their information. In practice, however, claims about private information are often verified. Sellers implementing third-degree price discrimination use online services such as SheerID to confirm buyer identities. Lenders ask borrowers for pay stubs to confirm their income reports. Insurers investigate the legitimacy of insurance claims. In all of these examples, verification is imperfect and noisy.

A parsimonious model of probabilistic verification specifies for any types θ and θ' the probability $\alpha(\theta'|\theta)$ with which type θ can "pass" as type θ' (Caragiannis et al., 2012; Ferraioli and Ventre, 2018). This reduced-form model runs into difficulties, however, as Green and Laffont (1986) observe in the special case where passage is either certain or impossible. To appreciate the problem, consider the example pictured in Figure 1. There is a single agent with three possible types, denoted $\theta_1, \theta_2, \theta_3$. The directed graph (shown twice) represents the verification technology: there is an edge from θ to θ' if type θ can "pass" as type θ'. The principal decides whether to allocate a good to the agent. Every type of agent prefers to get the good. There are no transfers.

Each copy of the graph illustrates an allocation rule. On the left, the good is allocated to types θ_1 and θ_2 (which are shaded). The principal could try to implement this rule by providing the good to the agent if he passes either as type θ_1 or as type θ_2. But this mechanism does not exclude type θ_3 since he can pass as type θ_1. Instead, the principal must provide the good if and only if the agent passes as type θ_2. Types θ_1 and θ_2 can do so, but type θ_3 cannot. On the right, the good is allocated to types θ_2 and θ_3 (which are shaded). Symmetrically, this rule must be implemented by providing the good if and only if the agent passes as type θ_3. Types θ_2 and θ_3 can do so, but type θ_1 cannot.

According to the directed graph in this example, type θ_1 can pass as type θ_2. But can type θ_1 copy type θ_2's equilibrium strategy? The answer depends on the allocation rule and the mechanism—yes, in the left mechanism (where type θ_2 passes as type θ_2); no, in the right mechanism (where type θ_2 passes as type θ_3).

The issue is that the graph—or the authentication rate $\alpha(\cdot|\cdot)$ in the general case—implicitly (a) introduces a family of statistical tests, and (b) assigns to each type θ a test, so that "passing as type θ" means passing the test assigned to type θ. In the example, type θ_2 must be given different tests in order to implement different allocation rules. Therefore, the principal's capacity to screen type θ_1 away from type θ_2 cannot be represented by a primitive probability $\alpha(\theta_2|\theta_1)$.

We model *probabilistic verification* by endowing the principal with a testing technology. There is a set of pass–fail tests. Each test has a type-dependent passage rate: type θ can pass test τ with probability $\pi(\tau|\theta)$. We consider the following protocol. The principal elicits a type report from the agent. Based on the report, the principal selects one test to conduct. The agent sees the test and privately chooses whether to try on the test or skip it. This choice is costless. If the agent tries, then his passage probability depends on the test and his type, according to π. If the agent skips the test, he fails with certainty. The principal observes whether the agent passes or fails—but not whether the agent tried—and then takes a decision.

Our paper makes two contributions. First, we formalize whether one test is "better" than another test at identifying a particular type, no matter the social choice rule. This allows us to characterize whether the testing technology can be represented by a reduced-form authentication rate. Second, we use this tractable reduced-form authentication rate to solve for optimal mechanisms with verification.

In the first part of our analysis, we introduce for each type θ an order between tests. Intuitively, test τ is *more θ-discerning* than test ψ if the *relative* performance of type θ compared to *every* other type is better on test τ than on test ψ. The formal definition requires that there is a "conversion" from τ-scores to ψ-scores that is fair for type θ but disadvantageous for all other types. This score conversion is analogous to the garbling in the definition of Blackwell's (1953) order between statistical experiments.

Our main characterization is the analogue of Blackwell's theorem in our testing setting. Consider tests τ and ψ such that τ is more θ-discerning than ψ. Theorem 1 says that any social choice rule that the principal can implement by conducting test ψ on type θ can also be implemented by conducting test τ on type θ. We apply this logic repeatedly to obtain Theorem 2: If each type θ has an associated test that is most θ-discerning, then there is no loss in assuming that the principal conducts on each type the associated test. In this case, the testing model can be represented by an authentication rate, where the probability that type θ can pass as type θ' is the probability that type θ can pass the most θ'-discerning test associated to type θ'.

Authentication rates that can be induced in this way are called *most-discerning*. We characterize these authentication rates. Our result gives a unified generalization of various conditions imposed in the literature on deterministic verification (Green and Laffont, 1986) and evidence (Lipman and Seppi, 1995; Bull and Watson, 2007).

For the second part of our analysis, we take as primitive a most-discerning authentication rate. In any mechanism, the agent must consider how perturbing his report away from the the truth reduces his probability of being authenticated. Unlike in models of deterministic verification, small deviations are detected with positive probability, so the *local* incentive constraints are relaxed. We therefore use the first-order approach. To solve for profit-maximizing mechanisms in a quasilinear setting, we compute the virtual value of each type. The virtual value of a given type consists of the value that type gets from receiving the good, net the additional information rent the allocation generates for higher types. Verification attenuates the effect on information rents by discouraging higher types from deviating downward. As verification ranges from uninformative to perfect, the virtual value increases from the classical virtual value to the true value. The solution continuously interpolates from the no-verification solution to full surplus extraction.

Verification changes the structure of the profit-maximizing mechanism for allocating a single indivisible good. With one agent, a posted price mechanism is no longer optimal. Instead, the price depends on the agent's type. With multiple symmetric agents, the optimal auction charges between the first- and second-price; the agents are still truthful. If the authentication rates are asymmetric, the optimal allocation favors a bidder if his type can be authenticated more precisely.

The rest of the paper is organized as follows. Section 2 presents our model of testing and defines implementation in our environment. Section 3 introduces the discernment orders and characterizes whether a single testing rule suffices for all implementation. Section 4 characterizes the class of authentication rates that can be induced by testing. Section 5 solves for revenue-maximizing mechanisms. Section 6 generalizes the model to nonbinary tests. Section 7 compares our model with other models of verification in economics and computer science. The conclusion is Section 8. The main proofs are in Appendix A.

2 Model

2.1 Setting

Principal–agent environment There are two players: a principal (she) and an agent (he).[1] The agent has a private type $\theta \in \Theta$, drawn from a commonly known distribution. The principal controls a decision $x \in X$.[2]

[1] We generalize to multiple agents in Section 5.2.3, where we study auctions.
[2] Transfers, if allowed, are included as a component of the decision.

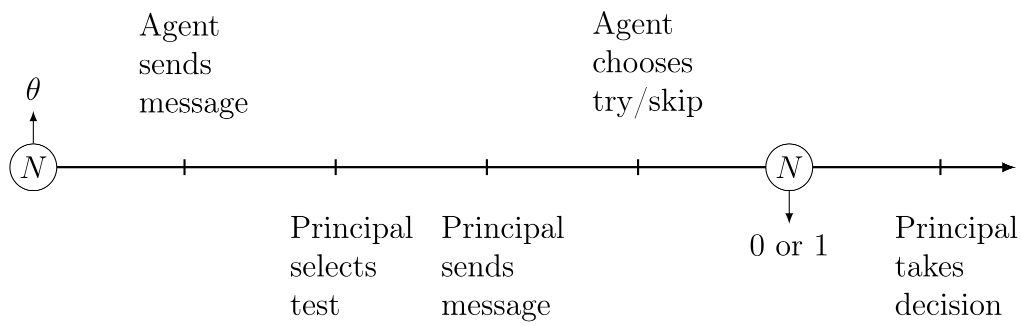

Figure 2. Timing

The agent and the principal have type-dependent utilities $u(x, \theta)$ and $v(x, \theta)$, respectively. We extend these functions linearly to $\Delta(X) \times \Theta$.

Verification The testing technology (T, π) is given by a set T of pass–fail tests and a passage rate[3]

$$\pi \colon T \times \Theta \to [0, 1],$$

which specifies the probability $\pi(\tau|\theta)$ with which type θ can pass test τ.

The principal can conduct one[4] test from the set T. The agent observes the selected test and chooses whether to *try* or *skip*. If the agent tries, his passage probability is determined by π. If he skips, he fails with certainty. The test score ("pass" or "fail") is observed by the principal, but the agent's try–skip decision is not. Because of the try–skip decision, the labels "pass" and "fail" are not interchangeable.

Mechanisms and strategies The principal can commit to an arbitrary dynamic mechanism. We consider protocols of the following form, shown in Figure 2. First the principal elicits a message from the agent. Based on the message, the principal selects a test to conduct and then sends a message to

[3]We implicitly assume that all sets are Polish and all mappings are universally measurable. The details are in Appendix C.1.

[4]Conducting multiple tests can be represented as a single compound test, but this test would likely have more scores than "pass" and "fail." Section 6 extends the model to allow for nonbinary tests.

the agent. The agent sees the realized test and the message and then privately chooses whether to try on the test or skip it. Nature draws the test score—"pass" (denoted 1) or "fail" (denoted 0). The principal observes this score, but not whether the agent tried, and then takes a decision.

Formally, a mechanism is a tuple $(M_1, M_2; t, r_2, g)$ consisting of message spaces M_1 and M_2, a testing rule $t\colon M_1 \to \Delta(T)$, a messaging rule $r_2\colon M_1 \times T \to \Delta(M_2)$, and an outcome rule $g\colon M_1 \times T \times M_2 \times \{0,1\} \to \Delta(X)$. Such a mechanism induces a dynamic decision problem for the agent. A strategy for the agent consists of a messaging strategy $r_1\colon \Theta \to \Delta(M_1)$ and an action strategy $a\colon \Theta \times M_1 \times T \times M_2 \to [0,1]$, which specifies the probability with which the agent tries on the test.

2.2 Implementation

We introduce two social choice objects. A *social choice rule* is a map from Θ to $\Delta(X)$, which specifies for each type a decision lottery. Sometimes we need to keep track of which test is conducted. An *extended social choice rule* is a map from Θ to $\Delta(T \times X)$, which specifies for each type a joint lottery over tests and decisions. A mechanism and a strategy together *implement* an (extended) social choice rule f if (i) the strategy is a best response to the mechanism and (ii) the composition of the mechanism and the strategy induce f. An (extended) social choice rule is *implementable* if there exist a mechanism and a strategy that implement it.

We show that it is without loss to focus on a special kind of implementation in which the agent reports his type truthfully and tries on whatever test he is given (and the second communication stage is omitted). Formally, a *reduced mechanism* consists of a testing rule $t\colon \Theta \to \Delta(T)$ and an outcome rule

$$g\colon \Theta \times T \times \{0,1\} \to \Delta(X).$$

In such a mechanism, a strategy for the agent consists of a reporting strategy

$r\colon \Theta \to \Delta(\Theta)$ and an action strategy

$$a\colon \Theta \times \Theta \times T \to [0,1],$$

which specifies the probability with which the agent tries as a function of his true type, his reported type, and the test. An (extended) social choice rule is *canonically implementable* if it can be implemented by a reduced mechanism (t, g) and a strategy (r, a) in which r is the identity and $a(\theta, \theta, \tau) = 1$ for all types θ and all tests τ in $\operatorname{supp} t(\theta)$.

Proposition 1 (Revelation principle)
Every implementable (extended) social choice rule is canonically implementable.

The proof has two steps. First, a standard argument (see Myerson, 1986) shows that every implementable social rule can be implemented with truthful reporting and obedient recommendations. For the second step, the idea is that the discriminatory power of testing is wasted when the agent skips a test on path. We modify the truthful and obedient implementation as follows. The principal always recommends that the agent tries. When the principal would have recommended skipping, she skips on the agent's behalf by selecting the outcome that she would have selected in the old mechanism if the agent had failed the test. The principal's recommendation is always the same, so it conveys no information and can be dropped. This new mechanism preserves truth-telling and obedience.

3 Ordering tests

In this section, we introduce a family of orders over tests. We use these orders to identify a smaller class of testing rules that suffices for all implementation.

3.1 Discernment orders

For a fixed type θ, our order captures whether one test is better than another at distinguishing type θ from all other types.

Definition 1 (θ-discernment)**.** Fix a type θ. Test τ is *more θ-discerning* than test ψ, denoted $\tau \succeq_\theta \psi$, if there exist probabilities k_1 and k_0 with $k_1 \geq k_0$ such that

(i) $\pi(\tau|\theta)k_1 + (1 - \pi(\tau|\theta))k_0 = \pi(\psi|\theta)$;

(ii) $\pi(\tau|\theta')k_1 + (1 - \pi(\tau|\theta'))k_0 \leq \pi(\psi|\theta')$ for all types θ' with $\theta' \neq \theta$.

The interpretation is that after test τ is conducted, the score $s_\tau \in \{0, 1\}$ is converted into a score $s_\psi \in \{0, 1\}$ according to the transition probabilities $\mathbf{P}(s_\psi = 1|s_\tau = 1) = k_1$ and $\mathbf{P}(s_\psi = 1|s_\tau = 0) = k_0$. The inequality $k_1 \geq k_0$ ensures that $s_\psi = 1$ is more likely if $s_\tau = 1$ than if $s_\tau = 0$. Condition i says that this score conversion is fair for type θ. If type θ tries on test τ and his score is converted, he is just as likely to pass as if he tries on test ψ directly. Condition ii says that this score conversion is weakly disadvantageous for any other type θ'. If type θ' tries on test τ and his score is converted, he is less likely to pass than if he tries on test ψ directly.

In the language of statistical hypothesis testing, we can think of failing a test as rejecting the null hypothesis. Then our definition requires that test τ can be used to construct an hypothesis test of the null θ against the alternative $\Theta \setminus \{\theta\}$ with significance $1 - \pi(\psi|\theta)$ that is uniformly more powerful than test ψ. The additional requirement that $k_1 \geq k_0$ preserves incentives, which are not an issue in the setting of hypothesis testing.

Theorem 1 (Test replacement)
Fix a type θ and tests τ and ψ such that $\tau \succeq_\theta \psi$. If a social choice rule can be canonically implemented by a mechanism in which $t(\theta) = \psi$, then it can also be canonically implemented by a mechanism in which $t(\theta) = \tau$.

The proof directly applies the definition of the relation $\tau \succeq_\theta \psi$. Start with a canonical implementation in which $t(\theta) = \psi$. Adjust the mechanism after the report θ as follows. The principal conducts test τ and then converts the agent's score s_τ into a new score s_ψ using the transition probabilities k_1 and k_0. Then the principal takes the decision that she would have taken in the old mechanism after score s_ψ on test ψ.

This new mechanism implements the same social choice rule. If type θ is truthful and then tries on test τ, he will get the decision lottery $f(\theta)$ by condition i. If any other type θ' reports type θ and then tries on test τ, then by condition ii he will get a decision lottery that he could have gotten in the original mechanism by reporting type θ and then skipping test ψ with some probability. Finally, the inequality $k_1 \geq k_0$ ensures that skipping test τ does not improve the agent's converted score.

For each θ, the definition of $\succeq_\theta$ parallels Blackwell's (1953) order. If the binary tests are re-interpreted as statistical experiments, then Blackwell's order takes the same form as Definition 1 except (a) the inequality $k_1 \geq k_0$ is dropped, and (b) the inequality in (ii) is strengthened to equality. Blackwell's order is neither stronger nor weaker than $\succeq_\theta$ since (a) weakens the definition and (b) strengthens it. By (b), no type is privileged in Blackwell's order. In a statistical experiment, the labels of the realizations are interchangeable. On a test, the scores are not interchangeable because of the agent's try–skip decision. In particular, every type can fail every test. Our order extends to tests that generate scores in an arbitrary partially ordered set; see Section 6.

Like Blackwell's order, each θ-discernment order $\succeq_\theta$ is reflexive and transitive but not generally anti-symmetric. Tests τ_1 and τ_2 are θ-*equivalent*, denoted $\tau_1 \sim_\theta \tau_2$, if $\tau_1 \succeq_\theta \tau_2$ and $\tau_1 \preceq_\theta \tau_2$. Our next result characterizes θ-equivalence. Type θ is *minimal* on test τ if $\pi(\tau|\theta) \leq \pi(\tau|\theta')$ for types all θ'. If type θ is minimal on a test τ, then test τ has no power to screen other types away from type θ and hence $\tau \preceq_\theta \psi$ for every test ψ.

Proposition 2 (θ-discernment equivalence)
Fix a type θ. Tests τ_1 and τ_2 are θ-equivalent if and only if (a) $\pi(\tau_1|\cdot) = \pi(\tau_2|\cdot)$ or (b) type θ is minimal on τ_1 and τ_2.

Corollary 1 (Discernment equivalence)
Tests τ_1 and τ_2 are θ-equivalent for every type θ if and only if (a) $\pi(\tau_1|\cdot) = \pi(\tau_2|\cdot)$ or (b) τ_1 and τ_2 have type-independent passage rates.[5]

[5] That is, there are probabilities α_1 and α_2 such that $\pi(\tau_1|\theta') = \alpha_1$ and $\pi(\tau_2|\theta') = \alpha_2$ for all types θ'.

3.2 Implementation with most-discerning testing

Theorem 1 is particularly useful if there is a single test that can replace every other test for type θ.

Definition 2 (Most discerning)**.** A test τ is *most θ-discerning* if $\tau \succeq_\theta \psi$ for all ψ in T. A function $t\colon \Theta \to T$ is *most discerning* if for each type θ the test $t(\theta)$ is most θ-discerning.

In the language of hypothesis testing, a most θ-discerning test corresponds to a uniformly most powerful test for the null θ against the alternative $\Theta \setminus \{\theta\}$.

Whether a test is most θ-discerning depends on the feasible set T. The only test that is more θ-discerning than *every* test is the perfect test $\hat{\tau}_\theta$ that exactly identifies whether the agent's type is θ, i.e., $\pi(\hat{\tau}_\theta|\theta') = [\theta' = \theta]$.

To state the main result, we define a *decision environment* to consist of a decision set X and a utility function $u\colon X \times \Theta \to \mathbf{R}$ for the agent.

Theorem 2 (Most-discerning implementation)
Fix a type space Θ and a testing technology (T, π). For a testing function $\hat{t}\colon \Theta \to T$, the following are equivalent.

1. *$\hat{t}$ is most discerning.*
2. *In every decision environment (X, u), every implementable social choice rule is canonically implementable with $\hat{t}$.*[6]

The main result is the forward implication from condition 1 to condition 2. If a testing function is most-discerning, then it suffices for all implementation. The proof applies the procedure from Theorem 1 to replace any test ψ conducted on type θ with test $\hat{t}(\theta)$.

The backward implication confirms that the most-discerning property is the right one. With any weaker definition, the main result would fail. Intuitively, any violation of $\tau \succeq_\theta \psi$ means that replacing ψ with τ must introduce a new

[6] A social choice rule f is *canonically implementable with* $\hat{t}$ if there exists a decision rule g such that $(\hat{t}, g)$ canonically implements f.

deviation. The proof constructs a decision environment in which this new deviation is profitable.[7]

Even if the testing technology does not admit a most-discerning testing function, we can still use the replacement theorem (Theorem 1) to reduce the class of tests that need to be considered. Suppose there is a subset $\hat{T}(\theta)$ of tests with the property that for every test ψ there is some test τ in $\hat{T}(\theta)$ with $\tau \succeq_\theta \psi$. Then there is no loss in assuming that on type θ the principal uses only (mixtures over) tests in $T(\theta)$. See Appendix B.1 for a formal statement, which requires some technical conditions.

3.3 Characterizing the discernment orders

We now characterize θ-discernment in terms of relative performance.

Proposition 3 (Relative performance)
Fix a type θ and tests τ and ψ.[8]

1. *Suppose $\pi(\tau|\theta) \geq \pi(\psi|\theta) > 0$. We have $\tau \succeq_\theta \psi$ if and only if there exists $\lambda \in [0,1]$ such that for all types $\theta' \neq \theta$,*

$$\lambda \frac{\pi(\tau|\theta')}{\pi(\tau|\theta)} + (1-\lambda) \leq \frac{\pi(\psi|\theta')}{\pi(\psi|\theta)}. \tag{1}$$

2. *Suppose $\pi(\tau|\theta) \leq \pi(\psi|\theta) < 1$. We have $\tau \succeq_\theta \psi$ if and only if there exists λ in $[0,1]$ such that for all types $\theta' \neq \theta$,*

$$\lambda \frac{1-\pi(\tau|\theta')}{1-\pi(\tau|\theta)} + (1-\lambda) \geq \frac{1-\pi(\psi|\theta')}{1-\pi(\psi|\theta)}. \tag{2}$$

To build intuition, first take $\lambda = 1$ in (1) and (2) to get a sufficient condition for $\tau \succeq_\theta \psi$ in each case. In the first case, where type θ is more likely to pass τ than ψ, we have $\tau \succeq_\theta \psi$ if the relative *passage* rate of type θ' compared to type θ is *lower* on test τ than on test ψ. In the second case, where type θ is

[7] The formal argument is similar in spirit to de Oliveira's (2018) proof of Blackwell's theorem using diagrams.

[8] Two edge cases are excluded. If $\pi(\psi|\theta) = 0$, then $\tau \succeq_\theta \psi$. If $\pi(\tau|\theta) < \pi(\psi|\theta) = 1$, then $\tau \not\succeq_\theta \psi$.

more likely to fail τ than ψ, we have $\tau \succeq_\theta \psi$ if the relative *failure* rate of type θ' compared to type θ is *higher* on test τ than on test ψ.

In Proposition 3, the parameter λ controls the size of the difference between the transition probabilities k_1 and k_0 in Definition 1. In particular, $\lambda = 1$ makes $k_1 - k_0$ as large as possible, and $\lambda = 0$ corresponds to $k_1 = k_0$. Consider a particular deviating type θ'. If $\pi(\tau|\theta) > \pi(\psi|\theta')$, then (1) and (2) are loosest with $\lambda = 1$. Since type θ' performs worse than type θ on test τ, type θ' is most disadvantaged if the converted ψ-score is most sensitive to the true τ-score. If $\pi(\tau|\theta) < \pi(\tau|\theta')$, then (1) and (2) are loosest with $\lambda = 0$. Since type θ' performs better than type θ on test τ, type θ' is most disadvantaged if the converted ψ-score is insensitive to the true τ-score. In Proposition 3, a single conversion must deter deviations by all types, so a single value of λ must work for every deviator θ'. Appendix B.2 gives an example where the required λ is strictly between 0 and 1.

4 Testing in reduced form

Suppose the testing technology (T, π) admits a most-discerning testing function $\hat{t}$. By Theorem 2, there is no loss of generality in restricting the principal to use $\hat{t}$. With testing function $\hat{t}$, the principal must specify for each report θ two decisions—decision $g_0(\theta)$ if the agent fails test $\hat{t}(\theta)$ and decision $g_1(\theta)$ if the agent passes test $\hat{t}(\theta)$.

The *authentication rate induced by* $\hat{t}$ is defined by

$$\alpha(\theta'|\theta) = \pi(\hat{t}(\theta')|\theta), \qquad \theta, \theta' \in \Theta. \tag{3}$$

For any types θ and θ', set

$$u(\theta'|\theta) = \alpha(\theta'|\theta)u(g_1(\theta'), \theta) + (1 - \alpha(\theta'|\theta))u(g_0(\theta'), \theta).$$

The principal's problem becomes

$$\begin{aligned} &\text{maximize} \quad \mathbf{E}[\alpha(\theta|\theta)v(g_1(\theta),\theta) + (1-\alpha(\theta|\theta))v(g_0(\theta),\theta)] \\ &\text{subject to} \quad u(\theta|\theta) \geq u(\theta'|\theta) \vee u(g_0(\theta'),\theta), \qquad \theta,\theta' \in \Theta. \end{aligned} \tag{4}$$

The incentive constraint requires that for each type θ, reporting θ and trying on test $\hat{t}(\theta)$ is weakly preferable to reporting θ' and either trying or skipping test $\hat{t}(\theta')$. In particular, taking $\theta' = \theta$ ensures that type θ prefers to try on test $\hat{t}(\theta)$ than to skip it. The cost of misreporting is determined endogenously by the outcome rule, in contrast to models of exogenous lying costs.[9]

We consider the program (4) only for authentication rates α that are induced by some testing technology with a most-discerning testing rule $\hat{t}$. Call such authentication rates *most-discerning*. To be sure, any authentication rate α implicitly defines a testing technology (T, π) given by

$$T = \{\tau_{\theta'} : \theta' \in \Theta\}, \qquad \pi(\tau_{\theta'}|\theta) = \alpha(\theta'|\theta). \tag{5}$$

If α is not most-discerning, then Theorem 2 guarantees that for some decision environment and principal objective, the solution of (4) is strictly suboptimal within the class of all mechanisms using the testing technology (T, π) in (5). If α is most-discerning, then it is easy to check that in the induced testing technology (5), the map $\theta \mapsto \tau_\theta$ is most-discerning.[10] Therefore, the solution of (4) is optimal among all dynamic mechanisms using the testing technology (T, π) in (5).

Applying Proposition 3, we immediately get the following characterization of most-discerning authentication rates.

Proposition 4 (Authentication rate characterization)

[9] In models of lying costs, the agent pays a cost $c(\theta'|\theta)$ from reporting θ' when his true type is θ. See, for example, Lacker and Weinberg (1989), Maggi and Rodriguéz-Clare (1995), Crocker and Morgan (1998), Kartik et al. (2007), Kartik (2009), and Deneckere and Severinov (2017). In particular, Kephart and Conitzer (2016) show that, with lying costs, the revelation principle holds if the lying cost function satisfies the triangle inequality.

[10] By definition, α is most-discerning if α is induced by some most-discerning testing rule, but it follows that α is induced by the rule $\theta\tau_\theta$ with the testing technology (T, π) in (5). The test set in T is as small as possible, so the most-discerning condition is as weak as possible.

An authentication rate α is most discerning *if and only if the following hold for all distinct types θ_2 and θ_3.*

1. *If $\alpha(\theta_2|\theta_2) \geq \alpha(\theta_3|\theta_2)$ and $\alpha(\theta_2|\theta_2) > 0$, then there exists $\lambda \in [0,1]$ such that, for all types $\theta_1 \neq \theta_2$,*

$$\lambda \frac{\alpha(\theta_2|\theta_1)}{\alpha(\theta_2|\theta_2)} + (1-\lambda) \leq \frac{\alpha(\theta_3|\theta_1)}{\alpha(\theta_3|\theta_2)}.$$

2. *If $\alpha(\theta_2|\theta_2) \leq \alpha(\theta_3|\theta_2)$ and $\alpha(\theta_2|\theta_2) < 1$, then there exists $\lambda \in [0,1]$ such that, for all types $\theta_1 \neq \theta_2$,*

$$\lambda \frac{\bar{\alpha}(\theta_2|\theta_1)}{\bar{\alpha}(\theta_2|\theta_2)} + (1-\lambda) \geq \frac{\bar{\alpha}(\theta_3|\theta_1)}{\bar{\alpha}(\theta_3|\theta_2)}.$$

If $\alpha(\theta|\theta) \geq \alpha(\theta|\theta')$ for all types θ and θ', then α is most discerning if and only if

$$\alpha(\theta_3|\theta_2)\alpha(\theta_2|\theta_1) \leq \alpha(\theta_3|\theta_1)\alpha(\theta_2|\theta_2), \tag{6}$$

for all $\theta_1, \theta_2, \theta_3 \in \Theta$.

If α is $\{0,1\}$-valued, then we can define for each type θ the set $M(\theta) = \{\theta' : \alpha(\theta'|\theta) = 1\}$. Each type θ can pass as any type in $M(\theta)$. If $\alpha(\theta|\theta) = 1$, then $M(\theta)$ contains θ, as assumed in Green and Laffont (1986). In terms of M, (6) becomes

$$\theta_3 \in M(\theta_2) \quad \& \quad \theta_2 \in M(\theta_1) \implies \theta_3 \in M(\theta_1),$$

which is exactly Green and Laffont's (1986) nested range condition.

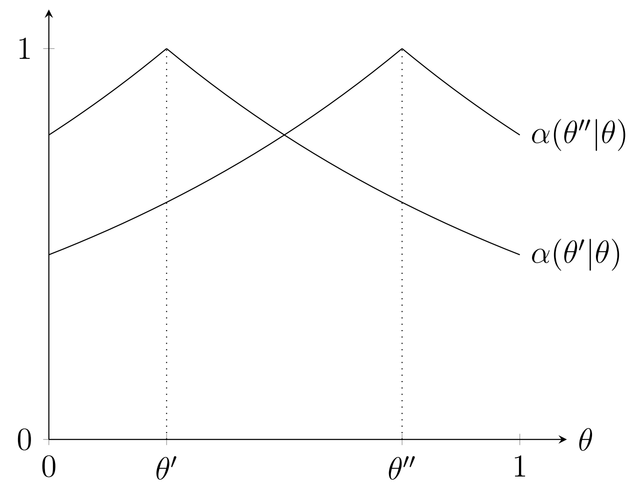

Figure 3. Exponential authentication rate

5 Applications: Profit-maximization with verification

5.1 Setting

5.1.1 Authentication rate

We represent the verification technology with a most-discerning authentication rate α. Assume the authentication rate α is given by

$$\alpha(\theta'|\theta) = \exp\left(-\left|\int_{\theta'}^{\theta} \lambda(\xi)\,\mathrm{d}\xi\right|\right), \qquad \theta, \theta' \in \Theta,$$

for some integrable function $\lambda \colon [\underline{\theta}, \bar{\theta}] \to \mathbf{R}_+$. The value of $\lambda(\theta)$ quantifies the local precision of the verification technology near type θ. The function $\alpha(\theta|\cdot)$ has a kink at type θ if and only if $\lambda(\theta) > 0$. Figure 3 plots the authentication rate when $\lambda(\theta) = 1$ for all θ. It shows the authentication probability, as a function of the agent's true type, for two reports θ' and θ''.

Every authentication rate in this exponential family is most discerning. We work with exponential authentication rates in order to get a cleaner characterization of the optimal mechanisms. Appendix B.3 shows how to extend the

the results to most-discerning authentication rates that are not exponential. In the general case, more care is needed to ensure that the global incentive constraints are satisfied.

5.1.2 Quasilinear environment

The agent's type $\theta \in \Theta = [\underline{\theta}, \bar{\theta}]$ is drawn from a distribution function F with strictly positive density f. The principal allocates a quantity $q \in Q \subset \mathbf{R}_+$ and receives a transfer $t \in \mathbf{R}$.[11] Utilities for the agent and the principal are

$$u(q, t, \theta) = \theta q - t \quad \text{and} \quad v(q, t) = t - c(q),$$

for some weakly concave cost function $c\colon Q \to \mathbf{R}_+$.

The agent is free to walk away at any time, so we impose an ex post participation constraint.[12] If the principal could impose arbitrarily severe punishments for failed authentication, then probabilistic verification would essentially reduce to perfect verification; see Caragiannis et al. (2012).

Since $\alpha(\theta|\theta) = 1$ for all θ, the agent is always authenticated on path. Thus, there is no loss in holding the agent to his outside option utility if he is not authenticated. We set $g_0(\theta) = (0, 0)$ for all θ, and we optimize over the decision rule g_1. Without loss of optimality, we restrict g_1 to be deterministic. Denote the quantity and transfer components of g_1 by q and t.

The principal selects a quantity function $q\colon \Theta \to \mathbf{R}_+$ and a transfer function $t\colon \Theta \to \mathbf{R}_+$ to solve

$$\begin{aligned} &\text{maximize} && \int_{\underline{\theta}}^{\bar{\theta}} [t(\theta) - c(q(\theta))] f(\theta)\, \mathrm{d}\theta \\ &\text{subject to} && \theta q(\theta) - t(\theta) \geq \alpha(\theta'|\theta)[\theta q(\theta') - t(\theta')], \quad \theta, \theta' \in \Theta \\ &&& \theta q(\theta) - t(\theta) \geq 0, \quad \theta \in \Theta. \end{aligned} \tag{7}$$

[11]The pair (q, t) is the decision x in the general model. In applications, t always denotes transfers (and we make no direct reference to tests).

[12]In particular, this rules out upfront payments like those used in Border and Sobel (1987).

The first constraint is incentive compatibility.[13] The second is ex post participation, conditional upon being authenticated.

5.1.3 Virtual value

We sketch the derivation of the virtual value in this quasilinear setting with verification. In the classical setting without verification, the envelope theorem pins down the derivative of the indirect utility function U in terms of the allocation rule: $U'(\theta) = q(\theta)$, so

$$U(\theta) = U(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} q(\xi)\,\mathrm{d}\xi. \tag{8}$$

With verification, the derivative of the indirect utility function U is no longer pinned down by the quantity function because of the kink in $\alpha(\theta|\cdot)$.[14] Instead, the envelope formula gives the inequality $U'(\theta) \geq q(\theta) - \lambda(\theta)U(\theta)$, so

$$U(\theta) \geq U(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} \alpha(\xi|\theta)q(\xi)\,\mathrm{d}\xi. \tag{9}$$

The exponential authentication rate α appears in the integrand as part of the solution of a linear differential equation. This is a convenient mathematical property of the exponential—it does not reflect global incentive constraints. Indeed, the envelope formula for U depends only on the *local* behavior of α around the diagonal, which is captured by the function λ.

In both (8) and (9), it is optimal to choose the smallest solution U consistent with the participation constraint. After substituting U into the objective and changing the order of integration, the principal's objective can be expressed as a linear functional in q. The coefficient on $q(\theta)$ is the virtual value of type θ. From (8) and (9), respectively, we obtain the Myersonian

[13] This is the constraint from (4), but it now takes a simple form because $u(g_0(\theta'), \theta) = 0$ for all θ' and θ.

[14] See Carbajal and Ely (2013) for a general characterization of the incentive compatible indirect utility functions with kinked utility. Carbajal and Ely (2016) apply their characterization in a model of reference-dependent utility.

(no-verification) virtual value $\varphi^M(\theta)$ and the (verification) virtual value $\varphi(\theta)$:

$$\varphi^M(\theta) = \theta - \frac{1}{f(\theta)} \int_\theta^{\bar{\theta}} f(\xi)\, \mathrm{d}\xi,$$
$$\varphi(\theta) = \theta - \frac{1}{f(\theta)} \int_\theta^{\bar{\theta}} \alpha(\theta|\xi) f(\xi)\, \mathrm{d}\xi.$$

In both cases, the virtual value of type θ is the marginal revenue from allocating to type θ. There are two parts. First, the principal can extract the additional consumption utility θ from type θ. Second, the allocation pushes up the indirect utility of each type ξ above type θ. This marginal effect on type ξ, which equals 1 in (8) and $\alpha(\theta|\xi)$ in (9), is then integrated against the relative density $f(\xi)/f(\theta)$.

Comparing the virtual values, we immediately see that

$$\varphi^M(\theta) \leq \varphi(\theta) \leq \theta.$$

The virtual value $\varphi(\theta)$ tends towards these bound in the limiting cases.

Proposition 5 (Testing precision)
As λ converges to 0 pointwise, $\varphi(\theta)$ converges to $\varphi^M(\theta)$ for each type θ.
As λ converges to ∞ pointwise, $\varphi(\theta)$ converges to θ for each type θ.

5.2 Optimal mechanisms

5.2.1 Nonlinear pricing

For nonlinear pricing (Mussa and Rosen, 1978), we have $Q = \mathbf{R}_+$. Assume that the principal's cost function c satisfies $c'(0) = 0$ and c' is strictly increasing with $\lim_{q\to\infty} c'(q) > \bar{\theta}$.

Proposition 6 (Optimal nonlinear pricing)
Suppose that the virtual value φ is increasing. The optimal quantity function

q^* and transfer function t^* are given by[15]

$$c'(q^*(\theta)) = \varphi(\theta)_+, \qquad t^*(\theta) = \theta q^*(\theta) - \int_{\underline{\theta}}^{\theta} \alpha(\xi|\theta) q^*(\xi)\,\mathrm{d}\xi.$$

The optimal allocation rule has the same form as in the classical case, except the new virtual value appears in place of the classical virtual value. Transfers are determined by the indirect utility function U, which is given by the minimal solution of (9).

Each type θ receives the quantity that is efficient for type $\varphi(\theta)_+$. Therefore, quantity is distorted downward, below the efficient level, for every type except the higherst type. As λ increases pointwise, so that the verification technology becomes more precise, downward distortion is attenuated. In the limit of perfect verification, the good is allocated efficiently and the principal extracts the full surplus.

5.2.2 Selling a single indivisible good

For a single indivisible good (Riley and Zeckhauser, 1983), take $Q = [0, 1]$. Here, quantity is interpreted as the probability of receiving the good. Hence, $c(q) = cq$, where c is the cost of producing a single good. Assume $0 \le c \le \bar{\theta}$.

Without verification, the profit-maximizing mechanism is a posted price. With verification, the optimal price depends on the agent's report.

Proposition 7 (Optimal sale of a single good)
Suppose that the virtual value φ is strictly increasing. The optimal quantity function q^* and transfer function t^* are given by

$$q^*(\theta) = [\theta \ge \theta^*], \qquad t^*(\theta) = \theta q^*(\theta) - \int_{\underline{\theta}}^{\theta} \alpha(\xi|\theta) q^*(\xi)\,\mathrm{d}\xi,$$

where $\theta^* = \inf\{\theta : \varphi(\theta) \ge c\}$.

[15] We refer to *the* optimal mechanism here and in Proposition 7, even though there is some flexibility for a knife-edge set of types.

As in the classical solution, there is a cutoff type θ^* who receives the good and pays his valuation. Each type below the cutoff is excluded; each type above the cutoff receives the good and pays less than his valuation. The allocation probability takes values 0 and 1 only—there is no randomization.

Verification increases the virtual value relative to the classical virtual value, so the cutoff type is lower and more types receive the good. The price is no longer uniform. As long as λ is strictly positive, the price is strictly increasing in the agent's report. Nevertheless, types above the cutoff cannot profit by misreporting downward—the benefit of a lower price is outweighed by the risk failing authentication and getting nothing.

5.2.3 Auctions

Consider an auction for a single indivisible good (Myerson, 1981). We extend the quasilinear setting to multiple agents. Each agent i has an exponential authentication rate α_i, with integrable precision function λ_i. The solution concept is Bayes–Nash equilibrium. We maintain the ex post participation constraint for each player. See Appendix B.4 for the formal extension of our testing model to multiple agents.

Each agent i independently draws his type $\theta_i \in \Theta_i = [\underline{\theta}_i, \bar{\theta}_i]$ from a distribution function F_i with positive density f_i. The principal allocates the good to each agent i with probability $q_i \in [0, 1]$, where $q_1 + \cdots + q_n \leq 1$. The principal receives transfers $t_1, \ldots, t_n \in \mathbf{R}$. Write $q = (q_1, \ldots, q_n)$ and $t = (t_1, \ldots, t_n)$. The principal's cost of providing the good is c, where $0 \leq c \leq \max_i \bar{\theta}_i$. Utilities are given by

$$u_i(q, t, \theta) = \theta_i q_i - t_i \quad \text{and} \quad v(q, t) = \sum_{i=1}^{n} (t_i - c q_i).$$

Proposition 8 (Optimal auction)
Suppose that each virtual value φ_i is strictly increasing. An optimal mechanism is given as follows. For each bidder i and each θ_{-i} in Θ_{-i}, let

$$r_i(\theta_{-i}) = \varphi_i^{-1}\big(\max\nolimits_{j \neq i} \varphi_j(\theta_j)_+\big).$$

If $\theta_i \leq r_i(\theta_{-i})$*, then* $q_i^*(\theta) = t_i^*(\theta) = 0$*. If* $\theta_i > r_i(\theta_{-i})$*, then* $q_i^*(\theta) = 1$ *and*

$$t_i^*(\theta) = r_i(\theta_{-i}) + \int_{r_i(\theta_{-i})}^{\theta_i} (1 - \alpha_i(\xi_i|\theta_i))\, \mathrm{d}\xi_i.$$

Only the interim allocation and transfers are pinned down. In Proposition 8, we have selected the version of the optimal mechanism that is ex-post incentive compatible.

First consider the symmetric case in which $\alpha_i = \alpha$ and $F_i = F$ for all bidders i. Here, $r_i(\theta_{-i}) = r^* \vee \max_{j \neq i} \theta_j$, where $r^* = \varphi^{-1}(0)$. If at least one bidder's valuation is above the reserve r^*, then the bidder i^* with the highest valuation gets the good and pays an amount between $r^* \vee \max_{j \neq i^*} \theta_j$ and his own valuation θ_{i^*}, with the payment increasing as the authentication rate α becomes more precise.

Next, consider ex ante asymmetric bidders. In the no-verification case, the allocation rule rewards bidders with value distributions that are weaker in the sense of the hazard rate order. With verification, the allocation rule further rewards bidders who can be verified more precisely. Allocating the good to a given type θ_i has a smaller effect on expected information rents if the bidder is weaker (so that higher types are relatively are) or more precisely verifiable (so that higher types find downward deviations less attractive).

6 Beyond pass–fail tests

The main model considers pass–fail tests. We now consider tests that generate scores in a finite score set S. The type-dependent performance on each test is given by

$$\pi \colon T \times \Theta \to \Delta(S),$$

which specifies for each type θ and test τ a distribution $\pi_{\tau|\theta}$ over S.

In order to generalize the try–skip decision, we take as primitive a partial order $\succeq$ on S. The interpretation is that the agent can reduce his score by changing s to s' if $s \succeq s'$. As before, the agent's choice is costless and unobservable to the principal.

The binary setting of the model corresponds to $S = \{0, 1\}$ with the usual order $\succeq$. If type θ mixes between trying and skipping on test τ, he can pass with any probability below $\pi(\tau|\theta)$. In the general setting, type θ can achieve on test τ any score distribution p in $\Delta(S)$ satisfying $\pi_{\tau|\theta} \succeq_{\text{st}} p$, where $\succeq_{\text{st}}$ is the stochastic order between measures on a partially ordered space (Kamae et al., 1977). That is, $\mu \succeq_{\text{st}} \nu$ if and only if $\mu(U) \geq \nu(U)$ for every upper set U.[16]

Now we define the order $\succeq_\theta$ in this setting. A function $k\colon S \to \Delta(S)$ is increasing if $k(s) \succeq_{\text{st}} k(s')$ whenever $s \succeq s'$. We interpret k as a Markov transition, and we use the following notation from Markov chains. Given μ in $\Delta(S)$ and $k\colon S \to \Delta(S)$, we write μk for the measure on $\Delta(S)$ defined by $(\mu k)(A) = \sum_s \mu(s) k(A|s)$, for $A \subset S$.

Definition 3 (θ-discernment for general tests)**.** Fix a type θ. Test τ is *more θ-discerning* than test ψ, denoted $\tau \succeq_\theta \psi$, if there exists an increasing function $k\colon S \to \Delta(S)$ such that

(i) $\pi_{\tau|\theta} k = \pi_{\psi|\theta}$;

(ii) $\pi_{\tau|\theta} k \preceq_{\text{st}} \pi_{\psi|\theta}$ for all types θ' with $\theta' \neq \theta$.

The partial order $\succeq$ determines how the agent can control his score through an unobservable action.[17] If $s \succeq s'$ holds only if $s = s'$, then the agent has no influence over his score. In this case, each order $\succeq_\theta$ reduces to Blackwell's order.[18]

In this more general setting, the replacement theorem (Theorem 1) and the forward implication in Theorem 2 go through with essentially the same proofs. Of course, this testing technology cannot be represented by an authentication rate. See Section 6 for details.

[16] An *upper set* is a set with the property that if s is in U and $s' \succeq s$, then s' is also in U. The order $\succeq_{\text{st}}$ can be defined dually in terms of lower sets (since a set is an upper set if and only if its complement is a lower set).

[17] The substantive requirements are that $\succeq$ is reflexive and transitive. If $\succeq$ is not anti-symmetric, pass to the quotient $S/\sim$. This pools all scores that the agent can freely interchange.

[18] In this case, every Markov transition k is vacuously increasing, and every subset U is vacuously an upper set, hence $\succeq_{\text{st}}$ reduces to equality.

	Reduced form	**Microfoundation**
partial	message correspondence $M\colon \Theta \twoheadrightarrow \Theta$ Green and Laffont (1986)	evidence correspondence $E\colon \Theta \twoheadrightarrow \mathcal{E}$ Bull and Watson (2007) Lipman and Seppi (1995)
probabilistic	authentication rate $\alpha\colon \Theta \times \Theta \to [0,1]$ Our paper Caragiannis et al. (2012)	passage rate $\pi\colon T \times \Theta \to [0,1]$ Our paper

Table 1. Taxonomy of verification models

7 Related literature on verification

Verification has been modeled in many ways, in both economics and computer science. Here, we focus on costless, imperfect verification.[19] Our discussion is organized around the taxonomy in Table 1.

Green and Laffont (1986) introduce *partial verification.*[20] They restrict their analysis to direct mechanisms. Verification is represented as a correspondence $M\colon \Theta \twoheadrightarrow \Theta$ satisfying $\theta \in M(\theta)$ for all types θ. Each type θ can "report" any type θ' in $M(\theta)$. They show that truthful implementation is without loss if and only if the correspondence M satisfies a transitivity condition that they call nested range condition.[21] Green and Laffont (1986) interpret their result as a revelation principle. In our framework, we would re-interpret their "failure of the revelation principle" as a consequence of implicitly identify

[19]In economics, "verification" traditionally means that the principal can learn the agent's type perfectly by taking some action, e.g., paying a fee or allocating a good. This literature began with Townsend (1979) who studied *costly verification* in debt contracts. Ben-Porath et al. (2019b) connect costly verification and evidence. When monetary transfers are infeasible, costly verification is often used as a substitute; see Ben-Porath et al. (2014); Erlanson and Kleiner (2015); Halac and Yared (2017); Li (2020); Mylovanov and Zapechelnyuk (2017).

[20]A precursor to their work is Postlewaite (1979), which considers exchange mechanisms when endowments are hidden. Each agent can benefit by hiding part of his endowment.

[21]More precisely, it is transitivity of the relation $\succeq$ on Θ defined by $\theta \succeq \theta'$ if and only if $\theta' \in M(\theta)$.

some type θ with a test that is not most θ-discerning. [22]

Bull and Watson (2004, 2007) and Lipman and Seppi (1995) study *hard evidence*.[23] They introduce an evidence set $\mathcal{E}$ and an evidence correspondence $E : \Theta \twoheadrightarrow \mathcal{E}$. Type θ possesses the evidence in $E(\theta)$; he can present one piece of evidence from $E(\theta)$ to the principal. The evidence environment is *normal* if each type θ has a piece of evidence $e(\theta)$ in $E(\theta)$ that is maximal in the following sense: Every other type θ' who has $e(\theta)$ also has every other piece of evidence in $E(\theta)$. Therefore, type θ' can mimic type θ if and only if $E(\theta')$ contains $e(\theta)$. A normal evidence environment induces an abstract mimicking correspondence that satisfies the nested range condition. Normality is a special case of our most-discerning condition.

In computer science, Caragiannis et al. (2012) and Ferraioli and Ventre (2018) study a reduced-form model of *probabilistic verification* in mechanism design. Their analysis is implicitly restricted to direct mechanisms, and they focus on truthful implementation.[24] Our paper shows that the restriction to direct, truthful mechanisms is without loss if α is most-discerning. If α is not most-discerning, then even untruthful direct mechanisms may be insufficient. Caragiannis et al. (2012) allow the principal to use arbitrarily severe punishments to deter the agent from misreporting. If one type cannot mimic another type perfectly, then he risks being detected and facing a prohibitive fine. Therefore, this setting reduces to partial verification. In our model, the agent can walk away at any time, so punishment is limited.

Closest to us is the independent paper of Ben-Porath et al. (2019a). In their model, the principal does not have the power to control which test is conducted. Instead, there is an abstract set of pieces of evidence, and the

[22] Strausz (2016) also re-interprets Green and Laffont's (1986) framework. They model verification as a component of the outcome and recover the revelation principle.

[23] Evidence was introduced in games (without commitment) by Milgrom (1981) and Grossman (1981); for recent work on evidence games, see Hart et al. (2017), Ben-Porath et al. (2017), and Koessler and Perez-Richet (2017).

[24] Dziuda and Salas (2018) and Balbuzanov (2019) study a setting without commitment in which these probabilities are constant. For some fixed $p \in (0, 1)$, the authentication probability $\alpha(\theta'|\theta)$ equals p if $\theta' \neq \theta$ and 1 if $\theta' = \theta$. This authentication satisfies (6) and hence is most discerning.

agent's type determines the agent's feasible set of distributions over subsets of pieces of evidence. Our setting fits within their framework if the agent's choice of evidence-gather activity is observable.[25] Our papers have different aims. Ben-Porath et al. (2019a) study the relationship between the evidence-acquisition and the signal-choice model in the most general setting. Our focus is on obtaining a tractable framework in which we can apply the first-order approach.

If the environment is most discerning, tests can also be interpreted as *stochastic evidence*. Each test τ in T corresponds to a request for a particular piece of evidence. The agent is asked to send a cheap talk message to the principal after he learns his payoff type but *before* he learns which evidence is available to him. With probability $\pi(\tau|\theta)$, type θ will have the evidence requested by test τ. Deneckere and Severinov (2008) study a different kind of stochastic evidence. In their model the agent simultaneously learns his payoff type and his set of feasible evidence messages.

8 Conclusion

We model probabilistic verification as a family of noisy tests that are available to the principal. With probabilistic verification, small lies are detected with positive probability. Therefore, verification relaxes the *local* incentive constraints, and our model is amenable to the first-order approach. We illustrate this approach in a few classical revenue-maximization problems. We believe this local approach will make verification tractable in other settings as well.

As the precision of the verification technology varies, our setting continuously interpolates between private information and complete information. Thus we can quantify the value to the principal of a particular verification technology. This is the first step toward analyzing a richer setting in which

[25] Formally, each piece of evidence corresponds to a pair (s, τ) of a score and a test. For each test τ and type θ, we view the measure $\pi_{\tau,\theta}$ as concentrating on the copy of S embedded in $S \times T$ as $S \times \{\tau\}$. Formally, $\mu_{\tau,\theta}(A) = \pi_{\tau,\theta}(\{s : (s, \tau) \in A\})$ for $A \subset S \times T$. With this notation, test τ is more θ-discerning than test ψ in our framework if and only if $\mu_{\tau,\theta}$ is more informative than $\mu_{\psi,\theta}$ in theirs.

the principal decides how much to invest in verification.

A Proofs

A.1 Proof of Proposition 1

Let $S = \{0, 1\}$. Consider a mechanism $(M_1, M_2; t, r_2, g)$ and a strategy (r_1, a). For each fixed type θ, the sequence (m_1, τ, m_2, s, x) in $M_1 \times T \times M_2 \times S \times X$ is realized according to the following procedure:

- Agent sends $m_1 \sim r_1(\theta)$
- Principal selects $\tau \sim t(m_1)$
- Principal sends $m_2 \sim r_2(m_1, \tau)$
- Agent tries with probability $a(m_1, \tau, m_2)$
- Nature draws s according to π and Agent's try/skip choice
- Principal selects $x \sim g(m_1, \tau, m_2, s)$

This distribution of the sequence (m_1, τ, m_2, s, x) is replicated by the following canonical procedure:

- Agent sends $\theta' = \theta$
- Principal privately draws $m_1 \sim r_1(\theta')$ and then selects $\tau \sim t(m_1)$
- Agent tries
- Nature draws s according to π and Agent's try/skip choice
- Principal privately draws $m_2 \sim r_2(m_1, \tau)$ and then privately draws s' from s by applying the transition with $k_1 = a(m_1, \tau, m_2)$ and $k_0 = 0$, and finally selects $x \sim g(m_1, \tau, m_2, s')$

We check that the outcome of any deviation by type θ in the new mechanism can be replicated by a deviation in the old mechanism. It follows that such a deviation cannot be profitable. If type θ (i) reports $\theta' \sim \rho$ and (ii) tries with probability $\alpha(\theta', \tau)$, this can be replicated in the old mechanism by (i) drawing $\theta' \sim \rho$ and then sending $m_1 \sim r_1(\theta')$; and (ii) trying with probability $a(m_1, \tau, m_2)\alpha(\theta', \tau)$.[26]

[26] This argument relies in two places on a form of randomization that our model does not technically allow. In the canonical mechanism, the principal remembers her privately drawn m_1 and uses it to select x. In the replicating deviation in the original mechanism, the

A.2 Proof of Theorem 1

Let f be a social choice rule that is canonically implemented by a mechanism (t, g) in which $t(\theta) = \psi$. Consider a new mechanism $(\hat{t}, \hat{g})$ that coincides with (t, g) except for the following modifications. Set $\hat{t}(\theta) = \tau$. Choose k_1 and k_0 from the definition of $\tau \succeq_\theta \psi$. For each $s = 0, 1$, set

$$\hat{g}(\theta, \tau, s) = k_s g(\theta, \psi, 1) + (1 - k_s) g(\theta, \psi, 0) \in \Delta(X).$$

Under this new mechanism, if type θ' reports type θ and tries on test τ with probability a, the resulting decision will be

$$p(a|\theta') g(\theta, \psi, 1) + (1 - p(a|\theta')) g(\theta, \psi, 0) \in \Delta(X),$$

where

$$p(a|\theta') = a \left[\pi(\tau|\theta') k_1 + (1 - \pi(\tau|\theta')) k_0\right] + (1 - a) k_0.$$

From the definition of $\tau \succeq_\theta \psi$, we have $p(a|\theta') \leq \pi(\psi|\theta')$ for all a in $[0, 1]$ and all types θ', with equality if $a = 1$ and $\theta' = \theta$. Therefore, the new mechanism replicates the social choice rule f without introducing any new deviations.

A.3 Proof of Proposition 2

Fix a type θ and tests τ_1 and τ_2.

One direction is clear. If $\pi(\tau_1|\cdot) = \pi(\tau_2|\cdot)$, then setting $(k_0, k_1) = (0, 1)$ shows that $\tau_1 \succeq_\theta \tau_2$ and $\tau_2 \succeq_\theta \tau_1$. If θ is minimal on test τ_1 then we get $\tau_2 \succeq_\theta \tau_1$ by setting $k_0 = k_1 = \pi(\tau_1|\theta)$. Symmetrically, if θ is minimal on τ_2, then we have $\tau_1 \succeq_\theta \tau_2$ by setting $k_0 = k_1 = \pi(\tau_2|\theta)$.

For the converse, assume $\tau_1 \sim_\theta \tau_2$. Choose (k_0, k_1) from the definition of $\tau_1 \succeq_\theta \tau_2$ and (k'_0, k'_1) from the definition of $\tau_2 \succeq_\theta \tau_1$. Suppose type θ is not minimal on one of the tests, say τ_1. Hence there exists θ' such that $\pi(\tau_1|\theta) > \pi(\tau_1|\theta')$.

agent remembers his privately drawn θ' before choosing whether to try. This can be handled with conditional probabilities; see Appendix C.3. The principal redraws m_1 conditional on (θ', τ). The agent redraws θ' conditional on (θ, m_1).

In Markov transition notation, writing $\pi_{\tau|\theta}$ for the measure that puts probability $\pi(\tau|\theta)$ on $s = 1$, we have k and k' such that

$$\pi_{\tau_1|\theta} k k' = \pi_{\tau_2|\theta} k' = \pi_{\tau_1|\theta} \qquad \text{and} \qquad \pi_{\tau_1|\theta} k k' \preceq_{\text{st}} \pi_{\tau_2|\theta'} k' \preceq_{\text{st}} \pi_{\tau_1|\theta'}.$$

In terms of probabilities, we get the system

$$\begin{aligned} k_0 k_1' + (1 - k_0) k_0' + \pi(\tau_1|\theta)(k_1 - k_0)(k_1' - k_0') &= \pi(\tau_1|\theta), \\ k_0 k_1' + (1 - k_0) k_0' + \pi(\tau_1|\theta')(k_1 - k_0)(k_1' - k_0') &\leq \pi(\tau_1|\theta'). \end{aligned}$$

Subtracting we conclude that

$$[\pi(\tau_1|\theta) - \pi(\tau_1|\theta')](k_1 - k_0)(k_1' - k_0') \geq \pi(\tau_1|\theta) - \pi(\tau_1|\theta').$$

Since $\pi(\tau_1|\theta) - \pi(\tau_1|\theta') > 0$, it follows that $(k_0, k_1) = (k_0', k_1') = (0, 1)$, and hence $\pi(\tau_1|\cdot) = \pi(\tau_2|\cdot)$.

A.4 Proof of Theorem 2

Sufficiency Let $\hat{t}$ be a most-discerning testing function. For each type θ and test ψ, select probabilities $k_0(\theta, \psi)$ and $k_1(\theta, \psi)$ as in the definition of $\hat{t}(\theta) \succeq_\theta \psi$; Appendix C.4 shows that there exists a *measurable* selection. Let f be an implementable social choice rule. By the revelation principle (Proposition 1), there exists a canonical mechanism $t \colon \Theta \to \Delta(T)$ and $g \colon \Theta \times T \times \{0, 1\} \to \Delta(X)$ that implements f. Define a mechanism with testing rule $\hat{t}$ and outcome rule $\hat{g}$ satisfying

$$\hat{g}(\theta, \hat{t}(\theta), s) = \mathbf{E}_{\psi \sim t(\theta)} \left[k_s(\theta, \psi) g(\theta, \psi, 1) + (1 - k_s(\theta, \psi)) g(\theta, \psi, 0) \right] \in \Delta(X),$$

for all types θ and scores $s = 0, 1$.

Under this new mechanism, if type θ' reports type θ and tries on test $\hat{t}(\theta)$ with probability a, the resulting decision will be

$$\mathbf{E}_{\psi \sim t(\theta)} \left[p(a, \theta, \psi|\theta') g(\theta, \psi, 1) + (1 - p(a, \theta, \psi|\theta')) g(\theta, \psi, 0) \right] \in \Delta(X),$$

where

$$p(a, \theta, \psi|\theta') = a\left[\pi(\hat{t}(\theta)|\theta')k_1(\theta, \psi) + (1 - \pi(\hat{t}(\theta)|\theta'))k_0(\theta, \psi)\right] + (1 - a)k_0(\theta, \psi).$$

For each type θ and test ψ, the definition of $\hat{t}(\theta) \succeq_\theta \psi$ guarantees that $p(a, \theta, \psi|\theta') \leq \pi(\psi|\theta')$ for all a in $[0, 1]$ and all types θ', with equality if $a = 1$ and $\theta' = \theta$. Therefore, the new mechanism replicates the social choice rule f without introducing any new deviations.

Necessity Fix a type θ and a test ψ. We will prove that $\hat{t}(\theta) \succeq_\theta \psi$.

Construct a decision environment (X, u) as follows. The decision set X consists of three decisions, denoted $\bar{x}$, $\underline{x}$, and x_0. Every type gets utility 1 from decision $\bar{x}$ and utility 0 from decision $\underline{x}$. Each type θ' gets utility $\pi(\hat{t}(\theta)|\theta')$ from decision x_0.

Consider the following mechanism. If the agent reports $\theta' \neq \theta$, the principal selects x_0 (the test and score don't matter). If the agent reports θ, the principal conducts test ψ and then selects $\bar{x}$ if the agent passes and $\underline{x}$ if the agent fails. Truthtelling and trying is a best response for every type. Denote the induced social choice rule by f.

By assumption, f can be canonically implemented with $t(\theta) = \hat{t}(\theta)$. For $s = 0, 1$, let k_s be the probability that the principal selects $\bar{x}$ after the agent reports θ and receives score s on test τ. We must have $k_1 \geq k_0$, or else type θ could profitably deviate by skipping the test. Since this mechanism implements f, the probabilities k_0 and k_1 satisfy (i) in the definition of $\tau \succeq_\theta \psi$. Since each type θ' cannot profit from reporting θ and trying on test $\hat{t}(\theta)$, we get (ii). Therefore, $\hat{t}(\theta) \succeq_\theta \psi$ as desired.

A.5 Proof of Proposition 3

Fix type θ and tests τ and ψ. According to the definition, we have $\tau \succeq_\theta \psi$ if and only if there exist probabilities k_1 and k_0 satisfying

$$k_1 \geq k_0 \quad \text{and} \quad \pi(\tau|\theta)k_1 + (1 - \pi(\tau|\theta))k_0 = \pi(\psi|\theta) \tag{10}$$

such that for all $\theta' \neq \theta$, we have

$$\pi(\tau|\theta')k_1 + (1 - \pi(\tau|\theta'))k_0 \leq \pi(\psi|\theta'). \tag{11}$$

Taking $k_0' = 1 - k_1$ and $k_1' = 1 - k_0$, we can state this condition symmetrically in terms of failure rather than passage. Let $\bar{pi} = 1 - \pi$. We have $\tau \succeq_\theta \psi$ if and only if there exist probabilities k_1' and k_0' satisfying

$$k_1' \geq k_0' \quad \text{and} \quad \bar{\pi}(\tau|\theta)k_1' + (1 - \bar{\pi}(\tau|\theta))k_0' = \bar{\pi}(\psi|\theta) \tag{12}$$

such that for all $\theta' \neq \theta$ we have

$$\bar{\pi}(\tau|\theta')k_1' + (1 - \bar{\pi}(\tau|\theta'))k_0' \geq \bar{\pi}(\psi|\theta'). \tag{13}$$

We split into (overlapping) cases.

1. If $\pi(\tau|\theta) \geq \pi(\psi|\theta)$ and $\pi(\tau|\theta) > 0$ then the solutions of (10) are given by

$$\begin{bmatrix} k_0 \\ k_1 \end{bmatrix} = \lambda \begin{bmatrix} 0 \\ \pi(\psi|\theta)/\pi(\tau|\theta) \end{bmatrix} + (1 - \lambda) \begin{bmatrix} \pi(\psi|\theta) \\ \pi(\psi|\theta) \end{bmatrix}, \qquad \lambda \in [0, 1].$$

   Plug these parameterized solutions into (11) and group terms by λ. We have $\tau \succeq_\theta \psi$ if and only if there exists λ in $[0, 1]$ such that

$$\lambda \frac{\pi(\tau|\theta')\pi(\psi|\theta)}{\pi(\tau|\theta)} + (1 - \lambda)\pi(\psi|\theta) \leq \pi(\psi|\theta').$$

   Dividing by $\pi(\psi|\theta)$ gives the desired inequality.

2. If $\bar{\pi}(\tau|\theta) \geq \bar{\pi}(\psi|\theta)$ and $\bar{\pi}(\tau|\theta) > 0$, then the solutions of (12) are given by

$$\begin{bmatrix} k_0' \\ k_1' \end{bmatrix} = \lambda \begin{bmatrix} 0 \\ \bar{\pi}(\psi|\theta)/\bar{\pi}(\tau|\theta) \end{bmatrix} + (1 - \lambda) \begin{bmatrix} \bar{\pi}(\psi|\theta) \\ \bar{\pi}(\psi|\theta) \end{bmatrix}, \qquad \lambda \in [0, 1].$$

   Plug these parameterized solutions into (10) and group terms by λ. We have $\tau \succeq_\theta \psi$ if and only if there exists λ in $[0,1]$ such that

$$\lambda \frac{\bar{\pi}(\tau|\theta')\bar{\pi}(\psi|\theta)}{\bar{\pi}(\tau|\theta)} + (1-\lambda)\bar{\pi}(\psi|\theta) \geq \bar{\pi}(\psi|\theta').$$

Dividing by $\bar{\pi}(\psi|\theta)$ gives the desired inequality.

A.6 Proof of Proposition 4

Given an authentication rate α, define the induced testing technology by $T = \{\tau_{\theta'} : \theta' \in \Theta\}$ and $\pi(\tau_{\theta'}|\theta) = \alpha(\theta'|\theta)$. The function $\theta \mapsto \tau_\theta$ is most-discerning if and only if for all distinct types θ_2 and θ_3, we have $\tau_{\theta_2} \succeq_{\theta_2} \tau_{\theta_3}$. Now apply Proposition 3 to the relation $\tau_{\theta_2} \succeq_{\theta_2} \tau_{\theta_3}$ (with the deviating type θ' labelled as θ_1) and substitute α in for π.

A.7 Proof of Proposition 5

We apply the dominated convergence theorem. As λ converges to 0 pointwise, $\alpha(\theta|\xi)$ converges to 1 for all θ and ξ with $\theta \leq \xi$. Hence $\varphi(\theta)$ converges to $\varphi^M(\theta)$, for each θ. Likewise, as λ converges to ∞ pointwise, $\alpha(\theta|\xi)$ converges to 0 for all θ and ξ with $\theta < \xi$. Hence $\varphi(\theta)$ converges to θ, for each θ.

A.8 Proof of Proposition 6

The equality $U(\theta) = \theta q(\theta) - t(\theta)$ gives a one-to-one correspondence between the transfer function t and the indirect utility function U, given the quantity function q.

Lemma 1 (Envelope theorem lower bound)
Let q be a bounded quantity function, and let U be an indirect utility function. If (q, U) is incentive compatible, then for each type θ, we have

$$U(\theta) \geq \int_{\underline{\theta}}^{\theta} \alpha(\xi|\theta) q(\xi)\,\mathrm{d}\xi. \tag{14}$$

We prove the theorem, taking as given Lemma 1, which is proven in Appendix A.9. Pick a bounded quantity function $q\colon \Theta \to \mathbf{R}_+$. The principal's objective function can be decomposed as the difference between the total surplus and the agent's rents:

$$\int_{\underline{\theta}}^{\bar{\theta}} [\theta q(\theta) - c(q(\theta))] f(\theta)\,\mathrm{d}\theta - \int_{\underline{\theta}}^{\bar{\theta}} U(\theta)\,\mathrm{d}\theta.$$

Plug in the bound from Lemma 1 and switch the order of integration to obtain the following upper bound on the principal's objective:

$$V(q) \leq \int_{\underline{\theta}}^{\bar{\theta}} [\varphi(\theta)q(\theta) - c(q(\theta))] f(\theta)\,\mathrm{d}\theta,$$

with equality if (14) holds with equality for every type θ.

The quantity function q^* from the theorem statement maximizes the expression in brackets pointwise, and t^* is the corresponding transfer for which U satisfies (14) with equality for every type θ.

We check that global incentive compatibility holds if the quantity function q^* satisfies the following monotonicity condition: Whenever $\xi_1 \leq \xi_2 \leq \theta$, we have

$$\alpha(\xi_1|\theta)q^*(\xi_1) \leq \alpha(\xi_2|\theta)q^*(\xi_2). \tag{15}$$

This condition holds if the virtual value φ is increasing.

The incentive constraint requires that for all all types θ and θ', we have

$$\begin{aligned} U(\theta) &\geq \alpha(\theta'|\theta)[\theta q^*(\theta') - t^*(\theta')] \\ &= \alpha(\theta'|\theta)[U(\theta') + (\theta - \theta')q^*(\theta')], \end{aligned} \tag{16}$$

or equivalently,

$$U(\theta) - \alpha(\theta'|\theta)U(\theta') \geq (\theta - \theta')\alpha(\theta'|\theta)q^*(\theta'). \tag{17}$$

Plug in the expression for U to get

$$\int_{\underline{\theta}}^{\theta} \alpha(\xi|\theta)q^*(\xi)\,\mathrm{d}\xi - \int_{\underline{\theta}}^{\theta'} \alpha(\xi|\theta')\alpha(\theta'|\theta)q^*(\xi) \geq (\theta - \theta')\alpha(\theta'|\theta)q^*(\theta').$$

It is easy to check that this inequality is implied by (15).

A.9 Proof of Lemma 1

First, we check that U is absolutely continuous. Choose θ and θ' such that $U(\theta') \geq U(\theta)$. From incentive-compatibility (16), we have

$$\begin{aligned} 0 &\leq U(\theta') - U(\theta) \\ &\leq (1 - \alpha(\theta'|\theta))U(\theta') + \alpha(\theta'|\theta)(\theta - \theta')q(\theta') \\ &\leq (1 - \alpha(\theta'|\theta))\theta' q(\theta') + \alpha(\theta'|\theta)(\theta - \theta')q(\theta') \\ &\leq (1 - \alpha(\theta'|\theta))\bar{\theta}\|q\|_\infty + (\theta - \theta')\|q\|_\infty. \end{aligned}$$

Since $1 - e^{-x} \leq x$, it follows that

$$|U(\theta) - U(\theta')| \leq C \left| \int_{\theta'}^{\theta} (\lambda(\xi) + 1)\,\mathrm{d}\xi \right|,$$

where $C = \max\{\bar{\theta}, 1\}\|q\|_\infty$. Since $\lambda + 1$ is integrable over $[\underline{\theta}, \bar{\theta}]$, we conclude that U is absolutely continuous.

Now we prove (14). Define the auxiliary function Δ on $[\underline{\theta}, \bar{\theta}]$ by

$$\begin{aligned} \Delta(\theta) &= \alpha(\theta|\bar{\theta})\left(U(\theta) - \int_{\underline{\theta}}^{\theta} \alpha(\xi|\theta)q(\xi)\,\mathrm{d}\xi \right) \\ &= \alpha(\theta|\bar{\theta})U(\theta) - \int_{\underline{\theta}}^{\theta} \alpha(\xi|\bar{\theta})q(\xi)\,\mathrm{d}\xi. \end{aligned}$$

The function Δ is absolutely continuous since it is the product of absolutely continuous functions on a compact set. Let $u(\theta'|\theta) = \alpha(\theta'|\theta)[\theta q(\theta) - t(\theta)]$.

Theorem 1 in Milgrom and Segal (2002), whenever U is differentiable, we have

$$U'(\theta) \geq D_{2+}u(\theta|\theta) = q(\theta) - \lambda(\theta)U(\theta),$$

where $D_{2+}u(\theta|\theta)$ denotes the right derivative with respect to the second argument.[27] At each point θ where all the functions involved are differentiable, which holds almost surely, we have

$$\begin{aligned}\Delta'(\theta) &= \lambda(\theta)\alpha(\theta|\bar{\theta})U(\theta) + \alpha(\theta|\bar{\theta})U'(\theta) - \alpha(\theta|\bar{\theta})q(\theta) \\ &= \alpha(\bar{\theta}|\theta)\left[U'(\theta) - (q(\theta) - \lambda(\theta)U(\theta))\right] \\ &\geq 0.\end{aligned}$$

Since $\Delta(\underline{\theta}) = 0$, the fundamental theorem of calculus implies that $\Delta(\theta) \geq 0$ for all θ, as desired.

A.10 Proof of Proposition 7

The same argument from Proposition 6 shows that the principal's value as a function of q can be written as

$$V(q) = \int_{\underline{\theta}}^{\bar{\theta}} (\varphi(\theta) - c)q(\theta)\,\mathrm{d}\theta.$$

The quantity function q^* from the theorem statement maximizes this quantity pointwise, and t^* is the corresponding transfer function. Since q^* is monotone, this mechanism is incentive compatible and hence optimal.

A.11 Proof of Proposition 8

Applying the same argument player by player gives

$$V(Q) = \int_{\Theta} \Big[\sum_{i=1}^{n} (\varphi_i(\theta_i) - c)q_i(\theta_i)\Big] f(\theta)\,\mathrm{d}\theta.$$

[27] That is, $D_{2+}u(\theta|\theta) = \lim_{h\downarrow 0} h^{-1}(u(\theta|\theta + h) - u(\theta|\theta))$.

This is maximized by the quantity functions q^* in the theorem statement, which induces monotone interim quantity functions.

The interim utility functions are then given by

$$U_i(\theta_i) = \int_{\underline{\theta}_i}^{\theta_i} \alpha_i(\xi_i|\theta_i) Q_i^*(\xi_i) \, \mathrm{d}\xi_i.$$

One choice of ex post utility functions consistent with this is

$$U_i(\theta_i, \theta_{-i}) = \int_{\underline{\theta}_i}^{\theta_i} \alpha_i(\xi_i|\theta_i) q_i^*(\xi_i, \theta_{-i}) \, \mathrm{d}\xi_i.$$

If $\theta_i \leq r_i(\theta_{-i})$, we have $q_i^*(\theta) = t_i^*(\theta) = 0$. IF $\theta_i > r_i(\theta_{-i})$, then $q_i^*(\theta) = 1$ and

$$U_i(\theta_i, \theta_{-i}) = \int_{r_i(\theta_{-i})}^{\theta_i} \alpha_i(\xi_i|\theta_i) \, \mathrm{d}\xi_i,$$

hence

$$t_i^*(\theta) = \theta_i - U_i(\theta_i, \theta_{-i}) = r_i(\theta_{-i}) + \int_{r_i(\theta_{-i})}^{\theta_i} (1 - \alpha_i(\xi_i|\theta_i)) \, \mathrm{d}\xi_i.$$

References

Aliprantis, Charalambos D. and Kim C. Border, *Infinite Dimensional Analysis: A Hitchhiker's Guide*, 3 ed., Springer, 2006. [49, 50, 51]

Balbuzanov, Ivan, "Lies and Consequences," *International Journal of Game Theory*, 2019, *48* (4), 1203–1240. [25]

Ben-Porath, Elchanan, Eddie Dekel, and Barton L. Lipman, "Optimal Allocation with Costly Verification," *American Economic Review*, 2014, *104* (12), 3779–3813. [24]

_ , _ , and Barton L Lipman, "Disclosure and choice," *The Review of Economic Studies*, 2017, *85* (3), 1471–1501. [25]

_ , _ , and _ , "Acquisition of/Stochastic Evidence," 2019. Working paper. [25, 26]

_ , _ , and _ , "Mechanisms With Evidence: Commitment and Robustness," *Econometrica*, 2019, *87* (2), 529–566. [24]

Blackwell, David, "Equivalent Comparisons of Experiments," *Annals of Mathematical Statistics*, 1953, *24* (2), 265–272. [4, 10]

Border, Kim C and Joel Sobel, "Samurai Accountant: A Theory of Auditing and Plunder," *Review of Economic Studies*, 1987, *54* (4), 525–540. [17]

Bull, Jesse and Joel Watson, "Evidence Disclosure and Verifiability," *Journal of Economic Theory*, 2004, *118* (1), 1–31. [25]

_ and _ , "Hard Evidence and Mechanism Design," *Games and Economic Behavior*, 2007, *58* (1), 75–93. [4, 24, 25]

Caragiannis, Ioannis, Edith Elkind, Mario Szegedy, and Lan Yu, "Mechanism Design: From Partial to Probabilistic Verification," in "Proceedings of the 13th ACM Conference on Electronic Commerce" EC '12 ACM New York, NY, USA 2012, pp. 266–283. [2, 17, 24, 25]

Carbajal, Juan Carlos and Jeffrey C. Ely, "Mechanism Design without Revenue Equivalence," *Journal of Economic Theory*, 2013, *148* (1), 104–133. [18]

_ and _ , "A Model of Price Discrimination under Loss Aversion and State-Contingent Reference Points," *Theoretical Economics*, 2016, *11*, 455–485. [18]

Cohn, Donald L, *Measure Theory* Birkhäuser Advanced Texts Basler Lehrbücher, 2 ed., Birkhäuser Basel, 2013. [48]

Crocker, Keither J. and John Morgan, "Is Honesty the Best Policy? Curtailing Insurance Fraud through Optimal Incentive Contracts," *Journal of Political Economy*, 1998, *106* (2), 355–375. [14]

de Oliveira, Henrique, "Blackwell's Informativeness Theorem Using Diagrams," *Games and Economic Behavior*, 2018, *109*, 126–131. [12]

Deneckere, Raymond and Sergei Severinov, "Mechanism Design with Partial State Verifiability," *Games and Economic Behavior*, 2008, *64* (2), 487–513. [26]

_ and _ , "Screening, Signalling and Costly Misrepresentation," 2017. Working paper. [14]

Dziuda, Wioletta and Christian Salas, "Communication with Detectable Deceit," *Available at SSRN 3234695*, 2018. [25]

Erlanson, Albin and Andreas Kleiner, "Costly Verification in Collective Decisions," 2015. Working paper. [24]

Ferraioli, Diodato and Carmine Ventre, "Probabilistic Verification for Obviously Strategyproof Mechanisms," in "Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems" AAMAS '18 International Foundation for Autonomous Agents and Multiagent Systems Richland, SC 2018, pp. 1930–1932. [2, 25]

Green, Jerry R. and Jean-Jacques Laffont, "Partially Verifiable Information and Mechanism Design," *Review of Economic Studies*, 1986, *53* (3), 447–456. [2, 4, 15, 24, 25]

Grossman, Sanford J, "The informational role of warranties and private disclosure about product quality," *The Journal of Law and Economics*, 1981, *24* (3), 461–483. [25]

Halac, Marina and Pierre Yared, "Commitment vs. Flexibility with Costly Verification," 2017. Working paper. [24]

Hart, Sergiu, Ilan Kremer, and Motty Perry, "Evidence Games: Truth and Commitment," *American Economic Review*, 2017, *107* (3), 690—-713. [25]

Kallenberg, Olav, *Random Measures, Theory and Applications*, Vol. 77 of *Probability Theory and Stochastic Modelling*, Springer, 2017. [48]

Kamae, T., U. Krengel, and G. L. O'Brien, "Stochastic Inequalities on Partially Ordered Spaces," *Annals of Probability*, 1977, *5* (6), 899–912. [23, 45]

Kartik, Navin, "Strategic Communication with Lying Costs," *Review of Economic Studies*, 2009, *76* (4), 1359–1395. [14]

_ , Marco Ottaviani, and Francesco Squintani, "Credulity, Lies, and Costly Talk," *Journal of Economic Theory*, 2007, *134* (1), 93–116. [14]

Kephart, Andrew and Vincent Conitzer, "The Revelation Principle for Mechanism Design with Reporting Costs," in "Proceedings of the 2016 ACM

Conference on Economics and Computation" 2016, pp. 85–102. [14]

Koessler, Frédéric and Eduardo Perez-Richet, "Evidence Reading Mechanisms," 2017. Working paper. [25]

Lacker, Jeffrey M. and John A. Weinberg, "Optimal Contracts under Costly State Falsification," *Journal of Political Economy*, 1989, *97* (6), 1345–1363. [14]

Li, Yunan, "Mechanism Design with Costly Verification and Limited Punishments," *Journal of Economic Theory*, 2020, *186*, 1–54. [24]

Lipman, Barton L. and Duane J. Seppi, "Robust Inference in Communication Games with Partial Provability," *Journal of Economic Theory*, 1995, *66* (2), 370–405. [4, 24, 25]

Maggi, Giovanni and Andrés Rodriguéz-Clare, "Costly Distortion of Information in Agency Problems," *RAND Journal of Economics*, 1995, *26* (4), 675–689. [14]

Milgrom, Paul and Ilya Segal, "Envelope Theorems for Arbitrary Choice Sets," *Econometrica*, 2002, *70* (2), 583–601. [35]

Milgrom, Paul R, "Good News and Bad News: Representation Theorems and Applications," *The Bell Journal of Economics*, 1981, (2), 380–391. [25]

Mussa, Michael and Sherwin Rosen, "Monopoly and Product Quality," *Journal of Economic Theory*, 1978, *18*, 301–317. [19]

Myerson, Roger B, "Optimal Auction Design," *Mathematics of Operations Research*, 1981, *6* (1), 58–73. [21]

_ , "Multistage Games with Communication," *Econometrica*, 1986, *54* (2), 323–358. [8]

Mylovanov, Tymofiy and Andriy Zapechelnyuk, "Optimal Allocation with Ex Post Verification and Limited Penalties," *American Economic Review*, 2017, *107* (9), 2666—2694. [24]

Postlewaite, Andrew, "Manipulation via Endowments," *Review of Economic Studies*, 1979, *46* (2), 255–262. [24]

Riley, John and Richard Zeckhauser, "Optimal Selling Strategies: When to Haggle, When to Hold Firm," *Quarterly Journal of Economics*, 1983, *98* (2), 267–289. [20]

Strausz, Roland, "Mechanism Design with Partially Verifiable Information," 2016. Cowles Foundation Discussion Paper No. 2040. [25]

Townsend, Robert M., "Optimal Contracts and Competitive Markets with Costly State Verification," *Journal of Economic Theory*, October 1979, *21* (2), 265–293. [24]

B For Online Publication: Additional results

B.1 Most-discerning correspondences

Definition 4 (Most-discerning correspondence)**.** A subset T_0 of T is *most θ-discerning* if for each test $\psi \in T$ there exists a test $\tau \in T_0$ such that $\tau \succeq_\theta \psi$. A correspondence $\hat{T}\colon \Theta \twoheadrightarrow T$ is *most discerning* if for each type θ the set $\hat{T}(\theta)$ is most θ-discerning.

A testing rule $\hat{t}\colon \Theta \to \Delta(T)$ is *supported on* a correspondence $\hat{T}\colon \Theta \twoheadrightarrow T$ if $\operatorname{supp} \hat{t}_\theta \subset \hat{T}(\theta)$ for each $\theta \in \Theta$. The next result says that if a correspondence is most discerning, then we can restrict attention to the testing rules supported on that correspondence. To avoid measurability problems, we impose additional regularity conditions.

Theorem 3 (Implementation with a most-discerning correspondence)
Suppose that the passage rate π is continuous. Let $\hat{T}$ be a correspondence from Θ to T with closed values and a measurable graph. If $\hat{T}$ is most discerning, then for every implementable social choice function f, there exists a testing rule $\hat{t}$ supported on $\hat{T}$ such that f is canonically implementable with $\hat{t}$.

The proof is similar to the proof of Theorem 2, but we must check that there is a measurable selection for each test ψ of a test τ in $\hat{T}(\theta)$ with $\tau \succeq_\theta \psi$. See Appendix C.4.

B.2 Example of more θ-discerning with $\lambda \neq 1$

Consider a type θ and two tests τ and ψ such that $\pi(\tau|\theta)$ and $\pi(\psi|\theta)$ are equal and nonzero. In this case, test τ is more θ-discerning than test ψ if and only if there exists $\lambda \in [0, 1]$ such that

$$\lambda\pi(\tau|\theta') + (1-\lambda)\pi(\tau|\theta) \leq \pi(\psi|\theta') \quad \text{for all } \theta' \in \Theta. \tag{18}$$

The passage rates for these tests are plotted in Figure 4. The type space is an interval, plotted on the horizontal axis. For test τ, the passage rate is

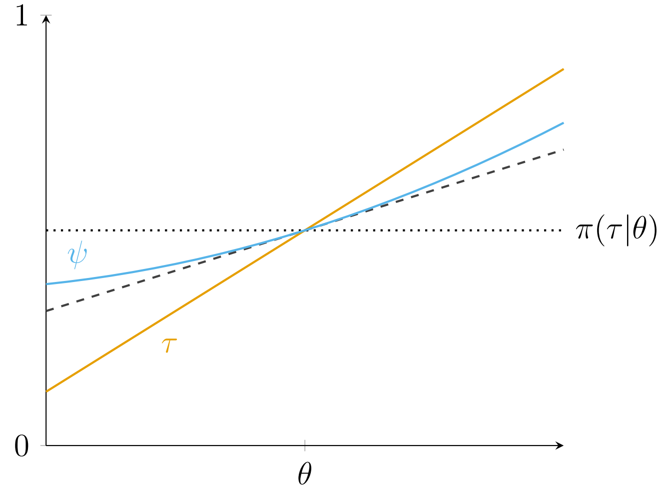

Figure 4. More θ-discerning with $\lambda \neq 1$

an increasing affine function; for test ψ, the passage rate is increasing and convex. The dotted line takes the constant value $\pi(\tau|\theta)$, and the dashed line is the average of the passage rate $\pi(\tau|\cdot)$ and the constant $\pi(\tau|\theta)$. From the graph we see that (18) is satisfied with $\lambda = 1/2$, so $\tau \succeq_\theta \psi$. Moreover, the tangency at the point $(\theta, \pi(\tau|\theta))$ shows that 1/2 is the only value of λ for which (18) holds.

B.3 Beyond exponential authentication rates

Suppose that the verification technology is represented by a measurable most-discerning authentication rate $\alpha\colon \Theta \times \Theta \to [0,1]$ that satisfies the following conditions.

(i) $\alpha(\theta|\theta) = 1$ for all types θ.

(ii) For each $\theta' \in \Theta$, the function $\theta \mapsto \alpha(\theta'|\theta)$ is absolutely continuous.

(iii) For each $\theta \in \Theta$, the right and left partial derivatives $D_{2+}\alpha(\theta|\theta)$ and $D_{2-}\alpha(\theta|\theta)$ exist, and the functions $\theta \mapsto D_{2+}\alpha(\theta|\theta)$ and $\theta \mapsto D_{2-}\alpha(\theta|\theta)$ are integrable.

Condition i ensures that the agent is authenticated if he reports truthfully. Conditions iii and ii allow us to apply the envelope theorem. Since α is most discerning, condition i implies that $\alpha(\theta_3|\theta_2)\alpha(\theta_2|\theta_1) \leq \alpha(\theta_3|\theta_1)$ for all

$\theta_1, \theta_2, \theta_3 \in \Theta$. In particular, the exponential authentication rates studied in the main text satisfy these assumptions.

Define the right and left local precision functions $\lambda_+, \lambda_- \colon \Theta \to \mathbf{R}_+$ by

$$\lambda_+(\theta) = -D_{2+}\alpha(\theta|\theta), \qquad \lambda_-(\theta) = D_{2-}\alpha(\theta, \theta).$$

Define the function Λ by

$$\Lambda(\theta'|\theta) = \begin{cases} \exp\left(-\int_{\theta'}^{\theta} \lambda_+(\xi)\,\mathrm{d}\xi\right) & \text{if } \theta \geq \theta', \\ \exp\left(-\int_{\theta}^{\theta'} \lambda_-(\xi)\,\mathrm{d}\xi\right) & \text{if } \theta < \theta'. \end{cases}$$

The function Λ is defined only by the local behavior of α near the diagonal.

Lemma 2 (Lower bound)
For all types θ and θ', we have $\alpha(\theta'|\theta) \geq \Lambda(\theta'|\theta)$.

For the exponential authentication rates α considered in the main text, we have $\lambda_+(\theta) = \lambda_-(\theta) = \lambda(\theta)$ for all types θ, and Lemma 2 holds with equality for all types θ and θ'. In this more general setting, Lemma 1 still holds with $\Lambda(\xi|\theta)$ in place of $\alpha(\xi|\theta)$.[28] In the main text, we study the relaxed problem in which only the local incentive constraints are required. Then we check that the solution of this relaxed problem is globally incentive compatible, provided that the virtual φ is increasing.[29]

Now, the solution given in the main text, with Λ in place of α remains correct, provided that

$$\alpha(\theta'|\theta) \leq \Lambda(\theta'|\theta) \frac{\int_{\underline{\theta}}^{\theta'} \Lambda(\xi|\theta) q^*(\xi)\,\mathrm{d}\xi + \int_{\theta'}^{\theta} \Lambda(\xi|\theta) q^*(\xi)\,\mathrm{d}\xi}{\int_{\underline{\theta}}^{\theta'} \Lambda(\xi|\theta) q^*(\xi)\,\mathrm{d}\xi + \int_{\theta'}^{\theta} \Lambda(\theta'|\theta) q^*(\theta')\,\mathrm{d}\xi},$$

[28] The proof is almost identical. To establish absolute continuity apply Lemma 2 and take $\lambda = \lambda_+ \vee \lambda_-$. To establish the bound, use Λ in place of α in the definition of the auxiliary function Δ.

[29] With verification, an increasing quantity function is sufficient, but no longer necessary for global incentive compatibiilty. Therefore, the condition that is increasing is sufficient but not necessary.

whenever $\theta > \theta'$. Intuitively, the more rapidly the map $\xi \mapsto \Lambda(\xi|\theta)q^*(\xi)$ increases over the interval $[0, \theta]$, the more slack there is for $\alpha(\theta'|\theta)$ to increase above $\Lambda(\theta'|\theta)$.

B.4 Testing multiple agents

We extend our model to allow for n agents, labeled $i = 1, \ldots, n$. Each agent i independently draws his type $\theta_i \in \Theta_i$ from a commonly known distribution. Set $\Theta = \prod_{i=1}^n \Theta_i$. The decision set is denoted by X, as before. Each agent i has utility function $u_i \colon X \times \Theta \to \mathbf{R}$; the principal has utility function $v \colon X \times \Theta \to \mathbf{R}$.

For each agent i, there is a set T_i of tests and a passage rate

$$\pi_i \colon T_i \times \Theta_i \to [0, 1],$$

where $\pi_i(\tau_i|\theta_i)$ is the probability with which type θ_i can pass test τ_i. Set $T = \prod_{i=1}^n T_i$. Each agent sees his own test—but not the tests of the other agents—and then chooses whether to exert effort. Nature try–skip score for each agent independently. The equilibrium concept is Bayes–Nash equilibrium.

In this multi-agent setting, the revelation principle (Proposition 1) holds. The discernment orders extend. For each agent i and each type θ_i in Θ_i, the θ_i-discernment order $\succeq_{\theta_i}$ over T_i is defined as in the baseline model with π_i in place of π. Given testing functions $t_i \colon \Theta_i \to T_i$ for each i, define the product testing function $\otimes_i t_i \colon \Theta \to T$ by $t(\theta_1, \ldots, \theta_n) = (t_1(\theta_1), \ldots, t_n(\theta_n))$. We get the natural generalization of our main result, with a similiar proof.

Theorem 4 (Most-discerning implementation with multiple players)
Fix a type space Θ and a testing environment (T, π). For a testing function $\hat{t} = \otimes_i \hat{t}_i$, the following are equivalent.

1. *$\hat{t}_i$ is most discerning for all i.*
2. *In every decision environment (X, u), every implementable social choice function is canonically implementable with $\hat{t}$.*

B.5 Nonbinary tests

For the pass–fail tests in the main model, the agent chooses a probability a in $[0, 1]$ of trying on the test. Mathematically, trying with probability a means that when the score $s = 1$ is realized, it is converted to $s = 0$ with probability a. The score $s = 0$ remains the same. With nonbinary tests, the agent formally chooses a transition $d\colon S \to \Delta(S)$ that is in the sense that $d(s'|s) = 0$ unless $s \succeq s'$.

With binary tests, we could equivalently allow the agent to directly chooses a passage probability on test τ in the interval $[0, \pi(\tau|\theta)]$. Similarly, in the general case, the agent can choose on test τ a score distribution p in $\Delta(S)$ satisfying $\pi_{\tau|\theta} \succeq_{\text{st}} p$. For probability measures μ and ν on S we have $\mu \succeq_{\text{st}} \nu$ if and only if there exists a downward transition d such that $\mu d = \nu$. (Kamae et al., 1977, Theorem 1, p. 900). To avoid technical issues with measurable selection, it is easiest to assume that the agent directly chooses a downward transition d.

In this setting with nonbinary tests, the following results go through: the revelation principle (Proposition 1), the replacement theorem (Theorem 1), and the forward implication in Theorem 2. In the proofs, simply replace the trying choice a in $[0, 1]$ with a downward transition $d\colon S \to \Delta(S)$. The key property for the proofs is that the composition of downward transitions is downward, which is easy to check.

All of the results in terms of authentication rates and passage rate ratios rely on the binary setting. We could obtain a similar reduction to an authentication distribution $\alpha\colon \Theta \times \Theta \to \Delta(S)$. The backward implication in Theorem 2 is more subtle. The proof uses the fact that there is a single increasing function f such that $\mu \succeq_{\text{st}} \nu$ if and only if $\mu f \le \nu f$. If $\hat{t}$ is not most-discerning, then any implementation will introduce a new deviation, but the type space must be sufficiently rich to ensure that every such deviation is profitable. For Proposition 2, say that type θ is minimal on test τ if $\pi_{\tau|\theta} \preceq_{\text{st}} \pi_{\tau|\theta'}$ for all types θ'. In this case conditions (a) and (b) are each sufficient for θ-equivalence, but they are no longer necessary.

B.6 Proof of Lemma 2

Fix θ and θ'. For each h, transitivity gives

$$\alpha(\theta'|\theta + h) \geq \alpha(\theta'|\theta)\alpha(\theta|\theta + h).$$

Subtract $\alpha(\theta'|\theta)$ from each side to get

$$\begin{aligned}\alpha(\theta'|\theta + h) - \alpha(\theta'|\theta) &\geq \alpha(\theta'|\theta)(\alpha(\theta|\theta + h) - 1) \\ &= \alpha(\theta'|\theta)[\alpha(\theta + h, \theta) - \alpha(\theta|\theta)].\end{aligned}$$

Dividing by h and passing to the limit as $h \downarrow 0$ and $h \uparrow 0$ gives

$$D_{2+}\alpha(\theta'|\theta) \geq -\lambda_{+}(\theta)\alpha(\theta'|\theta) \quad \text{and} \quad D_{2-}\alpha(\theta'|\theta) \leq \lambda_{-}(\theta)\alpha(\theta'|\theta).$$

Now we use absolute continuity to convert these local bounds into global bounds. Fix a report θ'. Define the function Δ on $[\underline{\theta}, \bar{\theta}]$ by

$$\Delta(\theta) = \frac{\alpha(\theta'|\theta)}{\Lambda(\theta'|\theta)}.$$

By construction, $\Delta(\theta') = 1$. We will argue that $\Delta(\theta) \geq 1$ for all θ. For $\theta' < \theta$, if $D_{2+}\Lambda(\theta'|\theta)$ exists, then

$$D_{+}\Delta(\theta) = \frac{1}{\Lambda(\theta'|\theta)} \left(D_{+}\alpha(\theta'|\theta) + \lambda_{+}(\theta)\alpha(\theta'|\theta)\right) \geq 0.$$

For $\theta' > \theta$, if $D_{2-}\Lambda(\theta'|\theta)$ exists, then

$$D_{-}\Delta(\theta) = \frac{1}{\Lambda(\theta'|\theta)} \left(D_{-}\alpha(\theta'|\theta) - \lambda_{-}(\theta)\alpha(\theta'|\theta)\right) \leq 0,$$

Since $\Lambda(\theta'|\cdot)$ is absolutely continuous, these inequalities hold almost surely. Moreover, the product of absolutely continuous functions on a compact set is absolutely continuous, so Δ is absolutely continuous, and hence the fundamental theorem of calculus gives $\Delta(\theta) \geq 1$, as desired.

C For Online Publication: Measurability

This section addresses potential measurability problems in the main text. There are two places that require arbitrary choices. To access results on analytic sets, we slightly weaken the measurability requirements to the mechanism, and then we apply measurable selection theorems to ensure that there exists selections that are measurable in this weaker sense.

C.1 Defining mechanisms and strategies

In the model, we make the following technical assumptions. The terms will be defined below. The sets Θ, T, and X are all Polish spaces. The finite signal space S is endowed with the discrete topology. The Markov transition π is from $(\Theta \times T, \mathcal{B}(\Theta \times T))$ to $(S, \mathcal{B}(S))$, where $\mathcal{B}$ denotes the Borel σ-algebra. In a mechanism, the message spaces M_1 and M_2 are Polish. Testing rules, decision rules, reporting strategies, and downward transitions are all Markov transitions, with the domain and codomain endowed with the universal completions of their Borel (product) σ-algebras.

C.2 Measurability and universal completions

Let $(X, \mathcal{X})$ and $(Y, \mathcal{Y})$ be measurable spaces. A function f from X to Y is $\mathcal{X}/\mathcal{Y}$-measurable if $f^{-1}(B)$ is in $\mathcal{X}$ for all B in $\mathcal{Y}$. This condition is written more compactly as $f^{-1}(\mathcal{Y}) \subset \mathcal{X}$. If the σ-algebra $\mathcal{Y}$ is understood, we say f is $\mathcal{X}$-measurable, and if both σ-algebras are understood, we say that f is measurable.

Let $(X, \mathcal{X}, \mu)$ be a probability space. A set A in $\mathcal{X}$ is a *μ-null set* if $\mu(A) = 0$. The *μ-completion* of σ-algebra $\mathcal{X}$, denoted $\overline{\mathcal{X}}_\mu$, is the smallest σ-algebra that contains every set in $\mathcal{X}$ and every subset of every μ-null set. It is straightforward to check that a subset A of X is a member of $\overline{\mathcal{X}}_\mu$ if and only if there are sets A_1 and A_2 in $\mathcal{X}$ such that $A_1 \subset A \subset A_2$ and $\mu(A_2 \setminus A_1) = 0$.

The *universal completion* $\overline{\mathcal{X}}$ of $\mathcal{X}$ is the σ-algebra on X defined by

$$\overline{\mathcal{X}} = \cap_\mu \overline{\mathcal{X}}_\mu,$$

where the intersection is taken over all probability measures on $(X, \mathcal{X})$.

It is convenient to work with the universal completion because of the following measurable projection theorem (Cohn, 2013, Proposition 8.4.4, p. 264).

Theorem 5 (Measurable projection)
Let $(X, \mathcal{X})$ be a measurable space, Y a Polish space, and C a set in the product σ-algebra $\mathcal{X} \otimes \mathcal{B}(Y)$. Then the projection of C on X belongs to $\overline{\mathcal{X}}$.

By taking universal completions, we do not lose any Markov transitions.

Lemma 3 (Completing transitions)
A Markov transition k from $(X, \mathcal{X})$ to $(Y, \mathcal{Y})$ can be uniquely extended to a Markov transition $\bar{k}$ from $(X, \overline{\mathcal{X}})$ to $(Y, \overline{\mathcal{Y}})$.

Next we consider the universal completion of a product σ-algebra.

Lemma 4 (Product spaces)
For measurable spaces $(X, \mathcal{X})$ and $(Y, \mathcal{Y})$,

$$\overline{\mathcal{X}} \otimes \overline{\mathcal{Y}} \subset \overline{\mathcal{X} \otimes \mathcal{Y}} = \overline{\overline{\mathcal{X}} \otimes \overline{\mathcal{Y}}}.$$

C.3 Disintegration of measures

In the proof of the revelation principle, we need to use conditional probabilities. More precisely, we need to use disintegration of measures (Kallenberg, 2017, Theorem 1.25, p. 39). We cannot directly apply the result to the universal completions of the Borel σ-algebras. Argue as follows. First, restrict $r \times t$ to a Markov transition from $(\Theta, \overline{\mathcal{B}(\Theta')})$ to $(M \times T, \mathcal{B}(M) \otimes \mathcal{B}(T))$ and $\hat{t}$ to a Markov transition from $(\Theta', \overline{\mathcal{B}(\Theta')})$ to $(T, \mathcal{B}(T))$. By Kallenberg (2017, Theorem 1.25, p. 39), there exists a Markov transition h from $(\Theta' \times T, \overline{\mathcal{B}(\Theta')} \otimes \mathcal{B}(T))$ to $(M, \mathcal{B}(M))$ that satisfies the desired equality for the restricted transitions from $(\Theta', \mathcal{B}(\Theta'))$ to $(M \times T, \mathcal{B}(M) \otimes \mathcal{B}(T))$. By Lemma 3, we can extend h to a Markov transition from $(\Theta \times T, \overline{\mathcal{B}(\Theta) \otimes \mathcal{B}(T)})$ to $(M, \overline{\mathcal{B}(M)})$.

C.4 Measurable selection of score conversion

Let $\mathcal{K}$ denote the space $\Delta(S)^S$ of Markov transitions on S, viewed as a subset of $\mathbf{R}^{S\times S}$, with the usual Euclidean topology and inner product $\langle\cdot,\cdot\rangle$. For $k \in \mathcal{K}$, denote by $k(s, s')$ the transition probability from s to s'.

Define the domain

$$D = \{(\theta, \tau, \psi) \in \Theta' \times T' \times T : \tau \succeq_\theta \psi\}.$$

Define the correspondence $K\colon D \twoheadrightarrow \mathcal{K}$ by putting $K(\theta, \tau, \psi)$ equal to the set of increasing maps $k\colon S \to \Delta(S)$ in $\mathcal{K}$ satisfying (i) $\pi_{\tau|\theta} k = \pi_{\psi|\theta}$, and (ii) $\pi_{\tau|\theta'} k \succeq_{\mathrm{st}} \pi(\psi|\theta')$ for all types $\theta' \neq \theta$. By the choice of domain D, the correspondence K is nonempty-valued.

Endow D with the restriction of the σ-algebra $\overline{\mathcal{B}(\Theta' \times T' \times T)}$. To prove that there exists a measurable selection $\hat{k}$ from K, we apply the Kuratowski–Ryll-Nardzewski selection theorem (Aliprantis and Border, 2006, 18.13, p. 600). The correspondence K has compact convex values, so it suffices to check that associated support functions for K are measurable (Aliprantis and Border, 2006, 18.31, p. 611).

Fix $\ell \in \mathbf{R}^{S\times S}$. Define the map $C\colon D \to \mathbf{R}$ by

$$C(\theta, \tau, \psi) = \max_{k \in K(\theta,\tau,\psi)} \langle k, \ell\rangle.$$

It suffices to show that C is $\overline{\mathcal{B}(\Theta' \times T' \times T)}$-measurable. Define a sequence of auxiliary functions $C_m\colon D \times (\Theta')^m \to \mathbf{R}$ as follows. Let $C_m(\theta, \tau, \psi, \theta'_1, \ldots, \theta'_m)$ be the value of the program

$$\begin{aligned} \text{maximize} \quad & \langle k, \ell\rangle \\ \text{subject to} \quad & k \in \mathcal{K} \\ & k \text{ is increasing} \\ & \pi_{\tau|\theta} k = \pi_{\psi|\theta} \\ & \pi_{\tau|\theta'_j} k \preceq_{\mathrm{st}} \pi_{\psi|\theta'_j}, \quad j = 1, \ldots, m. \end{aligned}$$

This is a standard linear programming problem with a compact feasible set. By Berge's theorem (Aliprantis and Border, 2006, 17.30, p. 569), the value of the linear program is upper semicontinuous (and hence Borel) as a function of the coefficients appearing in the constraints. Since π is Borel, so is each function C_m. By the measurable projection theorem (Theorem 5), each map

$$(\theta, \tau, \psi) \mapsto \inf_{\boldsymbol{\theta'} \in (\Theta')^m} C_m(\theta, \tau, \psi, \boldsymbol{\theta'})$$

is $\overline{\mathcal{B}(\Theta' \times T' \times T)}$-measurable. A compactness argument shows that[30]

$$C(\theta, \tau, \psi) = \inf_m \inf_{\boldsymbol{\theta'} \in (\Theta')^m} C_m(\theta, \tau, \psi, \boldsymbol{\theta'}),$$

so C is also $\overline{\mathcal{B}(\Theta' \times T' \times T)}$-measurable.

C.5 Measurable selection of more θ-discerning test

We prove that there exists a measurable function $\bar{t}$ from $(\Theta' \times T, \overline{\mathcal{B}(\Theta' \times T)})$ to $(T', \mathcal{B}(T'))$ such that the test $\bar{t}(\theta, \psi)$ is in $\hat{T}(\theta)$ and satisfies $\bar{t}(\theta, \psi) \succeq_\theta \psi$, for each $\theta \in \Theta$ and $\psi \in T$. Define a correspondence $H \colon \Theta' \times T \twoheadrightarrow T'$ by

$$H(\theta, \psi) = \{\tau \in T : \tau \succeq_\theta \psi\}.$$

[30]We claim that for each positive ε there exists a natural number m and a vector $\boldsymbol{\theta'} \in (\Theta')^m$ such that $C_m(\theta, \tau, \psi, \boldsymbol{\theta'}) < C(\theta, \tau, \psi) + \varepsilon$. Suppose not. For each $\theta' \in \Theta'$, let $K_{\theta'}$ be the compact set of monotone Markov transitions $k \in \mathcal{K}$ satisfying (i) $\pi_{\tau|\theta} k = \psi_{\psi|\theta}$, (ii) $\pi_{\tau|\theta'} k \preceq_{\text{st}} \pi_{\psi|\theta'}$, and (iii) $\langle k, \ell \rangle \geq C(\theta, \tau, \psi) + \varepsilon$. This family has the finite intersection property, but the intersection over all $\theta' \in \Theta'$ is empty, which is a contradiction.

Since π is continuous, the graph of H is closed in $\Theta' \times T \times T'$.[31] Since the graph of $\hat{T}$ is Borel, so is the set $\{(\theta, \psi, \tau) : \tau \in \hat{T}(\theta)\}$ and also the intersection

$$\{(\theta, \psi, \tau) : \tau \succeq_\theta \psi \text{ and } \tau \in \hat{T}(\theta)\}.$$

By the measurable projection theorem (Theorem 5), the associated correspondence from $(\Theta' \times T, \overline{\mathcal{B}(\Theta' \times T)})$ to T' is measurable. Moreover, this correspondence has closed values, so we can apply the Kuratowski–Ryll-Nardzewksi selection theorem (Aliprantis and Border, 2006, 18.13, p. 600) to obtain the desired function $\bar{t}$. ey

[31] Take a sequence $(\theta_n, \psi_n, \tau_n)$ in $\operatorname{gr} H$ converging to a limit (θ, ψ, τ) in $\Theta' \times T \times T'$. For each n, there is a monotone Markov transition k_n on S such that (i) $\pi_{\tau_n|\theta_n} k_n = \pi_{\psi_n|\theta_n}$, and (ii) $\pi_{\tau_n|\theta'} k_n \preceq_{\text{st}} \pi_{\psi_n|\theta'}$ for all $\theta' \in \Theta$. The space of Markov transitions on S is compact, so after passing to a subsequence, we may assume that k_n converges to a limit k, which must be monotone. Since π is continuous, taking limits gives (i) $\pi_{\tau|\theta} k = \pi_{\psi|\theta}$, and (ii) $\pi_{\tau|\theta'} k \preceq_{\text{st}} \pi_{\psi|\theta'}$ for each $\theta' \in \Theta$. Therefore, (θ, ψ, τ) is in $\operatorname{gr} H$.