Theoretical Economics 20 (2025), 1247–1284 1555-7561/20251247

Probabilistic verification in mechanism design

Ian Ball
Department of Economics, MIT

Deniz Kattwinkel
Department of Economics, UCL

We introduce a model of probabilistic verification in mechanism design. The principal elicits a message from the agent and then selects a test to give the agent. 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. If this condition holds, then 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 the verification technology varies, the solution continuously interpolates between the no-verification solution and full surplus extraction.

Keywords. Probabilistic verification, ordering tests, evidence.

JEL classification. D82, D86.

1. Introduction

In the standard paradigm of mechanism design, the principal elicits information from the agents, but the principal cannot *verify* whether the agents are being truthful. In many applications, however, claims about private information can be verified. Sellers have long offered discounts to certain groups such as students, seniors, or veterans, and sellers increasingly make targeted offers to different consumer segments. To verify a buyer's eligibility for an exclusive offer, many sellers use identity verification platforms such as ID.me and SheerID. If a buyer claims to be eligible for a particular offer, he is directed to a portal, which asks identifying questions or requests documentation, such as a student ID or a company email address. Depending on the buyer's responses, the

Ian Ball: ianball@mit.edu
Deniz Kattwinkel: denizkattwinkel@gmail.com
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, Princeton, 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, Stephen Morris, Jacopo Perego, Larry Samuelson, Sebastian Schweighofer-Kodritsch, Roberto Serrano, Philipp Strack, Roland Strausz, Juuso Välimäki, and Alex Wolitzky. An extended abstract for this paper appeared in EC'19. We are grateful for the opportunity to attend the 28th Jerusalem Summer School in Economic Theory, which planted the first seeds of this project. Funding from the German Research Foundation (DFG) through CRC TR 224 (Project B04) is gratefully acknowledged.

 https://doi.org/10.3982/TE6266

platform's proprietary algorithm either accepts or rejects the buyer's claim. In other contexts, governments verify income reports to determine eligibility for means-tested programs. Insurers verify the legitimacy of insurance claims. None of these verification systems is perfect—false claims sometimes go undetected. In this sense, verification is *probabilistic*.

The goal of this paper is to introduce a tractable model of probabilistic verification. A parsimonious model of probabilistic verification, directly generalizing Green and Laffont's (1986) deterministic model, would specify for any types θ and θ' the probability $\alpha(\theta'|\theta)$ with which type θ can "pass" as type θ'.[1] Call this function α the *authentication rate*. The difficulty is that with unrestricted communication, the authentication rate is generally endogenous. Whether type θ can "pass" as type θ' depends on what the principal demands of an agent who claims to be of type θ', e.g., which questions must be answered correctly or which documents must be provided.

We model *probabilistic verification* by endowing the principal with a set of pass–fail tests. A test could be a particular set of questions or a request for certain documentation. We represent a test by its type-dependent passage rate: type θ can pass test τ with probability $\pi(\tau|\theta)$. We assume that every type can intentionally fail any test. For example, the agent could leave a question blank (whether or not he knows the answer) or decline to provide the requested documentation (whether or not he has it).

The principal chooses how to utilize the testing technology within a mechanism. Formally, we consider the following protocol. The principal elicits a type report from the agent. Based on the report, the principal selects one test to give the agent. The agent sees the test and privately chooses whether to try on the test. This choice is costless. If the agent tries, then his passage probability depends on the test and his type, according to the function π. If the agent does not try, then he fails the test with certainty. The principal observes whether the agent passes or fails—but not whether the agent tried—and then makes a decision.

Our analysis proceeds in two parts—methodology and then applications.

In the first part of our analysis, we study whether there exists a canonical assignment of a test to each type. For each type θ, we introduce an associated order on tests. Intuitively, test τ is *more θ-discerning* than test ψ if type θ performs relatively better on test τ than on test ψ, compared to *every* other type. 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 similar to a Blackwell garbling of an experiment, but our order neither implies nor is implied by Blackwell's order.

We use our order on tests to simplify the principal's implementation problem. Consider two tests, τ and ψ, such that τ is more θ-discerning than ψ. Theorem 1 says that any social choice function that the principal can implement by giving test ψ to type θ can also be implemented by giving test τ to 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 gives each type the associated test. In a

[1] For example, Caragiannis, Elkind, Szegedy, and Yu (2012) and Ferraioli and Ventre (2018) take this approach.

contemporaneous paper, Ben-Porath, Dekel, and Lipman (2023) prove a related result in their model of stochastic evidence acquisition; we compare our model with theirs in Section 8.

If each type θ has an associated most-θ discerning test, then the testing technology can be represented by a single authentication rate: type θ is authenticated as type θ' if and only if he passes the most θ'-discerning test. The principal's design problem reduces to a tractable optimization problem involving this authentication rate. An authentication rate that represents a testing technology in this way is called *most-discerning*. Not all authentication rates are most discerning. We characterize the class of most-discerning authentication rates. Our condition generalizes the 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 turn to applications, taking a most-discerning authentication rate as a primitive. Unlike in models of *deterministic* verification, we can use the Myersonian local approach. Consider a seller who can imperfectly verify a potential buyer's membership in different market segments. It is more difficult for the seller to distinguish buyers who are in market segments with closer valuations. If the seller has a single indivisible good, it is no longer optimal for the seller to post one price. The seller prefers to charge higher prices to higher-valuation market segments. If a buyer in a higher segment claims to be in a lower segment, there is a chance that he is authenticated and charged a lower price. But there is also a chance that his misreport is detected and he does not receive the good. Under the optimal price schedule, these deviations are unprofitable.

To solve for profit-maximizing mechanisms in general quasilinear settings with verification, we derive a new expression for the virtual value that reflects the verification technology. As verification ranges from uninformative to perfectly informative, the virtual value increases from the classical virtual value to the true valuation. The associated revenue-maximizing mechanism continuously interpolates between the classical, no-verification solution and full surplus extraction.

The rest of the paper is organized as follows. Section 2 presents our model of testing. Section 3 discusses our modeling choices. Section 4 introduces the discernment orders and characterizes whether a single testing function suffices for all implementation. Section 5 characterizes the class of most-discerning authentication rates. Section 6 solves for revenue-maximizing mechanisms. Section 7 extends the model to allow for nonbinary tests. Section 8 discusses related literature. The conclusion is Section 9. Proofs are in the Appendix.

2. Model

We model probabilistic verification by endowing the principal with a testing technology. The principal can commit to use this technology (and communicate with the agent) however she wishes.

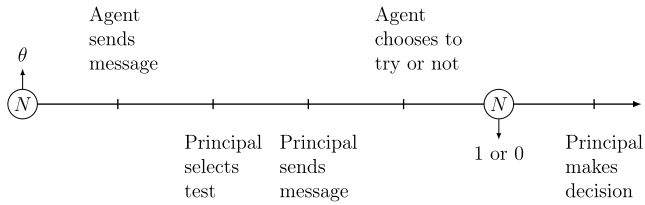

FIGURE 1. Timing.

2.1 *Setting*

Principal–agent environment There are two players: a principal (she) and an agent (he). The agent has a private type $\theta \in \Theta$, drawn from a commonly known distribution. The principal controls a decision $x \in X$.[2] The agent and the principal have bounded, type-dependent utilities $u(x, \theta)$ and $v(x, \theta)$, respectively. We extend these functions linearly to $\Delta(X) \times \Theta$.

Verification There is a testing technology (T, π), which consists of a set T of pass–fail tests and a passage rate

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

where $\pi(\tau|\theta)$ denotes the probability with which type θ can pass test τ. The spaces X, Θ, and T are assumed to be Polish spaces.[3]

The principal can give the agent one test from the set T.[4] The agent observes the selected test and chooses whether to *try* on the test. This choice is costless. If the agent tries, his passage probability is determined by π. If the agent does not try, then he fails with certainty. The principal observes the test score ("pass" or "fail"), but not whether the agent tried. Thus, there is moral hazard as well as adverse selection. In Section 3, we discuss why this modeling choice is natural in many applications. Section 7 considers tests with more than two scores.

Mechanisms and strategies The principal can commit to an arbitrary dynamic mechanism. We consider protocols of the following form, shown in Figure 1. First, the principal elicits a message from the agent. Based on the message, the principal selects a test and then sends a message to the agent. The agent sees the realized test and the message and then privately chooses whether to try on the test. 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 makes a decision.

Formally, a mechanism is a tuple $(M, M'; t, r', g)$ consisting of message spaces M and M' for the two rounds of messaging, a testing rule $t\colon M \to \Delta(T)$, a messaging rule $r'\colon M \times T \to \Delta(M')$, and an outcome rule $g\colon M \times T \times M' \times \{0, 1\} \to \Delta(X)$. A strategy for

[2]Transfers could be one component of the decision x.

[3]We further assume that the primitives u, v, and π are Borel measurable and that mechanisms are universally measurable. The details are in Appendix A.13.

[4]If the principal can give multiple tests, then the resulting compound test can be included in T. A compound test may have more scores than "pass" and "fail." Section 7 extends the model to allow for nonbinary tests.

the agent is a pair (r, a) consisting of a messaging strategy $r\colon \Theta \to \Delta(M)$ and an action strategy $a\colon \Theta \times M \times T \times M' \to [0, 1]$, which specifies the probability with which the agent tries on the test he is given.

2.2 *Implementation*

We introduce two social choice objects. A *social choice function* is a map from Θ to $\Delta(X)$, which specifies a decision lottery for each type. To keep track of which test is given, we define an *extended social choice function* to be 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 function f if (i) the strategy is a best response to the mechanism and (ii) the composition of the mechanism and the strategy induces f. An (extended) social choice function is *implementable* if there exist a mechanism and a strategy that implement it.

We show below that it is without loss to focus on a special class of direct mechanisms that induce the agent to (i) report his type truthfully and (ii) try on whichever test he is given. In these mechanisms, the principal's message to the agent is omitted. Formally, a *canonical mechanism* is a pair (t, g) consisting of a testing rule $t\colon \Theta \to \Delta(T)$ and an outcome rule $g\colon \Theta \times T \times \{0, 1\} \to \Delta(X)$, which specifies a decision lottery as a function of the reported type, the test given to the agent, and the agent's score on that test. Given such a mechanism, a strategy for the agent is a pair (r, a) consisting 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 function f is *canonically implementable* if f is implemented by some canonical mechanism (t, g) and some strategy (r, a) in which r is the identity and $a(\theta, \theta, \tau) = 1$ for all types θ and all tests τ in $\operatorname{supp} t(\theta)$. In this case, we say that (t, g) *canonically implements* f.

Proposition 1 (Revelation Principle). *Every implementable (extended) social choice function is canonically implementable.*

The proof has two parts. First, a standard argument (see Myerson (1982)) shows that every implementable social choice function can be implemented by a truthful and obedient mechanism. The second part is specific to testing. Consider a truthful, obedient mechanism. Whenever the principal recommends the agent to not try on a test, we modify the mechanism as follows. The principal recommends that the agent try on the test. Then, if the agent passes that test, the principal selects the decision as if the agent had failed. Now passing and failing result in the same decision. The agent is willing to follow the recommendation, and the resulting outcome is unchanged. Since every type can fail every test, this modification of the mechanism introduces no new deviation outcomes. Since the principal always recommends that the agent try, the principal's message conveys no information, and hence can be dropped.

3. Discussion of the model

We discuss two important features of the model: the agent's choice of whether to try on the test, and the principal's choice of a testing rule.

3.1 *Trying on the test*

When the principal gives the agent a test, the agent privately chooses whether to try or intentionally fail. If the agent fails, the principal cannot observe whether the failure was intentional. This assumption is reasonable in our motivating applications.[5] If the test asks the agent a question and the agent leaves the question blank, then the principal cannot tell whether the agent knows the answer. On an aptitude test, both high- and low-ability types are able to perform poorly.[6] If the agent does perform poorly, the principal cannot tell whether the agent is capable of performing well. Finally, if the test requests a document and the agent does not provide it, then the principal cannot tell whether the agent has the document. Indeed, our model nests previous models of deterministic hard evidence. In those models, the agent's choice to present evidence that he possesses is an "inalienable action" (Bull and Watson (2007, p. 76)).

The importance of allowing intentional failure is illustrated in the following example.

EXAMPLE 1 (Passing vs. Failing). Consider the problem of allocating a single desirable good to an agent with two possible types, θ_1 and θ_2. There is a single test. If the agent tries on this test, then type θ_1 passes with certainty and type θ_2 fails with certainty. To allocate the good to type θ_1 only, the principal can give the agent the good if and only if he passes the test. On the other hand, the principal cannot allocate the good to type θ_2 only. If the principal gives the good to the agent if and only if he fails the test, then type θ_1 would intentionally fail to receive the good.

The logic of Example 1 holds more generally. Without intentional failing, "pass" and "fail" would be arbitrary, interchangeable labels. On any test, if type θ_1 is more likely to pass than type θ_2, then type θ_2 is more likely to *fail* than type θ_1. Thus, without intentional failing, each test unavoidably links the ability of type θ_1 to mimic θ_2 with the ability of type θ_2 to mimic θ_1.

Finally, we discuss two alternative assumptions about the agent's control over the test result: (i) *observable skipping* and (ii) *exogenous scores*. Under (i), the agent cannot intentionally fail a test, but he can "skip" the test; skipping is observed by the principal. Under (ii), the agent can neither intentionally fail nor skip a test. These assumptions are nested in terms of the power afforded to the principal. Every social choice function that is implementable under our model is also implementable under (i),[7] and every social

[5] In their model of adaptive testing, Deb and Stewart (2018) make the same assumption about the agent's performance on each "task." The principal commits to an adaptive sequence of binary tasks and then assigns a final verdict—pass or fail. Each agent type wants to pass. In our model, by contrast, the principal chooses from a richer space of decisions, and different agent types have different preferences over those decisions.

[6] Myerson (1984, p. 74) gives the example of playing the piano. A good pianist can intentionally play poorly, but a bad pianist cannot play well.

[7] Given a canonical mechanism that is truthful and obedient in our model, the induced social choice function can be replicated under (i) by treating "skip" as "fail" on each test.

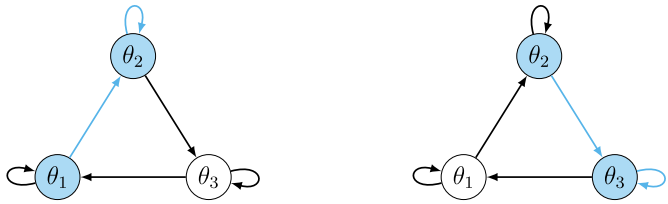

FIGURE 2. Directed graph representing a $\{0, 1\}$-valued authentication rate.

choice function implementable under (i) is also implementable under (ii).[8] In Example 1, allocating the good only to type θ_2 is implementable under both specifications (i) and (ii).[9]

3.2 *Test choice*

To cleanly separate verification from communication, we explicitly model the principal's choice of a test as part of the mechanism. As a result, the effective authentication rate faced by the agent is endogenous. Treating the authentication rate as an exogenous primitive can introduce difficulties, as illustrated in the following example, adapted from Green and Laffont (1986).

EXAMPLE 2 (Exogenous Authentication Rate). There is a single agent with three possible types, denoted θ_1, θ_2, θ_3. In Figure 2, 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 wants the good.

Each copy of the graph illustrates a social choice function. On the left, this function allocates the good to types θ_1 and θ_2 (which are shaded). This cannot be implemented by giving the good to the agent if and only if he passes as type θ_1 or as type θ_2. Then type θ_3 would pass as type θ_1 to get the good as well. Instead, the principal must give the good to the agent if and only if he passes as type θ_2. Types θ_1 and θ_2 can do so, but type θ_3 cannot. On the right, the social choice function allocates the good to types θ_2 and θ_3 (which are shaded). Symmetrically, this can be implemented only by giving the good to the agent if and only if he passes as type θ_3.

According to the directed graph, type θ_1 can "pass" as type θ_2. But type θ_1 can copy type θ_2's equilibrium strategy only in the equilibrium of the left mechanism (where type θ_2 passes as type θ_2), but not in the right mechanism (where type θ_2 passes as type θ_3).

[8]Under (i), the analogue of Proposition 1 still holds; the argument is essentially the same as in our model, with "skipping" in place of "intentionally failing." Under canonical implementation, the agent never skips a test, so the incentive constraints are preserved if the agent's option to skip a test is removed.

[9]The same social choice functions are implementable under (i) and (ii) whenever there is a decision that all types consider to be the worst (since this worst decision can be used to punish skipping). This condition is called TIWO for "type-independent worst option" in Strausz and Schweighofer-Kodritsch's (2023) model of *deterministic* evidence. For an example of a social choice function that is implementable under (ii), but not under (i), reinterpret their Example 1 (Strausz and Schweighofer-Kodritsch (2023, p. 16)).

As Example 2 illustrates, the authentication rate $\alpha(\cdot|\cdot)$ implicitly (a) introduces a family of tests, and (b) assigns to each type θ a test, so that "passing" as type θ means passing the test assigned to type θ. There is no guarantee, however, that this is the "right" assignment of tests to types. In Example 2, type θ_2 must be given different tests to implement different allocation rules. We model this test choice as part of the principal's protocol. Moreover, our model allows for an unrestricted test set (possibly larger than the type space), unrestricted communication, and test randomization.[10]

4. Ordering tests

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

4.1 *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 the agent takes test τ, his score $s_\tau \in \{0, 1\}$ can be 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 passing (rather than failing) test τ weakly increases the converted pass probability. 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 weakly 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. Our definition requires that the conversion of test τ constitutes 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 requirement that $k_1 \geq k_0$ preserves incentives, which are not relevant in the statistical framework.

Theorem 1 (Test Replacement). *Fix a type θ and tests τ and ψ such that $\tau \succeq_\theta \psi$. If a social choice function is canonically implemented by a mechanism (t, g) in which $t(\theta) = \psi$, then it is also canonically implemented by some mechanism (t', g') in which $t'(\theta) = \tau$.*

[10]Test randomization is useful if different tests are needed to deter deviations by different types; see Appendix A.2 for an example.

Here is a sketch of the proof. Start with a canonical implementation in which type θ is given test ψ. Adjust the mechanism after the report θ as follows. The principal gives the agent 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 makes the decision that she would have made in the old mechanism after score s_ψ on test ψ (following report θ).

This new mechanism implements the same social choice function. Suppose that type θ reports truthfully and tries on test τ. By (i), he will get the same decision as in the equilibrium of the original mechanism. Suppose another type θ' reports type θ and tries on test τ. By (ii), he will get a decision that he could have gotten in the original mechanism by reporting type θ and then trying on test ψ with some probability. The inequality $k_1 \geq k_0$ ensures that intentionally failing test τ also yields a decision that was achievable in the original mechanism.

For each fixed type θ, the θ-discernment order $\succeq_\theta$ is neither stronger nor weaker than Blackwell's (1953) order. 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 not suited to our setting because it does not consider the agent's incentives to intentionally fail a test. Indeed, Blackwell's order is invariant to relabeling the realizations "pass" and "fail." Our θ-discernment order is not.

Like Blackwell's order, each θ-discernment order $\succeq_\theta$ is reflexive and transitive but not generally antisymmetric. Tests τ_1 and τ_2 are *θ-equivalent*, denoted $\tau_1 \sim_\theta \tau_2$, if $\tau_1 \succeq_\theta \tau_2$ and $\tau_2 \succeq_\theta \tau_1$. If two tests have the same passage rates, then they are clearly θ-equivalent. We show that the converse holds except in the special case that neither test can screen any other type away from type θ. Formally, type θ is *minimal* on test τ if $\pi(\tau|\theta) \leq \pi(\tau|\theta')$ for all types θ'.

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 on τ_2.*

4.2 *Implementation with most-discerning testing*

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

Definition 2 (Most Discerning). A test τ is *most θ-discerning* if $\tau \succeq_\theta \psi$ for every 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.

Whether a test is most θ-discerning depends on the other tests in 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., $\pi(\hat{\tau}_\theta|\theta') = [\theta' = \theta]$, where $[\cdot]$ is the indicator function for the predicate it encloses.

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:*

(i) *$\hat{t}$ is most-discerning.*

(ii) *In every decision environment (X, u), every implementable social choice function can be canonically implemented with testing rule $\hat{t}$.*

The forward implication from condition i to condition (ii) says that a most-discerning testing function suffices for all implementation problems. In the proof, for each type θ, we apply the procedure from Theorem 1 to replace any test given to θ with the test $\hat{t}(\theta)$.

The backward implication from condition ii to condition i confirms that the most-discerning property is the right one. If $\tau \not\succeq_\theta \psi$, then replacing test ψ with test τ for type θ introduces a new deviation outcome for some type. The proof constructs a decision environment in which this deviation outcome is profitable.

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 set $\hat{T}(\theta)$ of tests with the following property: for every test ψ, there is some test τ in $\hat{T}(\theta)$ such that $\tau \succeq_\theta \psi$. Then there is no loss in assuming that the principal gives type θ only tests in $\hat{T}(\theta)$, though the principal may randomize over tests in $\hat{T}(\theta)$. See Appendix A.6 for a formal statement.

Remark 1 (Discernment Orders Under Alternative Specifications). Under the two alternative testing specifications described in Section 3.1—*observable skipping* and *exogenous scores*—the appropriate analogue of the θ-discernment order $\succeq_\theta$ is Blackwell's order, for each θ in Θ. With this redefinition of the discernment orders, it can be shown that Theorem 1 and Theorem 2 go through under each alternative specification.

4.3 *Sufficient conditions for discernment orders*

Checking whether one test τ is more-θ discerning than another test ψ amounts to verifying the feasibility of the system of linear inequalities in Definition 1. Here, we give a sufficient condition for θ-discernment in terms of relative performance.

Proposition 3 (Relative Performance). *Fix a type θ and tests τ and ψ.*[11]

(i) *Suppose $\pi(\tau|\theta) \geq \pi(\psi|\theta) > 0$. Test τ is more θ-discerning than test ψ if*

$$\frac{\pi(\tau|\theta')}{\pi(\tau|\theta)} \leq \frac{\pi(\psi|\theta')}{\pi(\psi|\theta)}, \quad \text{for all } \theta' \in \Theta. \tag{1}$$

(ii) *Suppose $\pi(\tau|\theta) \leq \pi(\psi|\theta) < 1$. Test τ is more θ-discerning than test ψ if*

$$\frac{1-\pi(\tau|\theta')}{1-\pi(\tau|\theta)} \geq \frac{1-\pi(\psi|\theta')}{1-\pi(\psi|\theta)}, \quad \text{for all } \theta' \in \Theta. \tag{2}$$

[11]The two (nonexclusive) cases exclude the following two edge cases. If $\pi(\tau|\theta) > \pi(\psi|\theta) = 0$, then $\tau \succeq_\theta \psi$. If $\pi(\tau|\theta) < \pi(\psi|\theta) = 1$, then $\tau \not\succeq_\theta \psi$, provided that $\pi(\psi|\cdot)$ is nonconstant.

In the first case, where type θ is more likely to pass test τ than test ψ, test τ is more θ-discerning than test ψ if for each type θ' the relative *passage* rate of type θ' compared with type θ is *lower* on test τ than on test ψ. In the second case, where type θ is more likely to fail test τ than test ψ, test τ is more θ-discerning than test ψ if for each type θ' the relative *failure* rate of type θ' compared with type θ is *higher* on test τ than on test ψ.

REMARK 2 (Sufficient Condition for Most θ-Discerning). In view of Proposition 3, a simple sufficient condition for test τ to be most θ-discerning is that τ maximizes $\pi(\cdot|\theta)$ and τ minimizes $\pi(\cdot|\theta')$ for each type θ' with $\theta' \neq \theta$. That is, among all tests in T, test τ is one that type θ is most likely to pass but every other type is most likely to fail.

5. TESTING IN REDUCED FORM

If the testing technology admits a most-discerning testing function, then the principal's design problem can be represented as a tractable optimization problem involving a single authentication rate α. In this section, we analyze this reduction.

Suppose that the testing technology (T, π) admits a most-discerning testing function $\hat{t}\colon \Theta \to T$. By Theorem 2, there is no loss of generality in restricting the principal to using $\hat{t}$ as the testing rule. With this testing rule, the principal selects two decisions for each report θ'—the decision, $g_1(\theta')$, if the agent passes test $\hat{t}(\theta')$ and the decision, $g_0(\theta')$, if the agent fails test $\hat{t}(\theta')$. Suppose type θ reports type θ' and then tries on test $\hat{t}(\theta')$. With probability $\pi(\hat{t}(\theta')|\theta)$, he passes and gets $g_1(\theta')$. With probability $1 - \pi(\hat{t}(\theta')|\theta)$, he fails and gets $g_0(\theta')$. Define the induced authentication rate α by

$$\alpha\big(\theta'|\theta\big) = \pi\big(\hat{t}\big(\theta'\big)|\theta\big), \quad \text{for all } \theta, \theta' \in \Theta. \tag{3}$$

For any reduced outcome rule $g = (g_0, g_1)\colon \Theta \to \Delta(X) \times \Delta(X)$, define the agent's associated utilities by

$$u\big(\theta'|\theta\big) = \alpha\big(\theta'|\theta\big)u\big(g_1\big(\theta'\big), \theta\big) + \big(1 - \alpha\big(\theta'|\theta\big)\big)u\big(g_0\big(\theta'\big), \theta\big), \quad \text{for all } \theta, \theta' \in \Theta.$$

The principal's problem is to choose a reduced outcome rule g to solve

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

The constraints capture truth-telling and obedience. They require that for each type θ, reporting θ and trying on test $\hat{t}(\theta)$ is weakly preferred to reporting any type θ' and either trying on test $\hat{t}(\theta')$ or intentionally failing it. In particular, with $\theta' = \theta$, the constraint ensures that type θ weakly prefers to try on test $\hat{t}(\theta)$ rather than intentionally failing it. In contrast to models of exogenous lying costs,[12] here the effective cost of misreporting

[12] In models of lying costs, reports have literal meanings. The agent pays a cost $c(\theta'|\theta)$ if he reports θ' when his true type is θ. See, for example, Lacker and Weinberg (1989), Maggi and Rodriguéz-Clare (1995), Crocker and Morgan (1998), Kartik, Ottaviani, and Squintani (2007), Kartik (2009), and Deneckere and Severinov (2022). Within mechanism design, Kephart and Conitzer (2016) show that if the lying cost function satisfies the triangle inequality, then there is no loss in restricting to truthful equilibria.

is determined jointly by the authentication rate and the principal's choice of decisions when the agent is not authenticated.

The approach described above motivates the following definition.

DEFINITION 3 (Most-Discerning Authentication Rate). An authentication rate α is *most-discerning* if there exists a testing technology (T, π) with a most-discerning testing function $\hat{t}$ such that

$$\alpha\big(\theta'|\theta\big) = \pi\big(\hat{t}\big(\theta'\big)|\theta\big), \quad \text{for all } \theta, \theta' \in \Theta. \tag{5}$$

If α is most discerning, then we can directly study the program (4), with the assurance that it represents the designer's full problem for some testing technology.

Now suppose that an *arbitrary* authentication rate α is specified directly; for examples of this approach, see Caragiannis et al. (2012) and Ferraioli and Ventre (2018). As long as communication is unrestricted, we believe the natural interpretation of a primitive authentication rate α is that for each report θ' there is an associated test, which each type θ can pass with probability $\alpha(\theta'|\theta)$. Formally, the principal has available the testing technology (T^α, π^α), defined by

$$T^\alpha = \big\{\tau^\alpha_{\theta'} : \theta' \in \Theta\big\}, \qquad \pi^\alpha\big(\tau^\alpha_{\theta'}|\theta\big) = \alpha\big(\theta'|\theta\big). \tag{6}$$

If this construction is applied to Example 2, then the testing function $\theta \mapsto \tau^\alpha_\theta$ is not most discerning.

REMARK 3 (Most Discerning). It is easily verified that an authentication rate α is most discerning if and only if, under the associated testing technology (T^α, π^α), the testing function $\theta \mapsto \tau^\alpha_\theta$ is most discerning.[13] Therefore, an authentication rate α is most discerning if and only if for all distinct types θ and θ', we have $\tau^\alpha_\theta \succeq_\theta \tau^\alpha_{\theta'}$, i.e., there exist $k_0 = k_0(\theta, \theta')$ and $k_1 = k_1(\theta, \theta')$ with $0 \le k_0 \le k_1 \le 1$ satisfying

$$\begin{aligned} \alpha(\theta|\theta)k_1 + \big(1 - \alpha(\theta|\theta)\big)k_0 &= \alpha\big(\theta'|\theta\big), \\ \alpha\big(\theta|\theta''\big)k_1 + \big(1 - \alpha\big(\theta|\theta''\big)\big)k_0 &\le \alpha\big(\theta'|\theta''\big), \quad \text{for all } \theta'' \in \Theta \setminus \{\theta\}. \end{aligned} \tag{7}$$

By Remark 3, checking whether an authentication rate α is most-discerning amounts to verifying whether a particular system of linear inequalities is feasible. We now give a simpler characterization of whether an authentication rate α is most-discerning, under one additional assumption on α.

PROPOSITION 4 (Most-Discerning Characterization). *Let α be an authentication rate satisfying $\alpha(\theta|\theta) \ge \max\{\alpha(\theta'|\theta), \alpha(\theta|\theta')\}$ for all $\theta, \theta' \in \Theta$. Then α is most discerning if and only if*

$$\alpha(\theta_3|\theta_2)\alpha(\theta_2|\theta_1) \le \alpha(\theta_3|\theta_1)\alpha(\theta_2|\theta_2), \quad \textit{for all } \theta_1, \theta_2, \theta_3 \in \Theta. \tag{8}$$

[13] If there exists a testing technology (T, π) with a most-discerning testing function $\hat{t}$ that satisfies (5), then for all distinct types θ and θ', we have $\tau^\alpha_\theta \sim \hat{t}(\theta) \succeq_\theta \hat{t}(\theta') \sim_\theta \tau^\alpha_{\theta'}$.

If $\alpha(\theta|\theta) = 1$ for all types θ, then it follows from Proposition 4 that α is most-discerning if and only if α is *supermultiplicative*: for all types θ_1, θ_2, θ_3, it is more likely that θ_1 is directly authenticated as type θ_3 than that type θ_1 is authenticated as type θ_2, and type θ_2 is (independently) authenticated as type θ_3.

To be sure, the principal's design problem cannot always be reduced to an optimization problem of the form (4). If the testing technology does not admit a most-discerning testing function, then the full protocol from Figure 1 must be considered.[14] Similarly, if an authentication rate α is not most discerning, then a solution of (4) need not be optimal among all dynamic mechanisms that use the associated testing technology (T^α, π^α). Still, there are many settings that can be reduced to the program in (4). We conclude this section with a few examples.

EXAMPLE 3 (Evidence Verification With Error). Suppose that each type θ has a distinct piece of evidence e_θ. The agent chooses whether to present the evidence he possesses. The principal has a system to check whether the presented evidence matches what was requested. A mismatch is detected with probability $1 - \varepsilon$, where $0 < \varepsilon < 1$.

For each type θ', let $\tau_{\theta'}$ denote the test that requests evidence $e_{\theta'}$. The agent passes this test if he presents a piece of evidence, and the system does not detect a mismatch between the presented evidence and $e_{\theta'}$. On this test, type θ can "try" by presenting evidence e_θ or "intentionally fail" by presenting no evidence. The associated passage rate is given by

$$\pi(\tau_{\theta'}|\theta) = \begin{cases} 1 & \text{if } \theta = \theta', \\ \varepsilon & \text{if } \theta \neq \theta'. \end{cases}$$

The map $\theta \mapsto \tau_\theta$ is most discerning; to see this, check the sufficient condition in Remark 2. The authentication rate representing this technology is analyzed in Dziuda and Salas (2018) and Balbuzanov (2019).

EXAMPLE 4 (Semimetric Authentication Rate). Let d be a semimetric on the type space Θ.[15] Consider the authentication rate α defined by

$$\alpha(\theta'|\theta) = \exp\{-d(\theta, \theta')\}.$$

The interpretation is that types that are closer in the semimetric d are more similar, and hence are more difficult to distinguish.[16] Using Proposition 4, it is easy to check that α is most discerning.

EXAMPLE 5 (Separate Verifiable and Payoff Components). Let $\Theta = \Theta^0 \times \Theta^1$, with a generic type denoted by $\theta = (\theta^0, \theta^1)$. Let α^0 be a most-discerning authentication rate

[14] In many cases, the class of testing rules can still be reduced; see Appendix A.6.

[15] Unlike a metric, a semimetric does not require that $d(\theta, \theta') \neq 0$ for $\theta \neq \theta'$.

[16] As a special case, suppose $\Theta = \mathbf{R}^k$ and d is induced by a norm. The resulting class of functions α is axiomatized in Billot, Gilboa, and Schmeidler (2008). They interpret these functions as measures of similarity in their model of belief-formation through similarity-weighted averaging.

on Θ^0. We obtain a most-discerning authentication rate α on Θ by defining

$$\alpha\big(\hat{\theta}^0, \hat{\theta}^1 | \theta^0, \theta^1\big) = \alpha^0\big(\hat{\theta}^0 | \theta^0\big),$$

for all $\theta^0, \hat{\theta}^0 \in \Theta^0$ and all $\theta^1, \hat{\theta}^1 \in \Theta^1$. Think of θ^0 as a verifiable attribute, and θ^1 as an unverifiable payoff type. Reuter (2023) considers this structure in a model of partial verification. The prior distribution over $\Theta^0 \times \Theta^1$ determines how informative the agent's verifiable attribute is about his payoff type. For example, the verifiable attribute could indicate whether the agent is a student, and the valuation distribution among students may be different from the valuation distribution among nonstudents.

Example 6 (Partial Verification). Suppose that the authentication rate α is $\{0, 1\}$-valued. For each type θ, let $M(\theta) = \{\theta' \in \Theta : \alpha(\theta'|\theta) = 1\}$. Hence, type θ can be authenticated as any type in $M(\theta)$. Following Green and Laffont (1986), suppose that each type can be authenticated as himself, i.e., $\alpha(\theta|\theta) = 1$ for all θ. In terms of M, (8) becomes

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

This is exactly Green and Laffont's (1986) nested range condition. Under this condition, they argue that the revelation principle is valid.

6. Applications: Profit-maximization with verification

As an application of the reduced-form representation of the principal's design problem, we solve for profit-maximizing mechanisms with verification in a few classical settings.

6.1 *Setting*

Authentication rate The type space is an interval $\Theta = [\underline{\theta}, \bar{\theta}]$, where $\bar{\theta} > \underline{\theta} \geq 0$. We represent the verification technology by a most-discerning authentication rate α. Assume that α takes the exponential form

$$\alpha\big(\theta'|\theta\big) = \exp\left(-\left|\int_{\theta'}^{\theta} \lambda(\xi)\, \mathrm{d}\xi\right|\right), \quad \text{for all } \theta, \theta' \in \Theta, \tag{9}$$

for some continuous function $\lambda\colon [\underline{\theta}, \bar{\theta}] \to \mathbf{R}_+$. It is easily verified that this function α is most discerning. The exponential authentication rate allows for a cleaner characterization of optimal mechanisms. With other most-discerning authentication rates, additional regularity conditions are needed to ensure that global deviations are not profitable; see Appendix A.11.

The parameter $\lambda(\theta)$ 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. The plot shows the authentication probability, as a function of the agent's true type, for two fixed reports θ' and θ''.

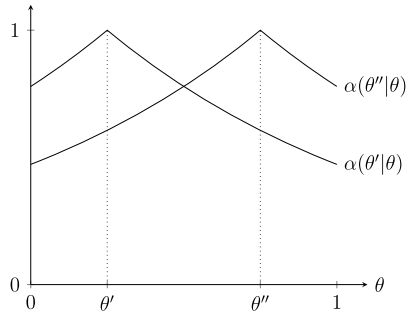

FIGURE 3. Exponential authentication rate.

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 $t \in \mathbf{R}$.[17] The set Q will be either $[0, 1]$ or $\mathbf{R}_+$, depending on the application. Utilities for the agent and the principal are given by

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

for some weakly convex cost function $c\colon Q \to \mathbf{R}_+$. For concreteness, we interpret the principal as the seller of a good and the agent as a potential buyer.[18] The agent's type represents his market segment, which the principal can imperfectly verify. To avoid the difficulties of mechanism design with multi-dimensional types, we make the stylized assumption that the agent's market segment pins down his valuation. With the specified authentication rate, it is more difficult for the seller to distinguish buyers who are in market segments with closer valuations.

The agent is free to walk away at any time, so we impose an ex post participation constraint.[19] If the principal could impose arbitrarily severe punishments for failed authentication, then probabilistic verification would be essentially as effective as perfect verification; see Caragiannis et al. (2012).

Since $\alpha(\theta|\theta) = 1$ for all θ, the agent is always authenticated if he is truthful. Therefore, failed authentication is off path. Given the ex post participation constraint, we may assume without loss that if the agent fails to be authenticated, then the principal excludes him—the agent pays nothing and does not get the good. Formally, we set $g_0(\theta) = (0, 0)$ for all θ, and we optimize over the decision rule g_1. Denote the quantity and transfer components of g_1 by q and t.

[17]The pair (q, t) corresponds to the decision x in the general model. Throughout Section 6, t always denotes transfers (and we make no direct reference to tests).

[18]An alternative interpretation of this setting is that the principal is the procurer of a good who can imperfectly verify the agent's production costs.

[19]Formally, after the agent observes the test result, he has the right to walk away, free and clear, with no payment obligation. This assumption rules out upfront payments like those used in Border and Sobel (1987).

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

$$\begin{aligned} \text{maximize} \quad & \int_{\underline{\theta}}^{\bar{\theta}} \big[t(\theta) - c\big(q(\theta)\big)\big] f(\theta)\,\mathrm{d}\theta \\ \text{subject to} \quad & \theta q(\theta) - t(\theta) \geq \alpha\big(\theta'|\theta\big)\big[\theta q\big(\theta'\big) - t\big(\theta'\big)\big], \quad \text{for all } \theta,\, \theta' \in \Theta \\ & \theta q(\theta) - t(\theta) \geq 0, \quad \text{for all } \theta \in \Theta. \end{aligned} \tag{10}$$

Here, the constraints from (4) take a simple form because $u(g_0(\theta'), \theta) = 0$ for all θ and θ'. These constraints guarantee ex post participation.[20]

6.1.1 *Virtual value* We derive a new expression for the virtual value in this quasilinear setting with verification. In the classical setting without verification, the envelope theorem pins down (almost everywhere) the derivative of the agent's indirect utility function U in terms of the allocation rule: $U'(\theta) = q(\theta)$. Hence,

$$U(\theta) = U(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} q(\xi)\,\mathrm{d}\xi \geq \int_{\underline{\theta}}^{\theta} q(\xi)\,\mathrm{d}\xi. \tag{11}$$

With verification, the derivative of the agent's indirect utility function U is no longer pinned down by the quantity function because of the kink in $\alpha(\theta|\cdot)$. Instead, the envelope formula gives the differential inequality[21]

$$q(\theta) - \lambda(\theta)U(\theta) \leq U'(\theta) \leq q(\theta) + \lambda(\theta)U(\theta). \tag{12}$$

This differential inequality depends only on the *local* behavior of α around the diagonal, which is captured by the function λ. Indeed, the left and right derivatives of the function $\alpha(\theta|\cdot)$, evaluated at θ, equal $\lambda(\theta)$ and $-\lambda(\theta)$, respectively. The greater the local verification precision $\lambda(\theta)$, the more permissive is the inequality in (12).

The lower bound in (12) can be shown to imply the bound

$$U(\theta) \geq \int_{\underline{\theta}}^{\theta} e^{-\int_{\xi}^{\theta} \lambda(z)\,\mathrm{d}z} q(\xi)\,\mathrm{d}\xi. \tag{13}$$

The right-hand side of (13) solves the differential equation $U'(\theta) = q(\theta) - \lambda(\theta)U(\theta)$. Since α takes the exponential form in (9), the integrand in (13) reduces to $\alpha(\xi|\theta)q(\xi)$. We will use this simpler expression below, but remember that the solution is pinned down by the envelope formula, not by global deviations.

It is optimal to choose U so that (13) holds with equality. After substituting this choice of 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

[20] Since $g_0(\theta') = (0, 0)$ for all θ', the agent gets his outside option whenever he fails to be authenticated. So in this case, the constraints in (4) imply the ex post participation constraints.

[21] See Carbajal and Ely (2013) for a general characterization of indirect utility functions, when the agent's primitive utility function is kinked. Carbajal and Ely (2016) apply this characterization in a model of reference-dependent utility.

value of type θ:

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

Myerson's virtual value is derived similarly in the no-verification problem, using (11) in place of (13). Myerson's virtual value can be expressed symmetrically as

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

The virtual value of type θ captures the marginal revenue from allocating to type θ. It has two parts. First, the principal can extract the consumption utility θ from type θ. Second, the allocation pushes up the indirect utility of each type ξ with $\xi > \theta$. This marginal effect on type ξ, which equals 1 (without verification) and $\alpha(\theta|\xi)$ (with verification), 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 toward these bounds in the limiting cases. As λ converges to 0 pointwise, $\varphi(\theta)$ converges to $\varphi^M(\theta)$ for each type θ. Conversely, as λ converges to ∞ pointwise, $\varphi(\theta)$ converges to θ for each type θ.

Below, we will characterize optimal mechanisms under the assumption that the virtual value φ is increasing. This is a joint assumption on the type distribution and the authentication rate. If $\lambda(\theta) = \lambda$ for all θ, then the virtual value has a simple expression for some standard distributions. In particular, for both uniform and exponential distributions, the virtual value is strictly increasing.

6.2 *Optimal mechanisms*

We find the optimal mechanism in two classical problems.

6.2.1 *Nonlinear pricing* For nonlinear pricing (Mussa and Rosen (1978)), the quantity space is $Q = \mathbf{R}_+$. Assume that the principal's cost function c satisfies the standard assumptions: $c'(0) = 0$, the derivative c' is strictly increasing, and $\lim_{q\to\infty} c'(q) > \bar{\theta}$. Say that the optimal mechanism is *essentially unique* if all optimal mechanisms agree at almost every type.

Proposition 5 (Optimal Nonlinear Pricing). *Assume that the virtual value φ is weakly increasing. The optimal quantity function q^* and transfer function t^* are essentially unique and given by*

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

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 (13).

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

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

Without verification, the profit-maximizing mechanism is a posted price. With verification, the seller charges different prices to consumers in different market segments.

Proposition 6 (Optimal Sale of a Single Good). *Assume that the virtual value φ is strictly increasing. The optimal quantity function q^* and transfer function t^* are essentially unique and given as follows. Let $\theta^* = \varphi^{-1}(c)$. If $\theta < \theta^*$, then $q^*(\theta) = t^*(\theta) = 0$. If $\theta \geq \theta^*$, then $q^*(\theta) = 1$ and*

$$t^*(\theta) = \theta^* + \int_{\theta^*}^{\theta} \big(1 - \alpha(\xi|\theta)\big)\, \mathrm{d}\xi.$$

As in the classical solution, the allocation probability takes values 0 and 1 only—there is no randomization.[22] 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 a price that is less than his valuation. 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 of failing to be authenticated and getting nothing. As verification becomes more precise, the price becomes more sensitive to the agent's type, and more types receive the good.

Remark 4 (Auctions). Our model can be extended to allow for multiple agents.[23] In the revenue-maximizing auction, the allocation rule takes the familiar form from Myerson (1981), with our generalized virtual value in place of the classical virtual value. In the asymmetric case, the allocation rule favors a bidder if his valuation distribution is lower or if his valuation can be verified more precisely.

7. Beyond pass–fail tests

The main model considers pass–fail tests. In this section, we consider tests that generate scores in a finite score set S. The agent's type-dependent performance on each test is

[22] Sher and Vohra (2015) study this selling problem with deterministic evidence, assuming the type space is finite. The optimal mechanism may involve lotteries. They give a condition on the evidence structure under which the optimal mechanism is deterministic.

[23] In this extension, we assume that the principal tests the agents simultaneously. In particular, the test given to one agent cannot depend on another agent's test score.

represented by a map $\pi\colon T \times \Theta \to \Delta(S)$, which specifies for each type θ and test τ a distribution $\pi_{\tau|\theta}$ over S.

To generalize the agent's choice of whether to try on a test, we take as primitive a partial order $\succeq$ on S. The interpretation is that the agent can shift probability from score s to score s' if and only if $s \succeq s'$. As before, the agent's choice is costless, and the principal observes only the final score. The main model of pass–fail testing corresponds to $S = \{0, 1\}$ with the usual order $\succeq$; by mixing, type θ can choose to pass test τ with any probability below $\pi(\tau|\theta)$. In the general model, 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 probability measures on the partially ordered space $(S, \succeq)$. That is, $\mu \succeq_{\text{st}} \nu$ if and only if $\mu(U) \geq \nu(U)$ for every upper set U;[24] see Kamae, Krengel, and O'Brien (1977).

We define the θ-discernment orders in this more general 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 $k\colon S \to \Delta(S)$, the measure μk on $\Delta(S)$ is defined by $(\mu k)(A) = \sum_s \mu(s)k(A|s)$, for $A \subset S$.

Definition 4 (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 $\theta' \neq \theta$.

This order $\succeq_\theta$ is reflexive and transitive; see Appendix A.12 for a proof. We can define *most θ-discerning tests* and *most-discerning testing functions* with respect to this definition of $\succeq_\theta$. With this generalized testing technology, the revelation principle (Proposition 1), the replacement theorem (Theorem 1), and the forward implication in the main implementation theorem (Theorem 2) go through with similar proofs.[25]

8. Related literature on verification

Verification has been modeled in many ways, in both economics and computer science. Here, we focus on costless, imperfect verification.[26]

Green and Laffont (1986) introduce *partial verification*.[27] They restrict their analysis to direct mechanisms. Verification is represented as a correspondence $M\colon \Theta \twoheadrightarrow \Theta$

[24] 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.

[25] A most-discerning testing function $\hat{t}$ induces a generalized authentication rate $\alpha\colon \Theta \times \Theta \to \Delta(S)$ defined by $\alpha(\theta'|\theta) = \pi_{\hat{t}(\theta')|\theta}$. We can set up an analogue of the program in (4), but the incentive constraints depend on the order $\succeq$ on S.

[26] 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 studies *costly verification* in debt contracts. Ben-Porath, Dekel, and Lipman (2019) connect costly verification and evidence. When monetary transfers are infeasible, costly verification is often used as a substitute; see Ben-Porath, Dekel, and Lipman (2014), Mylovanov and Zapechelnyuk (2017), Erlanson and Kleiner (2020), Halac and Yared (2020), and Li (2020).

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

satisfying $\theta \in M(\theta)$ for each type θ. Each type θ can "report" any type θ' in $M(\theta)$. This correspondence M can be represented within our model as a $\{0, 1\}$-valued authentication rate. We reinterpret the apparent failure of the revelation principle in Green and Laffont's (1986) model as a consequence of taking as primitive an authentication rate that is not most-discerning; see Example 6 for the formal connection.

Bull and Watson (2004, 2007), Deneckere and Severinov (2008), and Lipman and Seppi (1995) study *hard evidence*.[28] They introduce an abstract evidence set $\mathcal{E}$ and an evidence correspondence $E\colon \Theta \twoheadrightarrow \mathcal{E}$. Type θ possesses the evidence in $E(\theta)$, and he can present one piece of evidence from $E(\theta)$. Presenting evidence is costless.[29] Bull and Watson (2007) show that this evidence model can be represented in a reduced form if the evidence environment is *normal*, i.e., each type θ has a piece of evidence $e(\theta)$ in $E(\theta)$ that is *maximal* for type θ in the following sense: every other type θ' who has $e(\theta)$ also has every other piece of evidence in $E(\theta)$. This model of deterministic evidence can be represented within our model as follows. For each piece of evidence e in $\mathcal{E}$, define the test τ_e that requests evidence e. Type θ can pass test τ_e if and only if e is in $E(\theta)$. Every type can intentionally fail any test by withholding his evidence.[30] A piece of evidence e in $E(\theta)$ is maximal for type θ in the sense of Bull and Watson (2007) if and only if test τ_e is most-θ discerning in our sense.[31]

In computer science, Caragiannis et al. (2012) and Ferraioli and Ventre (2018) consider a primitive authentication rate, and they restrict attention to truthful equilibria of direct mechanisms. Our paper shows that the restriction to direct, truthful mechanisms is without loss if α is most-discerning. Caragiannis et al. (2012) allow the principal to use arbitrarily severe punishments to deter any report that is not authenticated with certainty. In our applications (Section 6), the agent can walk away at any time, so punishment is limited to the agent's outside option.

Closest to our paper is the independent paper of Ben-Porath, Dekel, and Lipman (2023). They consider an abstract evidence set $\mathcal{E}$. In their *signal-choice model*,[32] the primitive is a correspondence $A\colon \Theta \twoheadrightarrow \Delta(\mathcal{E})$. Type θ can choose any distribution a in $A(\theta)$. Then evidence e in $\mathcal{E}$ is realized according to the distribution a. For each type θ, they define an associated informativeness order over $A(\theta)$, which depends on the full correspondence A. In the spirit of our Theorem 1, they show that for implementation,

[28] Evidence was introduced in games (without commitment) by Milgrom (1981) and Grossman (1981); for more recent work on evidence games, see Hart, Kremer, and Perry (2017), Ben-Porath, Dekel, and Lipman (2017), and Koessler and Perez-Richet (2019).

[29] In Kartik and Tercieux (2012), the agent can provide evidence at a cost, which depends on the state. The focus of their paper is full implementation.

[30] Formally, in Bull and Watson (2007), the agent *must* present a piece of evidence from $E(\theta)$. Disclosing nothing can be represented in their framework as a distinguished piece of evidence that every type possesses. If there exists such "minimal evidence" (Bull and Watson (2007, p. 85)), then their evidence model is equivalent to our testing representation.

[31] Strausz and Schweighofer-Kodritsch (2023) revisit the setting of Bull and Watson (2007). They consider both the standard specification (termed "noncontrollable evidence") and an alternative specification ("controllable evidence") in which contracts can be written demanding particular evidence provision. They give conditions under which controllability has no value for the principal.

[32] They also consider a more general *evidence-acquisition model*, and they give conditions under which a given evidence-acquisition model can be represented as a signal-choice model.

each type's equilibrium choice of a less informative distribution can be replaced with a more informative distribution. It is possible to embed our testing protocol in their signal-choice model.[33] Our papers have different aims. Ben-Porath, Dekel, and Lipman (2023) study the relationship between different evidence protocols in the most general setting. We impose more structure to obtain a tractable verification framework that we can apply to classical mechanism design problems.

9. Conclusion

We model probabilistic verification as a technology—a family of tests that are available to the principal. The principal chooses how to use this testing technology within an arbitrary dynamic mechanism. We characterize whether this complex problem can be reduced to a static problem with relaxed incentive constraints. Then we solve this reduced problem using the first-order approach in a few classical profit-maximization applications. We believe this first-order approach will be useful for solving other mechanism design problems with probabilistic verification.

We have found the optimal mechanism for each fixed verification technology in a nonparametric family. We can therefore quantify the value of each technology to the principal. This is the first step toward analyzing a richer setting in which the principal chooses how much to invest in verification technologies. We leave this to future research.

Appendix: Proofs

A.1 *Proof of Proposition 1*

Let $S = \{0, 1\}$. Consider a mechanism $(M, M'; t, r', g)$ and a strategy (r, a). For each fixed type θ, the sequence (m, τ, m', s', x) in $M \times T \times M' \times S \times X$ is realized according to the following procedure. (Below, the symbol $\sim$ denotes "distributed according to.")

- Agent sends $m \sim r(\theta)$.
- Principal selects $\tau \sim t(m)$.
- Principal sends $m' \sim r'(m, \tau)$.
- Agent tries with probability $a(\theta, m, \tau, m')$.
- Nature draws s' according to $\pi(\tau|\theta)$ and whether the agent tried.
- Principal selects $x \sim g(m, \tau, m', s')$.

This distribution of (m, τ, m', s', x) is replicated by the following canonical procedure:

- Agent sends $\theta' = \theta$.

[33] Given a nonbinary testing technology (T, S, π), consider their signal-choice model with $\mathcal{E} = T \times S$. For each type θ, let $A(\theta)$ be the set of distributions $\delta_\tau \otimes p$ for all $\tau \in T$ and $p \in \Delta(S)$ satisfying $\pi_{\tau|\theta} \succeq_{\text{st}} p$. Under this embedding, τ is more θ-discerning than ψ in our framework if and only if $\delta_\tau \otimes \pi_{\tau|\theta}$ is more informative (for type θ) than $\delta_\tau \otimes \pi_{\psi|\theta}$ in theirs.

- Principal privately draws $m \sim r(\theta')$ and then selects $\tau \sim t(m)$.
- Agent tries.
- Nature draws s according to $\pi(\tau|\theta)$ and whether the agent tried.
- Principal privately draws $m' \sim r'(m, \tau)$; then privately draws $s' \in \{0, 1\}$ so that $s' = 1$ with probability $a(\theta', m, \tau, m')s$; and finally selects $x \sim g(m, \tau, m', s')$.

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) privately drawing $\theta' \sim \rho$ and then sending $m \sim r(\theta')$; and (ii) trying with probability $a(\theta', m, \tau, m')\alpha(\theta', \tau)$.[34]

A.2 *Insufficiency of mixed strategies with exogenous authentication rate*

Consider a modification of Example 2. The type space and the verification technology are as in Figure 2. Now there are three allocations—nothing, low-quality, and high-quality—with associated type-independent utilities 0, u_ℓ, and u_h. Assume $0 < u_\ell < u_h$ and $u_\ell \geq u_h/2$. Consider the social choice function that allocates the high-quality good to type θ_1 and the low-quality good to types θ_2 and θ_3.

We claim that this social choice function cannot be implemented in the reduced-form model, even if the agent uses a mixed strategy. Type θ_1 can pass only as type θ_1 or as type θ_2. So for some $\theta \in \{\theta_1, \theta_2\}$, the principal must give the high-quality good to the agent if he passes as type θ. But type θ_3 can pass as type θ_1, and type θ_2 can pass as type θ_2, so at least one of the types θ_3 and θ_2 has a strictly profitable deviation.

Now replace the authentication rate with a testing technology consisting of three tests, denoted τ_1, τ_2, τ_3. Test τ_i can be passed by those types that can pass as θ_i in the reduced-form model. In this testing model, the principal can implement the specified social choice function. If the agent reports type θ_2 or θ_3, he is given the low-quality good. If the agent reports type θ_1, then the principal gives either test τ_1 or test τ_2, each with probability $1/2$. Whichever test is given, the agent gets the high-quality good if he passes and nothing if he fails. If either type θ_2 or type θ_3 deviates by reporting θ_1, then he gets the high-quality good with probability at most $1/2$, and otherwise he gets nothing. This deviation is unprofitable since $u_\ell \geq u_h/2$.

[34]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 and uses it to select x. In the replicating deviation in the original mechanism, the agent remembers his privately drawn θ' before choosing whether to try. We can replace this memory with fresh draws from the correct conditional distributions. The principal redraws m conditional on (θ', τ). The agent redraws θ' conditional on (θ, m). To construct these conditional distributions, apply disintegration of measures (Kallenberg (2017, Theorem 1.25, p. 39)). This result applies to Borel probability measures, so we first restrict our measures to the Borel σ-algebra, then apply the theorem, and finally extend the resulting measures to the universal completion.

A.3 *Proof of Theorem 1*

Let f be a social choice function that is canonically implemented by a mechanism (t, g) in which $t(\theta) = \psi$. Define the mechanism (t', g') to coincide with (t, g) except for the following modifications. Set $t'(\theta) = \tau$. Choose k_1 and k_0 from the definition of $\tau \succeq_\theta \psi$. For each $s = 0, 1$, set

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

Under the mechanism (t', g'), if type θ' reports type θ and tries on test τ with probability a, the resulting decision will be

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

where

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

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, (t', g') replicates the social choice function f without introducing any new deviation outcomes for any type.

A.4 *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 we can set $(k_0, k_1) = (0, 1)$ to see that $\tau_1 \succeq_\theta \tau_2$ and $\tau_2 \succeq_\theta \tau_1$. If θ is minimal on test τ_1, then we see that $\tau_2 \succeq_\theta \tau_1$ by setting $k_0 = k_1 = \pi(\tau_1|\theta)$. Symmetrically, if θ is minimal on τ_2, then we see that $\tau_1 \succeq_\theta \tau_2$ 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 $\pi(\tau_1|\theta) > \pi(\tau_1|\theta')$. We prove that $\pi(\tau_1|\cdot) = \pi(\tau_2|\cdot)$.

We use Markov transition notation; see Section 7. Let k and k' denote the Markov transitions associated with (k_0, k_1) and (k_0', k_1'), respectively. Let $\pi_{\tau|\theta}$ denote the probability measure that puts probability $\pi(\tau|\theta)$ on $s = 1$. We have

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

In terms of the probability on $s = 1$, we can express this system as

$$k_0 k_1' + (1 - k_0) k_0' + \pi(\tau_1|\theta)(k_1 - k_0)\big(k_1' - k_0'\big) = \pi(\tau_1|\theta),$$
$$k_0 k_1' + (1 - k_0) k_0' + \pi\big(\tau_1|\theta'\big)(k_1 - k_0)\big(k_1' - k_0'\big) \leq \pi\big(\tau_1|\theta'\big).$$

After subtracting, we conclude that

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

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.5 *Proof of Theorem 2*

(i) $\implies$ (ii). Let $\hat{t}$ be a most-discerning testing function. For each type θ and test ψ, select probabilities $k_0(\theta, \psi)$ and $k_1(\theta, \psi)$ satisfying the definition of $\hat{t}(\theta) \succeq_\theta \psi$; Appendix A.14 shows that there exists a *measurable* selection. Fix a decision environment (X, u). Let f be an implementable social choice function. By the revelation principle (Proposition 1), f is canonically implemented by some mechanism (t, g). Consider a mechanism $(\hat{t}, \hat{g})$ in which $\hat{g}$ satisfies

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

for all types θ and scores $s = 0, 1$.[35]

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

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

where

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

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, $(\hat{t}, \hat{g})$ replicates the social choice function f without introducing any new deviation outcomes for any type.[36]

(ii) $\implies$ (i). 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 y. Every type gets utility 1 from decision $\bar{x}$ and utility 0 from decision $\underline{x}$. Each type θ' gets utility $\pi(\psi|\theta')$ from decision y.

Consider the following mechanism. If the agent reports θ' with $\theta' \neq \theta$, the principal selects y (the test and score do not matter). If the agent reports θ, the principal gives test ψ and then selects $\bar{x}$ if the agent passes and $\underline{x}$ if the agent fails. Observe that truth-telling and trying is a best response for every type. Denote the induced social choice function by f.

By (ii), f can be canonically implemented by $(\hat{t}, \hat{g})$, for some outcome rule $\hat{g}$. For $s = 0, 1$, let k_s be the probability that $\hat{g}(\theta, \hat{t}(\theta), s)$ assigns to $\bar{x}$. We must have $k_1 \geq k_0$; otherwise, type θ could profitably deviate by intentionally failing test $\hat{t}(\theta)$.[37] Since this mechanism implements f, the probabilities k_0 and k_1 satisfy (i) in the definition of $\hat{t}(\theta) \succeq_\theta \psi$. Since no type θ' can profit from reporting θ and trying on test $\hat{t}(\theta)$, we get (ii). Therefore, $\hat{t}(\theta) \succeq_\theta \psi$, as desired.

[35] In a slight abuse of notation, t maps Θ into $\Delta(T)$, while $\hat{t}$ maps Θ into T.

[36] Our argument is similar in spirit to de Oliveira's (2018) elegant proof of Blackwell's theorem using diagrams.

[37] If $\pi(\hat{t}(\theta)|\theta) > 0$, this holds because $\hat{g}(\theta, \hat{t}(\theta), 1)$ must concentrate on $\{\underline{x}, \bar{x}\}$, and type θ weakly prefers y to $\underline{x}$. If $\pi(\hat{t}(\theta)|\theta) = 0$, then we may assume $k_1 = k_0$ since implementation is preserved by redefining $\hat{g}(\theta, \hat{t}(\theta), 1)$ to equal $\hat{g}(\theta, \hat{t}(\theta), 0)$.

A.6 *Most-discerning correspondences*

Even if the testing technology does not admit a most-discerning testing function, we can still use the discernment orders to reduce the class of tests that need to be considered.

DEFINITION 5 (Most-Discerning Correspondence). A subset T_0 of T is *most θ-discerning* if for each test ψ in T there exists a test τ 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 type θ. The next result says that if a correspondence is most-discerning, then we can restrict attention to testing rules supported on that correspondence.

THEOREM 3 (Implementation With a Most-Discerning Correspondence). *Suppose that the passage rate π is continuous. Let $\hat{T}$ be a weakly measurable*[38] *correspondence from Θ to T with closed values. 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 essentially the same as the proof of Theorem 2. For each type θ and test ψ, there exists a test τ in $\hat{T}(\theta)$ such that $\tau \succeq_\theta \psi$. But we must check that there exists such a selection that is measurable; see Appendix A.14. The regularity conditions on π and $\hat{T}$ ensure that a measurable selection exists. If we can independently construct a measurable selection, then these conditions are not needed.

A.7 *Proof of Proposition 3*

There are two cases:

(i) Suppose $\pi(\tau|\theta) \geq \pi(\psi|\theta) > 0$. If (1) holds, then Definition 1 is satisfied with

$$k_0 = 0 \quad \text{and} \quad k_1 = \frac{\pi(\psi|\theta)}{\pi(\tau|\theta)}.$$

(ii) Suppose $\pi(\tau|\theta) \leq \pi(\psi|\theta) < 1$. If (2) holds, then Definition 1 is satisfied with

$$k_0 = \frac{\pi(\psi|\theta) - \pi(\tau|\theta)}{1 - \pi(\tau|\theta)} \quad \text{and} \quad k_1 = 1.$$

To see this, multiply each side of (2) by $1 - \pi(\psi|\theta)$. Subtract each side of the resulting inequality from 1 (and flip the direction of the inequality).

[38] That is, the lower inverse $\{\theta \in \Theta : T(\theta) \cap G \neq \emptyset\}$ is universally measurable for each open subset G of T; see (Aliprantis and Border (2006, p. 592)).

A.8 *Proof of Proposition 4*

Let α be an authentication rate satisfying $\alpha(\theta|\theta) \geq \max\{\alpha(\theta'|\theta), \alpha(\theta|\theta')\}$ for all $\theta, \theta' \in \Theta$. First, observe that (8) is trivially satisfied if $\theta_1 = \theta_2$ or $\theta_2 = \theta_3$. Fix distinct types θ_2 and θ_3. By Remark 3, it suffices to show that $\tau^\alpha_{\theta_2} \succeq_{\theta_2} \tau^\alpha_{\theta_3}$ if and only if

$$\alpha(\theta_3|\theta_2)\alpha(\theta_2|\theta_1) \leq \alpha(\theta_3|\theta_1)\alpha(\theta_2|\theta_2), \quad \text{for all } \theta_1 \in \Theta \setminus \{\theta_2\}. \tag{14}$$

There are two cases:

(i) Suppose $\alpha(\theta_2|\theta_2) = 0$. It follows from the assumption on α that $\alpha(\theta_3|\theta_2) = 0$. Thus, (14) is satisfied (because both sides are zero). Also, $\tau^\alpha_{\theta_2} \succeq_{\theta_2} \tau^\alpha_{\theta_3}$ because the system (7), with $\theta = \theta_2$ and $\theta' = \theta_3$, is solved by $k_0 = 0$ and $k_1 = 1$ (by the assumption on α).

(ii) Suppose $\alpha(\theta_2|\theta_2) > 0$. If (14) holds, then $\tau^\alpha_{\theta_2} \succeq_{\theta_2} \tau^\alpha_{\theta_3}$ because the system (7), with $\theta = \theta_2$ and $\theta' = \theta_3$, is solved by $k_0 = 0$ and $k_1 = \alpha(\theta_3|\theta_2)/\alpha(\theta_2|\theta_2)$; note that $\alpha(\theta_3|\theta_2)/\alpha(\theta_2|\theta_2) \leq 1$ by the assumption on α. Conversely, if $\tau^\alpha_{\theta_2} \succeq_{\theta_2} \tau^\alpha_{\theta_3}$, then the system (7), with $\theta = \theta_2$ and $\theta' = \theta_3$, has a nonnegative solution (k_0, k_1). We claim that this system is also solved by

$$k_0' = 0 \quad \text{and} \quad k_1' = k_1 + \frac{1 - \alpha(\theta_2|\theta_2)}{\alpha(\theta_2|\theta_2)} k_0.$$

To see this, note that this modification leaves the equality in (7) unchanged and changes the left-hand side of the θ''-inequality by

$$\left[\big(1 - \alpha(\theta_2|\theta_2)\big)\frac{\alpha\big(\theta_2|\theta''\big)}{\alpha(\theta_2|\theta_2)} - \big(1 - \alpha\big(\theta_2|\theta''\big)\big)\right] k_0,$$

which is nonpositive because $\alpha(\theta_2|\theta_2) \geq \alpha(\theta_2|\theta'')$, by the assumption on α. Now examine the new solution (k_0', k_1') of the system (7), with $\theta = \theta_2$ and $\theta' = \theta_3$. Since $k_0' = 0$, the equality gives $k_1' = \alpha(\theta_3|\theta_2)/\alpha(\theta_2|\theta_2)$. In each inequality, scale each side by $\alpha(\theta_2|\theta_2)$ to get (14), as desired.

A.9 *Proof of Proposition 5*

The following lemma is established at the end of the proof.

Lemma 1 (Bounded Mechanisms). *Let (q, t) be an incentive compatible mechanism. There exists a bounded, incentive compatible mechanism $(\bar{q}, \bar{t})$ such that either (i) $(\bar{q}, \bar{t})$ and (q, t) agree almost surely, or (ii) the principal strictly prefers $(\bar{q}, \bar{t})$ to (q, t).*

By Lemma 1, it suffices to prove that (q^*, t^*) is the essentially unique optimum among all bounded, incentive compatible mechanisms (q, t). By setting $U(\theta) = \theta q(\theta) - t(\theta)$, we can equivalently specify a mechanism (q, t) as a quantity-utility pair (q, U). Note that (q, t) is bounded if and only if (q, U) is.

LEMMA 2 (Envelope Theorem Bound). *Let (q, U) be a quantity-utility pair. If (q, U) is bounded and incentive compatible, then for each type θ, we have*

$$U(\theta) \geq \int_{\underline{\theta}}^{\theta} \alpha(\xi|\theta) q(\xi)\, \mathrm{d}\xi. \tag{15}$$

Lemma 2 is established at the end of the proof. We now prove Proposition 5, assuming Lemmas 1 and 2. Let (q, U) be a bounded, incentive compatible quantity–utility pair. We can bound the principal's objective by applying Lemma 2 and then switching the order of integration:

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

with equality if and only if (15) holds with equality for almost every type θ. For each type θ, the integrand in brackets on the right-hand side is uniquely maximized by $q^*(\theta)$. The transfer function t^* ensures that U satisfies (15) with equality for every type θ.

To complete the proof, we check that (q^*, t^*) satisfies global incentive compatibility if the quantity function q^* satisfies the following monotonicity condition: Whenever $\underline{\theta} \leq \xi_1 \leq \xi_2 \leq \theta$, we have

$$\alpha(\xi_1|\theta) q^*(\xi_1) \leq \alpha(\xi_2|\theta) q^*(\xi_2). \tag{16}$$

This monotonicity condition holds because q^* is weakly increasing (since φ is weakly increasing).

The global incentive constraints require that for all types θ and θ', we have

$$U(\theta) \geq \alpha\big(\theta'|\theta\big)\big[U\big(\theta'\big) + \big(\theta - \theta'\big) q^*\big(\theta'\big)\big],$$

or equivalently,

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

Plug in the right-hand side of (15) for U to get the condition

$$\int_{\underline{\theta}}^{\theta} \alpha(\xi|\theta) q^*(\xi)\, \mathrm{d}\xi - \int_{\underline{\theta}}^{\theta'} \alpha\big(\xi|\theta'\big) \alpha\big(\theta'|\theta\big) q^*(\xi)\, \mathrm{d}\xi \geq \big(\theta - \theta'\big) \alpha\big(\theta'|\theta\big) q^*\big(\theta'\big). \tag{18}$$

We separate into cases. If $\theta > \theta'$, then (18) is equivalent to

$$\int_{\theta'}^{\theta} \alpha(\xi|\theta) q^*(\xi) \geq \big(\theta - \theta'\big) \alpha\big(\theta'|\theta\big) q^*\big(\theta'\big).$$

If $\theta < \theta'$, then (18) holds if

$$\int_{\theta}^{\theta'} \alpha\big(\xi|\theta'\big) q^*(\xi) \leq \big(\theta' - \theta\big) q^*\big(\theta'\big).$$

In each case, the inequality is guaranteed by the monotonicity condition in (16).

Proof of Lemma 1. Let (q, t) be an incentive compatible mechanism. Let

$$\Theta_0 = \{\theta \in \Theta : t(\theta) - c(q(\theta)) \geq 0\}.$$

Since $\lim_{q\to\infty} c'(q) > \bar{\theta}$, we may choose L such that $\bar{\theta}q - c(q) < 0$ for all $q > L$. For all $\theta \in \Theta_0$, it follows from the participation constraint that $\theta q(\theta) - c(q(\theta)) \geq 0$, so $q(\theta) \leq L$, and hence $0 \leq t(\theta) \leq \bar{\theta}L$. For each type θ, let $\varphi(\theta)$ be the closure of the bounded set

$$\{(\alpha(\theta'|\theta)q(\theta'), \alpha(\theta'|\theta)t(\theta')) : \theta' \in \Theta_0\}.$$

By the measurable maximum theorem (Aliprantis and Border (2006, 18.19, p. 605)), the correspondence

$$\theta \mapsto \operatorname*{argmax}_{(q',t')\in\varphi(\theta)} (\theta q' - t')$$

admits a measurable selection $(\tilde{q}, \tilde{t})\colon \Theta \to [0, L] \times [0, \bar{\theta}L]$. Define $(\bar{q}, \bar{t})$ to equal (q, t) on Θ_0 and $(\tilde{q}, \tilde{t})$ on $\Theta \setminus \Theta_0$.

By the supermultiplicativity of α (see Proposition 4), we have

$$\alpha(\theta''|\theta) \geq \alpha(\theta''|\theta')\alpha(\theta'|\theta),$$

for all types $\theta \in \Theta$ and all reports $\theta' \in \Theta \setminus \Theta_0$ and $\theta'' \in \Theta_0$. Thus, it can be checked that $(\bar{q}, \bar{t})$ is incentive compatible.

Now we complete the proof. If $\Theta \setminus \Theta_0$ has measure zero, we get (i). If $\Theta \setminus \Theta_0$ has positive measure, we claim that (ii) holds. We show that for each $\theta \in \Theta \setminus \Theta_0$, the principal strictly prefers $(\tilde{q}(\theta), \tilde{t}(\theta))$ to $(q(\theta), t(\theta))$. Fix $\theta \in \Theta \setminus \Theta_0$. For each $\theta' \in \Theta_0$, we have

$$\alpha(\theta'|\theta)t(\theta') - c(\alpha(\theta'|\theta)q(\theta')) \geq \alpha(\theta'|\theta)t(\theta') - \alpha(\theta'|\theta)c(q(\theta')) \geq 0,$$

where the first inequality uses the convexity of c. We conclude that

$$\tilde{t}(\theta) - c(\tilde{q}(\theta)) \geq 0 > t(\theta) - c(q(\theta)).$$ □

Proof of Lemma 2. Let (q, U) be a bounded, incentive compatible quantity-utility pair. We first check that U is absolutely continuous. Choose θ and θ' such that $U(\theta') \geq U(\theta)$. By incentive compatibility,

$$U(\theta) \geq \alpha(\theta'|\theta)[U(\theta') + (\theta - \theta')q(\theta')].$$

Therefore,

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

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

$$0 \leq U(\theta') - U(\theta) \leq C\left|\int_{\theta'}^{\theta} (\lambda(\xi) + 1)\,\mathrm{d}\xi\right|,$$

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

Now we prove (15). Define the auxiliary function Δ on $[\underline{\theta}, \bar{\theta}]$ by

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

We prove that Δ is nonnegative. The function Δ is absolutely continuous since it is the product of absolutely continuous functions. Let $u(\theta'|\theta) = \alpha(\theta'|\theta)[\theta q(\theta') - t(\theta')]$. By 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.[39] Let $I(\theta) = \int_{\underline{\theta}}^{\theta} \alpha(\xi|\bar{\theta}) q(\xi)\,\mathrm{d}\xi$. At almost every θ in $[\underline{\theta}, \bar{\theta}]$, the absolutely continuous functions Δ, U, $\alpha(\cdot|\bar{\theta})$, and I are all differentiable, so we get

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

By the fundamental theorem of calculus, for $\underline{\theta} \leq \theta \leq \bar{\theta}$, we have

$$\Delta(\theta) \geq \Delta(\underline{\theta}) = U(\underline{\theta}) \geq 0,$$

where the last inequality follows from the participation constraint. □

A.10 *Proof of Proposition 6*

We follow the proof of Proposition 5 in Appendix A.9. As before, it suffices to prove essentially unique optimality among all bounded, incentive compatible mechanisms.[40] For any bounded, incentive compatible quantity-utility pair (q, U), we have

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

with equality if and only if (15) holds with equality for almost every type θ. For each type θ, the integrand on the right-hand side is maximized by $q^*(\theta)$, uniquely so if $\theta \neq \theta^*$. The transfer function t^* ensures that U satisfies (15) with equality for every type θ.

To check that (q^*, t^*) is globally incentive compatible, follow the argument from the proof of Proposition 5 in Appendix A.9.

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

[40] Any quantity function $q\colon \Theta \to [0, 1]$ is bounded. By Lemma 1, it suffices to consider bounded transfer functions.

A.11 *Beyond exponential authentication rates*

Suppose that the verification technology is represented by a Borel 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 type θ', the function $\theta \mapsto \alpha(\theta'|\theta)$ is absolutely continuous.

(iii) For each type θ, the right and left partial derivatives (with respect to the second argument) $D_{2+}\alpha(\theta|\theta)$ 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 (ii) and (iii) allow us to apply the envelope theorem. 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

$$\lambda_+(\theta) = -D_{2+}\alpha(\theta|\theta), \qquad \lambda_-(\theta) = D_{2-}\alpha(\theta, \theta). \tag{19}$$

Define the function Λ by

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

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

LEMMA 3 (Lower Bound on Authentication Rate). *For all types θ and θ', we have $\alpha(\theta'|\theta) \geq \Lambda(\theta'|\theta)$.*

Lemma 3 is established at the end of the proof. For the exponential authentication rate α considered in the main text, we have $\lambda_+(\theta) = \lambda_-(\theta) = \lambda(\theta)$, so $\alpha(\theta'|\theta) = \Lambda(\theta'|\theta)$ for all types θ and θ'. Therefore, among all most-discerning authentication rates satisfying (i)–(iii) with $-D_{2+}\alpha(\theta|\theta) = D_{2-}\alpha(\theta|\theta) = \lambda(\theta)$ for each θ, the exponential authentication rate with precision function λ makes the global incentive constraints weakest.

In this general setting, we show under further regularity conditions that the optimal mechanisms take the same form, except that the virtual value φ is defined with Λ in place of α:

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

Lemma 1 goes through with exactly the same proof. Lemma 2 can be shown to hold with $\Lambda(\xi|\theta)$ in place of $\alpha(\xi|\theta)$.[41] Therefore, Proposition 5 and Proposition 6 go through, with

[41] The proof is similar. To establish absolute continuity, apply Lemma 3 and put $\lambda_+ \vee \lambda_-$ in place of λ. To establish the bound, use Λ in place of α in the definition of the auxiliary function Δ. The rest of the proof goes through with λ_+ in place of λ.

the redefined virtual value, if (a) the monotonicity condition (16) holds with Λ in place of α, and (b) the following global bound is satisfied: For $\theta > \theta'$,

$$\alpha(\theta'|\theta) \le \Lambda(\theta'|\theta)\frac{\int_{\underline{\theta}}^{\theta'} \Lambda(\xi|\theta)q^*(\xi)\,d\xi + \int_{\theta'}^{\theta} \Lambda(\xi|\theta)q^*(\xi)\,d\xi}{\int_{\underline{\theta}}^{\theta'} \Lambda(\xi|\theta)q^*(\xi)\,d\xi + \int_{\theta'}^{\theta} \Lambda(\theta'|\theta)q^*(\theta')\,d\xi}. \tag{20}$$

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 $\Lambda(\theta'|\theta)$.

We check that (a) and (b) imply global incentive compatibility. By Lemma 2, the analogue of (18) is

$$\int_{\underline{\theta}}^{\theta} \Lambda(\xi|\theta)q^*(\xi)\,d\xi - \int_{\underline{\theta}}^{\theta'} \Lambda(\xi|\theta')\alpha(\theta'|\theta)q^*(\xi)\,d\xi \ge (\theta - \theta')\alpha(\theta'|\theta)q^*(\theta'). \tag{21}$$

We separate into cases. If $\theta < \theta'$, then (21) holds if

$$\int_{\theta}^{\theta'} \Lambda(\xi|\theta')q^*(\xi)\,d\xi \le (\theta' - \theta)q^*(\theta'),$$

which is guaranteed by (a). If $\theta > \theta'$, then (21) is equivalent to

$$\int_{\underline{\theta}}^{\theta} \Lambda(\xi|\theta)q^*(\xi)\,d\xi \ge \frac{\alpha(\theta'|\theta)}{\Lambda(\theta'|\theta)}\left[\int_{\underline{\theta}}^{\theta'} \Lambda(\xi|\theta)q^*(\xi)\,d\xi + (\theta - \theta')\Lambda(\theta'|\theta)q^*(\theta')\right].$$

Rearranging, we see that this inequality is equivalent to (20).

Proof of Lemma 3. Fix θ and θ'. For each h, supermultiplicativity (see Proposition 4) gives

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

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

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

Dividing by h and passing to the limit as $h \downarrow 0$ and $h \uparrow 0$, we see that whenever $D_2\alpha(\theta'|\theta)$ exists, we have

$$-\lambda_+(\theta)\alpha(\theta'|\theta) \le D_2\alpha(\theta'|\theta) \le \lambda_-(\theta)\alpha(\theta'|\theta).$$

Since α satisfies (ii) and (iii), we can 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 claim that $\Delta(\theta) \geq 1$ for all θ. Since $\Lambda(\theta'|\theta)$ is bounded away from 0, the function Δ is absolutely continuous. Therefore, the functions Δ, $\alpha(\theta'|\cdot)$, and $\Lambda(\theta'|\cdot)$ are simultaneously differentiable almost everywhere. If $\theta > \theta'$ and these three functions are simultaneously differentiable at θ, we have

$$\Delta'(\theta) = \frac{1}{\Lambda(\theta'|\theta)}\big[D_2\alpha(\theta'|\theta) + \lambda_+(\theta)\alpha(\theta'|\theta)\big] \geq 0.$$

If $\theta < \theta'$ and these three functions are simultaneously differentiable at θ, we have

$$\Delta'(\theta) = \frac{1}{\Lambda(\theta'|\theta)}\big[D_2\alpha(\theta'|\theta) - \lambda_-(\theta)\alpha(\theta'|\theta)\big] \leq 0.$$

Since Δ is absolutely continuous, it follows from the fundamental theorem of calculus that $\Delta(\theta) \geq \Delta(\theta') = 1$ for all θ. □

A.12 *Nonbinary tests*

First, we check that $\succeq_\theta$ is reflexive and transitive. Reflexivity is immediate by taking k to be the identity, which maps each score s to the point mass δ_s. For transitivity, it follows from (Kamae, Krengel, and O'Brien (1977, Proposition 1, pp. 901–902)) that (i) the order $\succeq_{\text{st}}$ is preserved by increasing Markov transitions; and (ii) the composition $k_1k_2\colon S \to \Delta(S)$ defined by $(k_1k_2)(s'|s) = \sum_{s''} k_2(s'|s'')k_1(s''|s)$ is increasing if k_1 and k_2 are increasing.

In the main model, if type θ tries on test τ with probability a, he passes with probability $a\pi(\tau|\theta)$. Therefore, type θ can achieve on test τ any passage probability p satisfying $p \leq \pi(\tau|\theta)$. In the general case, on a nonbinary test τ, type θ chooses a Markov transition $d\colon S \to \Delta(S)$ that is *downward* in the sense that $d(s'|s) = 0$ unless $s \succeq s'$. Then Nature draws the score from the distribution $\pi_{\tau|\theta}d$. By Kamae, Krengel, and O'Brien (1977, Theorem 1, p. 900), type θ can achieve on test τ a score distribution p in $\Delta(S)$ if and only if $p \preceq_{\text{st}} \pi_{\tau|\theta}$. Given a general mechanism $(M, M'; t, r', g)$, a strategy for the agent is a pair (r, d) consisting of a messaging strategy $r\colon \Theta \to \Delta(M)$ and an action strategy $d\colon \Theta \times M \times T \times M' \times S \to \Delta(S)$ such that $d_{\theta,m,\tau,m'}\colon S \to \Delta(S)$ is downward for each $(\theta, m, \tau, m') \in \Theta \times M \times T \times M'$.

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 the main implementation theorem (Theorem 2). The proofs are virtually identical, with the downward transition d in place of the trying probability a. The key property is that the composition of downward kernels is downward, which is easy to check.

A.13 *Universal measurability*

We begin by introducing universal measurability. For a more detailed discussion with proofs, see Bertsekas and Shreve (1996, Chapter 7). Let $(X, \mathcal{X})$ be a measurable space. Given a probability measure μ on $(X, \mathcal{X})$, let $\overline{\mathcal{X}}_\mu$ denote the μ-completion of $\mathcal{X}$, i.e.,

the σ-algebra generated by $\mathcal{X}$ and all μ-null sets of $\mathcal{X}$. The universal completion of $\mathcal{X}$, denoted $\overline{\mathcal{X}}$, is the intersection $\cap_\mu \overline{\mathcal{X}}_\mu$, where the intersection is taken over all probability measures μ on $(X, \mathcal{X})$. It can be shown that $\overline{\overline{\mathcal{X}}} = \overline{\mathcal{X}}$.

A function from $(X, \mathcal{X})$ to $(Y, \mathcal{Y})$ is universally measurable if it is $(\overline{\mathcal{X}}, \mathcal{Y})$-measurable. It can be shown that $(\overline{\mathcal{X}}, \mathcal{Y})$-measurability is equivalent to $(\overline{\mathcal{X}}, \overline{\mathcal{Y}})$-measurability. Similarly, it can be shown that any probability kernel from $(X, \overline{\mathcal{X}})$ to $(Y, \mathcal{Y})$ can be uniquely extended to a probability kernel from $(X, \overline{\mathcal{X}})$ to $(Y, \overline{\mathcal{Y}})$. Given $(X, \mathcal{X})$ and $(Y, \mathcal{Y})$, a probability kernel from $(X, \overline{\mathcal{X}})$ to $(Y, \overline{\mathcal{Y}})$ is called universally measurable.

On a topological space X, the Borel σ-algebra is denoted by $\mathcal{B}(X)$. For Polish spaces X and Y, we have $\mathcal{B}(X \times Y) = \mathcal{B}(X) \otimes \mathcal{B}(Y)$. The left-hand side is the σ-algebra generated by the product topology on $X \times Y$. The right-hand side is the σ-algebra generated by all rectangles with Borel-measurable sides.

Now we return to the model. We make the following standing technical assumptions. The sets Θ, T, and X are Polish spaces. The function $\pi\colon T \times \Theta \to \Delta(S)$ is Borel measurable (with $\Delta(S)$ viewed as a subset of $\mathbf{R}^S$).[42] In a mechanism, the message spaces M and M' are Polish, and all maps and probability kernels are universally measurable. Universally measurable sets are convenient because of the following measurable projection theorem (Cohn (2013, Proposition 8.4.4, p. 264)).

THEOREM 4 (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 $\overline{\mathcal{X}}$.*

The definition of θ-discernment imposes an inequality for each type θ'. If there are uncountably many types, this can create measurability problems. Using the measurable projection theorem, we can show that the score conversion in the definition of $\tau \succeq_\theta \psi$ can be selected in a universally measurable way.

A.14 *Measurable selection of score conversion*

For each triple $(\theta, \tau, \psi) \in \Theta \times T^2$ such that $\tau \succeq_\theta \psi$, there exists an associated score conversion satisfying Definition 4. Here, we show that this score conversion can be selected in a universally measurable way.

We represent the space of increasing Markov transitions $k\colon S \to \Delta(S)$ as a polytope $\mathcal{K}$ in $\mathbf{R}^{S\times S}$ consisting of vectors $k = (k(s'|s))_{s,s'\in S}$. Define the subset G of $\Theta \times T^2 \times \mathcal{K}$ to consist of all tuples (θ, τ, ψ, k) such that k satisfies the conditions in the definition of $\tau \succeq_\theta \psi$. We will show below that G is in $\overline{\mathcal{B}(\Theta \times T^2)} \otimes \mathcal{B}(\mathcal{K})$. Here, we use this claim to obtain the desired selection. By the measurable projection theorem, the projection of G onto $\Theta \times T^2$, which we call D, is in $\overline{\mathcal{B}(\Theta \times T^2)}$. For each $(\theta, \tau, \psi) \in D$, let $G_{\theta,\tau,\psi} = \{k \in \mathcal{K} : (\theta, \tau, \psi, k) \in G\}$. The measurable projection theorem also guarantees that the section correspondence $(\theta, \tau, \psi) \mapsto G_{\theta,\tau,\psi}$ on D is weakly measurable,[43] where

[42] We prove the measurability results in the nonbinary testing framework, which includes the main model as a special case.

[43] That is, the lower inverses of open sets are measurable. For each open subset A of $\mathcal{K}$, the lower inverse of A equals the projection of $G \cap (\Theta \times T^2 \times A)$ onto $\Theta \times T^2$.

D is endowed with the restriction of the σ-algebra $\overline{\mathcal{B}(\Theta \times T^2)}$. Finally, this section correspondence has nonempty, closed values, so we apply the Kuratowski–Ryll-Nardzewski selection theorem (Aliprantis and Border (2006, 18.13, p. 600)) to get the desired universally measurable selection.

Now we check that G is in $\overline{\mathcal{B}(\Theta \times T^2)} \otimes \mathcal{B}(\mathcal{K})$. On $\Theta \times T^2 \times \mathcal{K}$, define the real-valued functions f_s for each s in S, and g_U for each upper set $U \subset S$ by

$$f_s(\theta, \tau, \psi, k) = (\pi_{\tau|\theta} k)(s) - \pi_{\psi|\theta}(s),$$
$$g_U(\theta, \tau, \psi, k) = \sup_{\theta'} \big[(\pi_{\tau|\theta'} k)(U) - \pi_{\psi|\theta'}(U)\big].$$

The set G is the intersection of $\cap_s [f_s = 0]$ and $\cap_U [g_U \leq 0]$. Therefore, it suffices to check that these functions are all $(\overline{\mathcal{B}(\Theta \times T^2)} \otimes \mathcal{B}(\mathcal{K}), \mathcal{B}(\mathbf{R}))$-measurable. For each function f_s, this is implied by the Borel measurability of π. For each upper set U, we check that g_U is a Carathéodory function. For each fixed (θ, τ, ψ), the function $g_U(\theta, \tau, \psi, \cdot)$ is continuous. For each fixed k, the function $g_U(\cdot, k)$ is $(\overline{\mathcal{B}(\Theta \times T^2)}, \mathcal{B}(\mathbf{R}))$-measurable because the term in brackets, viewed as a function of $(\theta, \tau, \psi, \theta')$ is $(\mathcal{B}(\Theta \times T^2 \times \Theta), \mathcal{B}(\mathbf{R}))$-measurable. Hence, the supremum over θ' is $(\overline{\mathcal{B}(\Theta \times T^2)}, \mathcal{B}(\mathbf{R}))$-measurable by the measurable projection theorem.[44] Therefore, g_U is a Carathéodory function. By Aliprantis and Border (2006, 4.51, p. 153), g_U is $(\overline{\mathcal{B}(\Theta \times T^2)} \otimes \mathcal{B}(\mathcal{K}), \mathcal{B}(\mathbf{R}))$-measurable.

A.15 *Measurable selection of more θ-discerning test*

Consider the setting of Theorem 3. For each $(\psi, \theta) \in \Theta \times T$, there exists a test $\tau \in \hat{T}(\theta)$ such that $\tau \succeq_\theta \psi$. Here, we show that this test can be selected in a universally measurable way.

Consider the following subsets of $\Theta \times T^2$:

$$A = \big\{(\theta, \psi, \tau) : \tau \succeq_\theta \psi\big\}, \qquad B = \big\{(\theta, \psi, \tau) : \tau \in \hat{T}(\theta)\big\}.$$

It suffices to check that $A \cap B$ is in $\overline{\mathcal{B}(\Theta \times T)} \otimes \mathcal{B}(T)$ and that each section $(A \cap B)_{\theta,\psi}$ is closed. Then the section correspondence $(\theta, \psi) \mapsto (A \cap B)_{\theta,\psi}$ has a $(\overline{\mathcal{B}(\Theta \times T)}, \mathcal{B}(T))$-measurable selection by the same argument from Appendix A.14. We check that A and B are each in $\overline{\mathcal{B}(\Theta \times T)} \otimes \mathcal{B}(T)$ and have closed sections. The set $\mathcal{K}$ is compact, and by assumption π is continuous, so it is straightforward to check that A is closed. By assumption, $\hat{T}$ has closed values and is weakly $(\overline{\mathcal{B}(\Theta)}, \mathcal{B}(T))$-measurable, so the correspondence $(\theta, \psi) \mapsto \hat{T}(\theta)$ has closed values and is weakly $(\overline{\mathcal{B}(\Theta \times T)}, \mathcal{B}(T))$-measurable. Its graph, B, is therefore in $\overline{\mathcal{B}(\Theta \times T)} \otimes \mathcal{B}(T)$ by Aliprantis and Border (2006, 18.6, p. 596).

References

Aliprantis, Charalambos D. and Kim C. Border (2006), *Infinite Dimensional Analysis: A Hitchhiker's Guide*, third edition. Springer. [1271, 1274, 1280]

[44] For any bounded function $f\colon X \times Y \to \mathbf{R}$, define $F\colon X \to \mathbf{R}$ by $F(x) = \sup_{y \in Y} f(x, y)$. For any real t, the preimage $[F > t]$ is the projection of the preimage $[f > t]$ onto X.

Balbuzanov, Ivan (2019), "Lies and consequences." *International Journal of Game Theory*, 48, 1203–1240. [1259]

Ben-Porath, Elchanan, Eddie Dekel, and Barton L. Lipman (2014), "Optimal allocation with costly verification." *American Economic Review*, 104, 3779–3813. [1265]

Ben-Porath, Elchanan, Eddie Dekel, and Barton L. Lipman (2017), "Disclosure and choice." *Review of Economic Studies*, 85, 1471–1501. [1266]

Ben-Porath, Elchanan, Eddie Dekel, and Barton L. Lipman (2019), "Mechanisms with evidence: Commitment and robustness." *Econometrica*, 87, 529–566. [1265]

Ben-Porath, Elchanan, Eddie Dekel, and Barton L. Lipman (2023), "Mechanism design for acquisition of/stochastic evidence." Working paper. [1249, 1266, 1267]

Bertsekas, Dimitri and Steven E. Shreve (1996), *Stochastic Optimal Control: The Discrete-Time Case*. Athena Scientific. [1278]

Billot, Antoine, Itzhak Gilboa, and David Schmeidler (2008), "Axiomatization of an exponential similarity function." *Mathematical Social Sciences*, 55, 107–115. [1259]

Blackwell, David (1953), "Equivalent comparisons of experiments." *Annals of Mathematical Statistics*, 24, 265–272. [1255]

Border, Kim C. and Joel Sobel (1987), "Samurai accountant: A theory of auditing and plunder." *Review of Economic Studies*, 54, 525–540. [1261]

Bull, Jesse and Joel Watson (2004), "Evidence disclosure and verifiability." *Journal of Economic Theory*, 118, 1–31. [1266]

Bull, Jesse and Joel Watson (2007), "Hard evidence and mechanism design." *Games and Economic Behavior*, 58, 75–93. [1249, 1252, 1266]

Caragiannis, Ioannis, Edith Elkind, Mario Szegedy, and Lan Yu (2012), "Mechanism design: From partial to probabilistic verification." In *"Proceedings of the 13th ACM Conference on Electronic Commerce" EC'12*, 266–283, ACM, New York, NY. [1248, 1258, 1261, 1266]

Carbajal, Juan Carlos, and Jeffrey C. Ely (2013), "Mechanism design without revenue equivalence." *Journal of Economic Theory*, 148, 104–133. [1262]

Carbajal, Juan Carlos, and Jeffrey C. Ely (2016), "A model of price discrimination under loss aversion and state-contingent reference points." *Theoretical Economics*, 11, 455–485. [1262]

Cohn, Donald L. (2013), *Measure Theory*, second edition. Birkhäuser, Basel. [1279]

Crocker, Keither J. and John Morgan (1998), "Is honesty the best policy? Curtailing insurance fraud through optimal incentive contracts." *Journal of Political Economy*, 106, 355–375. [1257]

de Oliveira, Henrique (2018), "Blackwell's informativeness theorem using diagrams." *Games and Economic Behavior*, 109, 126–131. [1270]

Deb, Rahul and Colin Stewart (2018), "Optimal adaptive testing: Informativeness and incentives." *Theoretical Economics*, 13, 1233–1274. [1252]

Deneckere, Raymond and Sergei Severinov (2008), "Mechanism design with partial state verifiability." *Games and Economic Behavior*, 64, 487–513. [1266]

Deneckere, Raymond and Sergei Severinov (2022), "Screening, signalling and costly misrepresentation." *Canadian Journal of Economics/Revue canadienne d'économique*, 55, 1334–1370. [1257]

Dziuda, Wioletta and Christian Salas (2018), "Communication with detectable deceit." Available at SSRN 3234695. [1259]

Erlanson, Albin and Andreas Kleiner (2020), "Costly verification in collective decisions." *Theoretical Economics*, 15, 923–954. [1265]

Ferraioli, Diodato and Carmine Ventre (2018), "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*, 1930–1932, Richland, SC. [1248, 1258, 1266]

Green, Jerry R. and Jean-Jacques Laffont (1986), "Partially verifiable information and mechanism design." *Review of Economic Studies*, 53, 447–456. [1248, 1249, 1253, 1260, 1265, 1266]

Grossman, Sanford J. (1981), "The informational role of warranties and private disclosure about product quality." *Journal of Law and Economics*, 24, 461–483. [1266]

Halac, Marina and Pierre Yared (2020), "Commitment versus flexibility with costly verification." *Journal of Political Economy*, 128, 4523–4573. [1265]

Hart, Sergiu, Ilan Kremer, and Motty Perry (2017), "Evidence games: Truth and commitment." *American Economic Review*, 107, 690–713. [1266]

Kallenberg, Olav (2017), *Random Measures, Theory and Applications*, volume 77 of Probability Theory and Stochastic Modelling. Springer. [1268]

Kamae, Teturo, Ulrich Krengel, and George L. O'Brien (1977), "Stochastic inequalities on partially ordered spaces." *Annals of Probability*, 5, 899–912. [1265, 1278]

Kartik, Navin (2009), "Strategic communication with lying costs." *Review of Economic Studies*, 76, 1359–1395. [1257]

Kartik, Navin, Marco Ottaviani, and Francesco Squintani (2007), "Credulity, lies, and costly talk." *Journal of Economic Theory*, 134, 93–116. [1257]

Kartik, Navin and Olivier Tercieux (2012), "Implementation with evidence." *Theoretical Economics*, 7, 323–355. [1266]

Kephart, Andrew and Vincent Conitzer (2016), "The revelation principle for mechanism design with reporting costs." In *Proceedings of the 2016 ACM Conference on Economics and Computation*, 85–102. [1257]

Koessler, Frédéric and Eduardo Perez-Richet (2019), "Evidence reading mechanisms." *Social Choice and Welfare*, 53, 375–397. [1266]

Lacker, Jeffrey M. and John A. Weinberg (1989), "Optimal contracts under costly state falsification." *Journal of Political Economy*, 97, 1345–1363. [1257]

Li, Yunan (2020), "Mechanism design with costly verification and limited punishments." *Journal of Economic Theory*, 186, 1–54. [1265]

Lipman, Barton L. and Duane J. Seppi (1995), "Robust inference in communication games with partial provability." *Journal of Economic Theory*, 66, 370–405. [1249, 1266]

Maggi, Giovanni and Andrés Rodriguéz-Clare (1995), "Costly distortion of information in agency problems." *RAND Journal of Economics*, 26, 675–689. [1257]

Milgrom, Paul and Ilya Segal (2002), "Envelope theorems for arbitrary choice sets." *Econometrica*, 70, 583–601. [1275]

Milgrom, Paul R. (1981), "Good news and bad news: Representation theorems and applications." *The Bell Journal of Economics*, 12, 380–391. [1266]

Mussa, Michael and Sherwin Rosen (1978), "Monopoly and product quality." *Journal of Economic Theory*, 18, 301–317. [1263]

Myerson, Roger B. (1981), "Optimal auction design." *Mathematics of Operations Research*, 6, 58–73. [1264]

Myerson, Roger B. (1982), "Optimal coordination mechanisms in generalized principal–agent problems." *Journal of Mathematical Economics*, 10, 67–81. [1251]

Myerson, Roger B. (1984), "Cooperative games with imcomplete information." *International Journal of Game Theory*, 13, 69–96. [1252]

Mylovanov, Tymofiy and Andriy Zapechelnyuk (2017), "Optimal allocation with ex post verification and limited penalties." *American Economic Review*, 107, 2666–2694. [1265]

Postlewaite, Andrew (1979), "Manipulation via endowments." *Review of Economic Studies*, 46, 255–262. [1265]

Reuter, Marco (2023), "Revenue maximization with partially verifiable information." ZEW—Centre for European Economic Research Discussion Paper No. 51. [1260]

Riley, John and Richard Zeckhauser (1983), "Optimal selling strategies: When to haggle, when to hold firm." *Quarterly Journal of Economics*, 98, 267–289. [1264]

Sher, Itai and Rakesh Vohra (2015), "Price discrimination through communication." *Theoretical Economics*, 10, 597–648. [1264]

Strausz, Roland and Sebastian Schweighofer-Kodritsch (2023), "Principled mechanism design with evidence." Berlin School of Economics Discussion Paper 30. [1253, 1266]

Townsend, Robert M. (1979), "Optimal contracts and competitive markets with costly state verification." *Journal of Economic Theory*, 21, 265–293. [1265]

Co-editor Rakesh Vohra handled this manuscript.

Manuscript received 29 August, 2024; final version accepted 17 January, 2025; available online 30 January, 2025.