Probabilistic Verification in Mechanism Design*

Ian Ball† Deniz Kattwinkel‡

February 26, 2019

Abstract

We introduce a model of probabilistic verification into the standard mechanism design setting. The principal can verify the truthfulness of the reported type with a statistical test. Testing generates a binary outcome—pass or fail—that depends stochastically on the agent's true type and reported type. The principal commits to a mechanism, which assigns a test to each message and then a decision based on the test outcome. We solve for the optimal mechanism under quasilinear preferences. If verification is more accurate, then the optimal allocation is more efficient, and the principal extracts a greater share of the surplus.

Keywords: probabilistic verification, revelation principle, evidence.
JEL Codes: D82, D86.

*We thank our advisors, Dirk Bergemann and Stephan Lauermann, for continual guidance. For helpful discussions, we thank Tilman Börgers, Marina Halac, Navin Kartik, Bart Lipman, Jacopo Perego, Larry Samuelson, Sebastian Schweighofer-Kodritsch, Philipp Strack, Roland Strausz, and Juuso Välimäki.

†Department of Economics, Yale University, ian.ball@yale.edu.

‡Bonn Graduate School of Economics, Universität Bonn, denizkattwinkel@gmail.com.

Contents

1 Introduction

Private information is not entirely private—it can often be verified, at least partially. For instance, if a U.S. taxpayer claims a tax deduction, the IRS can demand receipts proving eligibility. A consumer applying for a new credit card is asked to report his income, and the credit card company can request a monthly pay stub as confirmation. Disability insurance claimants must undergo medical tests to confirm the legitimacy of their claims.

In the classical paradigm of mechanism design, the principal cannot perform such verification. Instead, the principal elicits messages from the agents. The mechanism must be designed so that the agents are willing to reveal their private information. How can the principal use verification to encourage agents to reveal their information? More generally, how does the verification technology change the set of implementable social choice functions?

These questions require a formal model of verification in mechanism design. Green and Laffont (1986) provide the first such model. They restrict the messages that each type can send, and famously show that these restrictions invalidate the revelation principle. Subsequent models have introduced hard evidence (Bull and Watson, 2004, 2007) or included verification as one dimension of the economic outcome (Strausz, 2016). These models of verification inherently assume that the ability of a type to mimicking another type is either perfect or completely lacking.

We take a distinct approach by modeling verification as a part of the principal's *technology* and by introducing randomness into the verification outcome.

In Section 2, we present our model of verification in a standard principal–agent setting. The agent has a private type and the principal controls decisions. The verification technology is represented by a family of *binary tests*. A test is characterized by the probability with which each type can pass it. The agent is free to fail the test, so really this probability is an upper bound.

Following the tradition in mechanism design, we assume that the principal has full commitment and can elicit messages from the agent. With the new verification technology, a mechanism has two parts—a testing rule and a decision rule. The testing rule specifies the test that will be conducted following each message. The decision rule species the decision that will be chosen following the agent's message, the selected test, and the binary test outcome. The agent's strategy also has two parts. First he chooses what message to send. Then, after observing the test that the principal has selected, he chooses the probability with which he will pass the test, subject

to the type-specific upper bound imposed by the technology.

In Section 3, we reduce the class of mechanisms in two stages. First, we apply the standard revelation principle (Proposition 1), which remains valid under our approach. Next, we reduce the class of testing functions that need to be considered (Theorem 1): For each fixed type, we define a partial order on the space of tests that compares how well tests can distinguish that type from other types. If for each type there exists a test that is most-discerning for that type, then there is no loss in restricting to testing functions that assign to each type report the corresponding most discerning test. From this identification, we define the *authentication rate*, which specifies the probability with which type θ can pass the test identified with type θ'. We also show that being most discerning for a certain type is necessary for a te t to be the sa In Section 4, we work with the authentication rate directly. We characterize whether a given authentication rate can be induced by an underlying testing technology (Theorem 2). Previous attempts to introduce verification have worked directly with the reduced form and then imposed various conditions on the reduced form. Our characterization result makes two contributions. First, it substantially generalizes the previous conditions to allow for probabilistic rather than partial verification, Second, our condition sheds new light on the interpretation of the previous conditions. In Section 5 we use the authentication rate micro-founded above to solve for optimal mechanisms in a few standard mechanism design settings. Previous models were difficult to analyze because nonrandom verification invalidated the first-order envelope approach. The structure of the optimal mechanism in these models reflects the perfectness of their verification technology: mechanism are often discontinuous and the local incentive constraints are not binding (Townsend, 1979; Ben-Porath et al., 2014; Erlanson and Kleiner, 2015); the few models of probabilistic verification are reduced to nonrandom verification through unbounded punishments (Ferraioli and Ventre, 2018; Caragiannis et al., 2012).

We use the first-order envelope approach to derive a suitable *virtual value* that reflects the verification technology. The solution methods are similar to the classical methods, except this new virtual value replaces Myerson's expression for the virtual value. Consequently, the impact of verification technology on the optimal allocation is cleanly encoded in this single expression.

We characterize the optimal mechanism with verification for the nonlinear pricing problem (Theorem 3), the selling of a single indivisible good and, the auction setting Theorem 4.

Section 6 connects our modelling approach through tests to previous ap-

proaches. In particular, our model can be interpreted as a model of stochastic evidence. We also discuss related literature that has proceeded, in parallel, in both the economics and computer science literature. Proofs are in Appendix A, along with formal measure-theoretic definitions.

2 Model

2.1 Setting

Consider a standard principal–agent setting. The agent draws a private type $\theta \in \Theta$ from a commonly known distribution $\mu \in \Delta(\Theta)$. The principal controls decisions $x \in X$. Preferences for both players depend on the decision x and the agent's type θ. The Bernoulli utility functions for the agent and principal, respectively, are denoted by

$$u\colon X \times \Theta \to \mathbf{R} \quad \text{and} \quad v\colon X \times \Theta \to \mathbf{R}.$$

The object of interest is a social choice function

$$f\colon \Theta \to \Delta(X),$$

which assigns a decision lottery to each type.[1]

2.2 Verification technology

To the principal–agent setting we add a verification technology, in the form of statistical testing. There is a set T of available tests, with generic element τ. The principal can conduct one test from this set. Each test generates a binary outcome—pass or fail. The agent's type determines the probability that he is *able to* pass each test, but the agent is always *free to fail* a test. This assumption breaks the symmetry between passage and failure; they are not arbitrary labels.

The testing technology is characterized by the *performance function*

$$p\colon \Theta \times T \to \Delta(\{0,1\}),$$

[1] Each set is assumed to be endowed with a σ-algebra, and all functions are assumed to be measurable. The space of probability measures on a measurable space $(Z, \mathcal{Z})$ is denoted $\Delta(Z)$. Every function into a space of probability distributions is assumed to satisfy the requirements of a *stochastic kernel*. That is, a function k from a measurable space Y into $\Delta(Z)$ is formally a map from $Y \times \mathcal{Z}$ to $[0,1]$ such that for each y, the map $k(y,\cdot)$ is in $\Delta(Z)$, and for each $E \in \mathcal{Z}$, the map $k(\cdot, E)$ is a measurable function from Y into $[0,1]$, where $[0,1]$ is endowed with the usual Borel σ-algebra.

which assigns to each type θ and test τ a distribution over the set $\{0, 1\}$ of outcomes, where 1 denotes passage and 0 denotes failure. This distribution is denoted $p_{\theta,\tau}$, and can be identified with the passage probability, denoted $p(\theta, \tau)$. If the agent has type θ and is given test τ, then he can elect to pass the test with any probability between 0 and $p(\theta, \tau)$.

Assume that for each test τ, the function $p(\cdot, \tau)$ is not constant, so that performance on test τ is not independent of the agent's type. This assumption is without loss of generality because we can remove all uninformative tests from the testing set without changing the set of implementable social choice rules.[2]

In practice, verification can take many forms: evidence, questioning, examinations. Our model abstracts from the *process* of verification; instead, we identify the verification technology with its statistical properties, which determine the agent's incentives.

Our abstract notion of testing has two features. First, the principal and agent understand the testing technology in the sense that they have a common understanding of the performance function p. As a special case, this allows for deterministic testing, in which case p simply specifies which types can pass which tests. More generally, we our model allows for test outcomes to be stochastic.

Second, the agent is *free to fail*. The agent plays an active role in determining his performance on a test, subject to the constraints imposed by the passage function p. One interpretation is that each test corresponds to a piece of evidence that the principal can ask for. Then freedom to fail reflects the fact that an agent with evidence can choose not to provide it. But this applies more broadly than evidence. Often, the agent knows at least one way to assure failure on a test, even if he does not know exactly how to pass.

2.3 Mechanisms, strategies, and implementation

As in the standard mechanism design framework, the principal has full commitment power and elicits (cheap-talk) messages from the agent. The principal commits to the test she will conduct as a function of the message. Then she observes the binary outcome of the test and commits to a decision as a function of the message, the test, and the test outcome. This is formalized as follows.

Definition 1 (Mechanism)**.** A *mechanism* consists of a message space M

[2] We use this assumption in Section 4 to rule out uninteresting cases where the authentication rate is not unique.

together with a *testing rule* $g^1 \colon M \to \Delta(T)$ and a *decision rule* $g^2 \colon M \times T \times \{0,1\} \to \Delta(X)$.[3] Such a mechanism is denoted (M, g).

The principal's mechanism induces a multistage decision problem for the agent, with the following timing.[4]

1. The principal commits to a mechanism (M, g).
2. Nature draws the agent's type θ according to μ.
3. The agent observes θ and chooses a message m to send.
4. The principal observes message m and selects test τ according to testing rule g^1
5. The agent observes test τ and chooses whether to intentionally fail the test.
6. The principal observes the test outcome and takes a decision $x \in X$ according to decision rule g_2.

We allow the agent to mix his actions, and we next define a (behavioral) strategy for the agent.

Definition 2 (Strategy)**.** A *strategy* $a = (a^1, a^2)$ for the agent consists of a message strategy $a^1 \colon \Theta \to \Delta(M)$ and a performance strategy $a^2 \colon \Theta \times M \times T \to [0,1]$ satisfying $a^2(\theta, m, \tau) \leq p(\theta, \tau)$ for all $\theta \in \Theta$, $m \in M$, and $\tau \in T$.

The performance function p imposes an upper bound on the passage rate that the agent selects. This constraint is incorporated into the definition of a strategy, so it is redundant to speak of a "feasible" strategy. The principal's mechanism and the agent's strategy together determine a stochastic process, represented as

$$\xrightarrow{\mu} \Theta \xrightarrow{a^1} M \xrightarrow{g^1} T \xrightarrow{a^2} \{0,1\} \xrightarrow{g^2} X. \tag{1}$$

This diagram should not be interpreted as a Markov process. The arrows indicate transitions,but not the full dependence. Transitions controlled by the agent depend on all previous states; transitions controlled by the principal depend on all previous states except the true type.

Because there are so many successive stages of randomization, writing out distributions of decisions quickly becomes unwieldy. To get around this problem, we adopt formalism from Markov processes to represent the

[3] Since the principal has commitment power, the decision rule need only specify decisions following history (m, t, z) if test t is in the support of g^1 after message m. To simplify notation, we do not restrict the domain of g^2, but sometimes we will not specify the decisions on these unsupported histories.

[4] Equivalently, an extensive form game in which the agent is the only strategy player.

composition of behavioral strategies. The meaning should be intuitively clear, and this is sufficient to follow the main text. The details of the notation are only needed for the proofs, and can be found in Appendix A.1.

In particular, the the mechanism g and the strategy a together induce the social choice function

$$\mu \otimes a^1 \otimes g^1 \otimes a^2 \otimes g^2 \in \Delta(\Theta \times M \times T \times \{0,1\} \times X).$$

This joint distribution is constructed in the natural way, which is defined via integration in Appendix A.1. We can also compute the induced social choice function as[5]

$$f = (a^1 \otimes g^1 \otimes a^2)g^2.$$

We conclude this section with the suitable definitions of implementation in our setting. Social choice functions depend both on the principal's mechanism and the agent's strategy. A *profile* consists of a mechanism (M, g) and a strategy a such that a is a best response to (M, g). A profile (M, g, a) implements the social function $f = (a^1 \otimes g^1 \otimes a^2)g^2$. We are interested in the class of all social choice functions that are implemented by some profile. Such social choice functions are *implementable*. The space of all mechanisms is large, so we next find a smaller classes of mechanisms that is sufficient to trace out all implementable social choice functions.

3 Reducing the class of mechanisms

Characterizing the class of implementable social choice rules is challenging because we must consider profiles (M, g^1, g^2, a^1, a^2). In this section, we show in stages that it is without loss to consider a special subclass of profiles. First, we use the classical revelation principal to show that, without loss, we may consider profiles that are *direct* ($M = \Theta$) and *truthful* ($a^1 = \mathrm{id}$). Second we show that without loss, we may consider profiles in which the agent does not voluntarily fail, that is, $a^1_{\theta,m,\tau} = p_{\theta,\tau}$ for all types θ, messages m, and tests τ. This leaves only the testing rule g^1 and the decision rule g^2. Third,

[5]Here, the transitions kernels are built up as follows:

$$\begin{aligned} a^1 &: \Theta \to \Delta(M), \\ a^1 \otimes g^1 &: \Theta \to \Delta(M \times T), \\ a^1 \otimes g^1 \otimes a^2 &: \Theta \to \Delta(M \times T \times \{0,1\}), \\ (a^1 \otimes g^1 \otimes a^2)g^2 &: \Theta \to X. \end{aligned}$$

we give a condition under which we may focus on a single, fixed testing rule g^1 that assigns to each type report the most discerning test for that type. In total, these simplifications reduces the characterizations from considering all five objects in the profile to simply the decision function g^2.

3.1 Revelation principle

First, we apply the classical revelation principle. The argument goes though as usual because we model verification as a technology, rather than a restriction on the message space. Such a message restriction can invalidate the revelation principle, as shown by Green and Laffont (1986).

Proposition 1 (Revelation principle)
If a social choice function f is implemented by a profile (M, g, a), then f is also implemented by a profile $(\hat{M}, \hat{g}, \hat{a})$ with $\hat{M} = \Theta$ and $\hat{a}_1 = \mathrm{id}$.

3.2 No failure

We can further simplify the problem by restricting attention to strategies in which the agent does not fail freely. Indeed, if the agent chooses to fail, redefine the decision lottery following passage so that by passing with maximal probability the agent replicates the previous decision distribution.

Proposition 2 (No voluntary failure)
If a social choice function f is implemented by a direct, truthful profile (M, g, a), then f is also implemented by a direct, truthful profile for which $\hat{a}^2_{\theta,m,\tau} = p_{\theta,\tau}$ for all $\theta \in \Theta$, $m \in M$, and $\tau \in T$.

We can restrict attention to a particular deviation. The agent can misreport his type and then freely fail, and we will compare this deviation to truthful reporting and then no voluntary failure. The challenge is that we don't know which test is being used. This is where the more discerning order comes in, as we describe next.

3.3 Most discerning tests

The revelation principle substantially reduces the class of mechanisms we need to consider. But characterizing the space of implementable social choice rules remains a challenge because the principal must jointly choose the testing function and the decision function. The agent's incentives to misreport depend on the testing function.

In this section, we introduce an order on the class of tests. Ultimately we provide a condition for a test to be a maximum with respect to this order, in which case we may restrict attention to mechanisms in which a particular mapping is chosen. If this condition is met it is without loss to restrict attention to a particular mapping g^1 from Θ into T. This means that irrespective of the preferences and the decision problem, the principal will always conduct the same test $g^1(\theta)$ on an agent who reports type θ.

To define the partial order, we introduce a few further definitions. A probability measure ν on $\mathbf{R}$ has an associated right-continuous cumulative distribution function $F\colon \mathbf{R} \to [0,1]$ and an associated left-continuous quantile function $Q\colon (0,1) \to \mathbf{R}$. If ν has compact support, then we can also define Q on the endpoints 0 and 1. We need to generalize these objects to kernels. The *cumulative distribution kernel*

$$\tilde{F}\colon \mathbf{R} \to \Delta([0,1])$$

assigns to each point s the uniform distribution on $[F(s-), F(s)]$, where $F(s-) = \lim_{r \downarrow s} F(r)$. If f is continuous at s, then this interval is a point, and the uniform distribution is the point mass $\delta_{F(s)}$ at continuity points s of F. Formally the *quantile kernel*

$$\tilde{Q}\colon [0,1] \to \Delta(\mathbf{R})$$

assigns to each point $q \in [0,1]$ the point mass $\delta_{Q(q)}$.

For each type–test pair (θ, τ), the probability measure $p_{\theta,\tau}$ has these associated kernels, which we denote by $\tilde{F}_{\theta,\tau}$ and $\tilde{Q}_{\theta,\tau}$, respectively. Finally, we define the composition of kernels as in the theory of Markov processes.[6] We can now introduce a new type-specify order on the test space.

Definition 3 (Discernment order)**.** Fix a type $\theta \in \Theta$. Test τ is *more θ-discerning* than test ψ, denoted $\tau \succeq_\theta \psi$, if for all types $\theta' \in \Theta$,

$$p_{\theta',\tau}\tilde{F}_{\theta,\tau}\tilde{Q}_{\theta,\psi} \preceq_1 p_{\theta',\psi}. \tag{2}$$

The idea of this condition is that if type θ is assigned to test ψ, then the principal can replace test ψ by test τ without introducing any new deviations. This condition can be replaced with the piecewise inequality

[6] Let P be a measure on $\mathbf{R}$ and F a cumulative distribuion kernel. $P\tilde{F}$ denotes the measure that is given by $P\tilde{F}(A) = \int_{\mathbf{R}} \tilde{F}(s)(A)\,\mathrm{d}P(s)$.

that

$$p(\theta,\tau) \geq p(\theta,\psi) \implies \frac{p(\theta,\tau)}{p(\theta',\tau)} \geq \frac{p(\theta,\psi)}{p(\theta',\psi)},$$
$$p(\theta,\tau) < p(\theta',\psi) \implies \frac{1-p(\theta',\tau)}{1-p(\theta,\tau)} \geq \frac{1-p(\theta',\psi)}{1-p(\theta,\psi)},$$

provided that the denominators do not vanish. Cross-multiplying gives a condition that makes sense even when some terms are zero.There are two separate cases, depending on the relative passage rate of type θ on tests τ and ψ. If type θ is more likely to pass test τ than test ψ, then the relative passage rate of type θ compared to type θ' must be greater on test τ than on test ψ. If type θ is more like to fail test τ than test ψ, then the relative failure rate of type θ compared to type θ' must be smaller on test τ than on test ψ.

Definition 4 (Most discerning test)**.** Test τ is *most θ-discerning* if $\tau \succeq_\theta \psi$ for all tests $\psi \in T$.

The definition of most discerning tests has a natural extension to testing fuctions.

Definition 5 (Most discerning testing function)**.** A testing function $g^1 : \Theta \to T$ is *most discerning* if for each type θ, the test $g^1(\theta)$ is most θ-discerning.

Recall that the revelation principle (Proposition 1) and the no-failure result (Proposition 2) tell us that we need only consider direct, truthful profiles with no voluntary failure. Hereafter, we will call this *canonical implementation*. The two propositions tell us that every implementable social choice function is canonically implementable.

Now we can turn to the main implementation result.

Theorem 1 (Implementation)
Let g^1 be a testing function.

1. *If g^1 is most discerning, then for every utility function u, the following holds: Every implementable social choice function can be canonically implemented with testing function g_1.*
2. *If g^1 is not most discerning, then for some utility function u, the following holds: There exists a social choice function f such that with respect to u, the function f is canonically implementable but not with the testing function g_1.*

Part 1 says that our discernment order is sufficient. When there exists a most discerning testing function, then mechanisms using this testing functions trace out the space of all implementable social choice functions. This is the third and final step in reducing the class of profiles that we need to consider.

Part 2 says that our discernment order is necessary. To be sure, preferences matter. As an extreme case, if every type of agent is indifferent between every decision, then every social choice function is implementable, even without any verification technology. But our condition is necessary if we want to restrict to mechanisms with a fixed testing function that works *no matter the agent's preferences.*

4 Reduced-form authentication rate

4.1 Characterization

Suppose that for every type there is at least one mostly discerning test. We find a reduced form authentication rate $\alpha\colon \Theta \times \Theta \to [0,1]$ that suffices to characterize the underlying testing environment.

Let g^1 be a mostly discerning testing rule. The *authentication rate induced by* g^1 is the function $\alpha\colon \Theta \times \Theta \to [0,1]$ defined by

$$\alpha(\theta, \theta') = p(\theta, g^1(\theta')).$$

Proposition 3 (Authentication rate)
Every mostly discerning testing rule induces the same authentication rate.

This common rate is called the *authentication rate induced by the testing technology*, and it is defined for any testing technology for which there exists a most discerning testing rule.

Four our main characterization result, we state the following condition.

Definition 6 (Normality)**.** An authentication rate is *normal* if, for all $\theta_1, \theta_2, \theta_3$, we have

$$\alpha(\theta_1, \theta_3)\alpha(\theta_2, \theta_2) \geq \alpha(\theta_1, \theta_2)\alpha(\theta_2, \theta_3),$$

whenever $\alpha(\theta_2, \theta_2) \geq \alpha(\theta_2, \theta_3)$, and we have

$$(1 - \alpha(\theta_1, \theta_3))(1 - \alpha(\theta_2, \theta_2)) \leq (! - \alpha(\theta_1, \theta_2))(1 - \alpha(\theta_2, \theta_3)),$$

whenever $\alpha(\theta_2, \theta_2) < \alpha(\theta_2, \theta_3)$.

Theorem 2 (Reduced-form characterization)
An authentication rate $\alpha\colon \Theta \times \Theta \to [0,1]$ is induced by some testing technology if and only if α is normal.

A function $\alpha\colon \Theta \times \Theta \to [0,1]$ is *transitive* if

$$\alpha(\theta_1, \theta_3)\alpha(\theta_2, \theta_2) \geq \alpha(\theta_1, \theta_2)\alpha(\theta_2, \theta_3),$$

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

Corollary 1 (Transitivity)
If an authentication rate $\alpha\colon \Theta \times \Theta \to [0,1]$ satisfies $\alpha(\theta, \theta) \geq \alpha(\theta, \theta')$ for all $\theta, \theta' \in \Theta$, then α is induced by some testing technology if and only if α is transitive.

The nested range condition used in Green and Laffont (1986) is a special case of the transitivity condition here. The message correspondence in Green and Laffont (1986) can be represented in our language as a $\{0,1\}$-valued authentication rate with $\alpha(\theta, \theta) = 1$ for all θ. In particular, $\alpha(\theta, \theta) = 1 \geq \alpha(\theta, \theta')$ for all types θ and θ' so Corollary 1 says that this message correspondence can be induced by some underlying testing technology if and only if

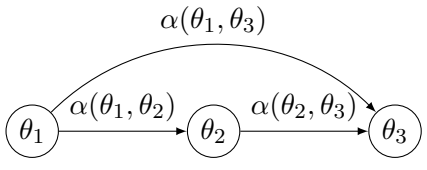
$$\alpha(\theta_1, \theta_3) \geq \alpha(\theta_1, \theta_2)\alpha(\theta_2, \theta_3)$$

for all $\theta_1, \theta_2, \theta_3 \in \Theta$, where we have suppressed the term $\alpha(\theta_2, \theta_2)$ on the left side because it equals 1. If either $\alpha(\theta_1, \theta_2)$ or $\alpha(\theta_2, \theta_3)$ vanishes, then the inequality is immediately satisfied. Thus, it requires that if θ_1 can mimic θ_2, and θ_2 can mimic θ_3, then θ_1 can mimic θ_3, which is precisely the nested range condition.

4.2 Example with three types

The agent has three possible types, $\Theta = \{\theta_1, \theta_2, \theta_3\}$. When can $\alpha\colon \Theta \times \Theta \to [0,1]$ be the reduced form of an underlying test technology? We assume here that $\alpha(\theta, \theta) = 1$ for all $\theta \in \Theta$: If α is a reduced form, this means that there is no false detection.

We further assume that $\alpha(\theta_i, \theta_j) = 0$ whenever $i > j$. With all these assumption in place the question whether α can be the reduced form of a mostly discerning test technology boils down to a single inequality:



$$\alpha(\theta_1, \theta_3) \geq \alpha(\theta_1, \theta_2)\alpha(\theta_2, \theta_3).$$

For α with $\alpha(\theta_1, \theta_2) = 1/3, \alpha(\theta_2, \theta_3) = 1/2$ and $\alpha(\theta_1, \theta_3) = 1/8$ the inequality is violated. This means that there cannot exists a test technology with mostly discerning tests that would induce α.

Suppose that α is induced by a selection of tests (that are not all mostly discerning nonetheless. Denote the test that is associated with type θ_i by τ_i. The assumptions on α lead to the following passage probabilities:

$$\begin{pmatrix} p(\theta_1, \tau_1) & p(\theta_1, \tau_2) & p(\theta_1, \tau_3) \\ p(\theta_2, \tau_1) & p(\theta_2, \tau_2) & p(\theta_2, \tau_3) \\ p(\theta_3, \tau_1) & p(\theta_3, \tau_2) & p(\theta_3, \tau_3) \end{pmatrix} = \begin{pmatrix} 1 & 1/3 & 1/8 \\ 0 & 1 & 1/2 \\ 0 & 0 & 1 \end{pmatrix}$$

Suppose that these are the only tests at the disposal of the principal, $T = \{\tau_1, \tau_2, \tau_3\}$, We want to demonstrate that in this setting the choice of the optimal test (g^1) for any type report (in a direct mechanism) can no longer be disentangled from the design of the subsequent decision rule (g^2).

Consider an allocation problem; $X = [0, 1]$ describes the probability of allocating a good to the agent. The agent wants to maximize the probability of receiving the good irrespective of his type. First consider the social choice function $f = (1/8, 1/2, 1)$, where component i is the value at θ_i. This function is implemented by the mechanism

$$\begin{aligned} (g^1(\theta_1), g^1(\theta_2), g^1(\theta_3)) &= (\tau_1, \tau_3, \tau_3), \\ g^2(\theta_1, \tau_1, 1), g^2(\theta_2, \tau_3, 1), g^3(\theta_3, \tau_3, 1)) &= (1/8, 1, 1), \end{aligned}$$

where $g^2(\theta_1, \tau_1, 1)$ denotes the allocation probability after a type report θ_1, the conduction of test τ_1 and the test outcome 1. The allocation probabilities after test outcome 0 is set to 0. Any direct mechanism that implements f must counter a type report θ_2 with test τ_3. Otherwise an agent of type θ_1 will always find a profitable to misreport his type as θ_2. Next consider the social choice function $\tilde{f} = (1/3, 1, 0)$. It is implemented by

$$\begin{aligned} (g^1(\theta_1), g^1(\theta_2), g^1(\theta_3)) &= (\tau_1, \tau_2, \tau_3), \\ g^2(\theta_1, \tau_1, 1), g^2(\theta_2, \tau_2, 1), g^3(\theta_3, \tau_3, 1)) &= (1/3, 1, 0). \end{aligned}$$

any implementation must have a type report θ_2 followed by a test τ_2. Otherwise the good cannot be allocated to type θ_2 with probability 1.

These examples of social choice functions show that there cannot be a fixed test that is assigned to report θ_2 in all problems.

The example shows that without mostly discerning tests, the test choice cannot be reduced into α. Correspondingly one cannot assume a reduced form mimicking relationship α that describes the ability of a type to mimic another type universally, that is in any possible preference and decision environment, if it does not obey the condition of Theorem 2.

5 Applications

We now take advantage of the reduced-form representation to solve a few classical problems in mechanism design with quasilinear utility. The solutions can be expressed in terms of a new expression for the virtual value that reflects the verification technology.

For the applications, consider an authentication rate $\alpha\colon \Theta \times \Theta \to [0,1]$ satisfying the following assumptions.

1. $\alpha(\theta,\theta) = 1$ for all θ.
2. α is transitive.
3. For each θ', the function $\alpha(\cdot,\theta')$ is absolutely continuous.
4. For each θ, the right and left partial derivatives $\partial_1\alpha(\theta+,\theta)$ and $\partial_1\alpha(\theta-,\theta)$ exist, and the functions $\theta \mapsto \partial_1\alpha(\theta+,\theta)$ and $\theta \mapsto \partial_1\alpha(\theta-,\theta)$ are integrable.

5.1 Nonlinear pricing setting

Consider the standard nonlinear pricing setting from Mussa and Rosen (1978). The agent's type θ is drawn from an interval $\Theta = [\underline{\theta}, \bar{\theta}]$ according to a cdf F with strictly positive density f. The agent's utility from consuming quantity q and making transfer t is

$$u(q,t,\theta) = \theta q - t.$$

The principal's utility from receiving transfer t and providing quantity q is

$$v(q,t) = t - c(q),$$

where the cost function c is strictly increasing and strictly convex. We further assume $c(0) = c'(0) = 0$ and $\lim_{q\to\infty} c'(q) = \infty$. The principal's cost is independent of the agent's type.

The novelty is that the principal has access to detection technology characterized by some function α. The agent is free to walk away at any time, so the participation constraint must be satisfied ex post, even if the agent is detected as lying. In particular, we rule out upfront payments.[7]

By the truthfulness principle, we may restrict attention to mechanisms in which the agent reports truthfully since doing so does not reduce the set of resulting social choice functions. Since $\alpha(\theta,\theta) = 1$ for all types θ, lies will not be detected on-path. Therefore, we may assume that if a lie

[7] For a model allowing upfront payments, see Border and Sobel (1987).

is detected, the principal holds the agent to zero utility, e.g., by refusing service ($q = t = 0$).

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

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

The first constraint is incentive compatibility. The agent is authenticated with probability $\alpha(\theta, \theta')$ and in this case his utility is $\theta q(\theta') - t(\theta')$. With complementary probability, the agent's lie is detected so he receives utility zero. The second constraint is the participation constraint.[8]

5.2 Virtual value

For each type θ, let

$$\lambda(\theta) = -\partial_1 \alpha(\theta+, \theta),$$

where the derivative is the right partial derivative with respect to the true type. This function λ parametrizes the local precision of the verificaiton technology. For $\theta \geq \theta'$, let

$$\Lambda(\theta, \theta') = \exp\left(- \int_{\theta'}^{\theta} \lambda(s)\, \mathrm{d}s \right).$$

In this section, we provide an informal derivation of the suitable virtual value in the presence of verification. Recall Myerson's *virtual value*

$$\varphi^M(\theta) = \theta - \frac{1 - F(\theta)}{f(\theta)}.$$

For intuition, it is convenient to express the virtual value in terms of an integral as

$$\theta - \frac{1}{f(\theta)} \int_{\theta}^{\bar{\theta}} f(s)\, \mathrm{d}s. \tag{3}$$

The virtual value $\varphi(\theta)$ is the marginal effect on total revenue from allocating an additional unit to type θ. This effect comes in two parts. First, the effect

[8] More precisely, this is the ex post participation constraint following authentication. We are assuming that, after detection, the ex post participation constraint holds with equality.

on the revenue from type θ is just θ because the principal can extract the additional utility that type θ gets from the higher allocation. The second reflects the loss of revenue on all higher types, weighted by the density of those types.

With verification, the marginal effect on the utility of higher types is much smaller. Indeed, the effect of type θ' on a higher type θ is proportional to $\Lambda(\theta, \theta')$. Formally, let

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

Comparing with Myerson's expression, it is immediate that

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

5.3 Nonlinear pricing solution

Theorem 3 (Optimal nonlinear pricing)
If φ is weakly increasing, then the optimal quantity and transfer functions $q^\star$ and t^* are unique and given by

$$c'(q^\star(\theta)) = \varphi(\theta)_+, \qquad t^\star(\theta) = \theta q^\star(\theta) - \int_{\underline{\theta}}^{\theta} \Lambda(\theta, s) q^\star(s) \, \mathrm{d}s.$$

5.4 Single indivisible good

Suppose there is only one potential buyer: $n = 1$. Now we can denote the single buyer's allocation by x and transfer by t, dropping the subscript i. Define θ^* by $\varphi(\theta^*) = \theta_0$. If $\theta < \theta^*$, then the buyer is not served and pays nothing. If $\theta \geq \theta^*$, then the buyer receives the good with probability one and pays

$$\hat{t}(\theta) = \theta - \int_{\theta^*}^{\theta} \Lambda(\theta, s) \, \mathrm{d}s.$$

That is, the price depends on the agent's type, so this is not longer a posted price mechanism.

With out expression for the virtual value, it is straightforward to allow for multiple players and obtain the analogous solution for auctions.

Suppose the principal has a single indivisible good. There are n buyers, labeled $i = 1, \ldots, n$. Each buyer's type $\theta_i \in \Theta_i = [\underline{\theta}_i, \bar{\theta}_i]$ is drawn independently from a cdf F_i with positive density f_i. The utility for buyer i is given by

$$u_i(q, t, \theta) = \theta_i q_i - t_i,$$

for $i = 1, \ldots, n$. The principal's utility is

$$v(q,t) = \sum_{i=1}^{n} t_i.$$

That is, the principal assigns no value to this good. This is without loss, but simplifies the notation.

Posted pricing is no longer optimal. This can be extended to auctions.

5.5 Auctions

Theorem 4 (Optimal auction)
Suppose φ_i is weakly increasing for each i. Then the seller's maximal revenue is achieved by the allocation function $q^\star$ and transfer function $t^\star$ given by

$$q_i^\star(\theta) = \begin{cases} 1 & \text{if } \varphi_i(\theta_i) > 0 \vee \max_{j \neq i} \varphi_j(\theta_j), \\ 0 & \text{otherwise,} \end{cases}$$

and

$$t_i^\star(\theta) = \theta_i Q_i^\star(\theta_i) - \int_{\underline{\theta}_i}^{\theta_i} \Lambda_i(\theta_i, s_i) Q_i^\star(s_i) \, \mathrm{d}s_i,$$

where $Q_i^\star$ is the interim expectation of $q_i^\star$:

$$Q_i^\star(\theta_i) = \int_{\Theta_{-i}} q_i^\star(\theta_i, \theta_{-i}) f_{-i}(\theta_{-i}) \, \mathrm{d}\theta_{-i}.$$

6 Discussion

6.1 Partial verification

Our model formalizes the interpretation of partially verifiable information originally suggested by Green and Laffont (1986, p. 55):

> To give the model some real content, the proper interpretation of $M(\cdot)$ is that the principal also has some information, and that the principal can act on this information, to inflict severe punishment on the agent in some circumstances .... It is not the case that the principal has an independent observation on θ. Rather, the principal can observe a binary variable whose value is (non-stochastically) jointly determined by the truth θ and the message θ', sent by the agent. Its value indicates whether or not $\theta' \in M(\theta)$.

We their model to reflect the fact that there is uncertainty about whether a misreport will be detected. Singh and Wittman (2001) use a model of deterministic partial verification, but address this issues in the conclusion and suggest fuzziness for future research. In our model no incentive constraint can be dropped. Each incentive constraint gets relaxed proportional to the probability that the corresponding type is able to mimic another type.

Partial verification has also been studied by computer scientists, beginning with Nisan and Ronen (2001) who examined distributed algorithms for shortest-path or scheduling. The computer science literature largely works within the quasi-linear setting: Fotakis and Zampetakis (2015); Auletta et al. (2011) show that it is NP-complete to determine whether a social choice function can be implemented in a non-truthful way; but truthful implementation can be recognized efficiently; Yu (2011) corrects the proof in the quasi-linear environment. Rochet (1987) and Vohra (2011) show that in the quasi-linear environments implementability of social choice function is equivalent to cyclical monotinicity.

Probabilistic verification Caragiannis et al. (2012) and Ferraioli and Ventre (2018) introduce probabilistic verification and characterize the set of implementable social choice functions in a quasi linear environment. Their assumptions about the verification technology reflect the properties of probabilistic verification algorithms, but are not explicitly microfounded. [9] Accordingly Caragiannis et al. (2012) assume that for any pair of type and report there is a probability of successful mimicking and that every type can successfully report the truth with probability one.

In contrast our reduced form mimicking relationship – the authetication rate α– allows for type I error and is not restricted to quasi-linear utilities. As a consequence, the characterization Caragiannis et al. (2012) provide does not apply to our setting. Under their assumptions the verification technology effectively degenerates to deterministic verification: Either a type can fully mimic another type or if the probability is less than one the agent can be completely deterred from mimicking, as possible fines following the detection of a lie can be chosen arbitrarily high. We allow for environments where the punishment is limited.[10] Since the agent may face the punishment with

[9] Examples for this algorithms are primality tests by Solovay and Strassen (1977) and Rabin (1980). With probability smaller than one this algorithms produces conclusive evidence that an inputted number is composite if the number is indeed composite. By constructing this algorithms never produces false evidence, mark a number as composite when it is prime.

[10] We do not impose any structure on the outcome space or restrict to a quasilinear

a low probability, the requirements on the magnitude of the punishment become greater and less realistic. But more fundamentally, the agent may experience this punishment on the equilibrium path, so it is not necessarily optimal to impose the greatest punishment.

6.2 Stochastic evidence

The literature on hard evidence (Bull and Watson (2004) and Lipman and Seppi (1995)) gives a microfoundation for the deterministic mimicking relation introduced in Green and Laffont (1986). We microfound the reduced form model with tests. But the connection to this part of the literature is deeper: Tests can be interpreted as stochastic evidence: Any test $\tau \in T$ corresponds to a piece of hard evidence. In this interpretation of tests, the principal can communicate with the agent before he learns himself what pieces of evidence are at his disposal. We assume that the agent can provide at most one piece of evidence or choose not to provide any evidence.

In a direct mechanism of this model, the principal asks the agent for a piece of evidence after a type report.

Since the provision of evidence is at the discretion of the agent a mechanism must specify an outcome for any piece of evidence provided (not only for the one that the agent was asked for) and for the case that no evidence was provided. [11]

But the commitment assumption allows us to restrict attention to mechanism where the principal implements the same outcome after she is provided with any piece of evidence she did not ask for or with no evidence at all. In this setting, free to fail corresponds to withholding evidence.

The interpretation as stochastic evidence allows us to connect our characterizations to the existing results on verification and hard evidence:

Our assumption on the existence of mostly discerning test for a type is a stochastic generalization of the assumption on the existence of maximal (or normal) pieces of evidence in Bull and Watson (2004) and Lipman and Seppi (1995). If we assume that the passage probabilities are all either zero or one, our assumption boils down to normality. Similarly the conditions of Theorem 2 are a stochastic generalization of the nested range condition in Green and Laffont (1986) and again reduces exactly to this condition if one assumes degenerate passage probabilities.

Deneckere and Severinov (2008) consider deterministic evidence but allow their revelation mechanism to be random. In such a random mechanism

setting. We view transfers, if available, as part of the outcome.

[11] Kartik and Tercieux (2012) and Strausz (2016) also discuss this issue

an agent who declares his type cannot be sure what evidence he is asked for. If he is lying, this effectively induces a probability of getting caught. By allowing for random verification technology, in an environment with most discerning test this uncertainty is independent of the mechanism, resulting in tractability.

6.3 Related literature

Lying costs Another approach to relax incentive constraints in a mechanism design framework is lying costs: Lacker and Weinberg (1989), Maggi and Rodriguéz-Clare (1995), Crocker and Morgan (1998), Kartik et al. (2007), Kartik (2009), and Deneckere and Severinov (2017). In these models, the costs of lying are exogenous. In contrast, we study the design of the mechanism to discourage lying when the agents have no inherent aversion to lying. Recently in computer science, Kephart and Conitzer (2016) studies various forms of the revelation principle. One condition that is most similar to ours is a condition that lying costs satisfy the triangle inequality.

Verification In economics "verification" usually describes models where the principal learns the true type of an agent after a certain action (paying a fee, allocating the good to particular agent). This literature was started by Townsend (1979) who studied a costly verification model for debt contracts. Ben-Porath et al. (2017) show a connection between costly verification models and evidence games. Most applications assume that monetary transfers are not feasible and use the verification technology as a substitute: Ben-Porath et al. (2014) (allocation problem), Erlanson and Kleiner (2015) (voting), Halac and Yared (2016) Optimal Delegation, Limited Awareness, and Financial Intermediation,(delegation) and Li (2017) (allocation, limited punishment). Costless verification with limited punishment is studied by Mylovanov and Zapechelnyuk (2017).

Finally, the literature on evidence games has been growing rapidly. Recent papers include Hart et al. (2017) and Koessler and Perez-Richet (2017). Our commitment assumption sets us apart from this literature, an interesting future research question is therefore how our technology would effect the outcomes of these games.

7 Conclusion

We have introduced a model of probabilistic verification that is founded on a model of tests. We define a test as a noisy signal about the type that is

chosen by the principal but can be strategically downward manipulated by the agent. Our framework can be interpreted as a generalization of hard evidence models; downward manipulations correspond to withholding evidence. We find conditions under which the optimal test choice can be disentangled from the subsequent decision problem. Under this conditions the underlying test technology can be compiled into a reduced form mimicking relation between types. We also characterize all mimicking relations between types, that can be microfounded with our framework.

The stochastic nature of the tests translates into uncertainty on the agent's side about the success of a misrepresentation of his type. This feature of our model stands in stark contrast to verification models that are based on hard evidence. It allows us to characterize the solution to standard mechanism design problems enhanced with verification by a first order approach.

This can be used to quantify the value of verification technology. We hope to use this quantification in future work to study models where the investment in verification technology is endogenous.

A Proofs

A.1 Kernels

We begin with some notation. If k is a kernel from $X \times Y$ to Z, then k_x is a kernel from Y to Z, $k_{x,y}$ is a measure on Z. In short, subscripts indicate "sections" of a function. If μ is a measure on X then μk is a kernel from Y to Z, and if ν is measure on $X \times Y$, then νk is a measure on Z. In short, measures take expectations pointwise, where a kernel from X to Y can be thought of as a function from $X \times \mathcal{Y}$.

The main proof requires the following facts about kernels, which we state without proof. Below, $\mathbf{R}$ is assumed to be equipped with its Borel σ-algebra. Similarly each subset of $\mathbf{R}$ is equipped with the induced σ-algebra. A *kernel on* $\mathbf{R}$ means a kernel from some Borel subset of $\mathbf{R}$ to $\mathbf{R}$. The domain of a such a kernel k is denoted $\operatorname{dom} k$. This domain is assumed to be large enough so that all products considered below are well-defined.

The first lemma shows that our definitions of a distribution kernel and its associated quantile kernel are suitable generalizations of the corresponding functions for continuous random variables.

Lemma 1 (Distribution kernels)**.** For measures μ and ν on $\mathbf{R}$, the following hold:

(i) $\mu \tilde{F}_\mu = U_{[0,1]}$;
(ii) $U_{[0,1]} \tilde{F}_\nu^{-1} = \nu$;
(iii) $\mu \tilde{F}_\mu \tilde{F}_\nu^{-1} = \nu$.

The notion of a downward kernel is intimately related to first-order stochastic dominance.

Definition 7. A kernel d on $\mathbf{R}$ is *downward* if $d(s, (-\infty, s]) = 1$ for all $s \in \operatorname{dom} d$.

Lemma 2 (Downward kernels)**.** For measures μ and ν on $\mathbf{R}$, the following are equivalent:

(i) $\mu \succeq_1 \nu$;
(ii) $\tilde{F}_\mu \tilde{F}_\nu^{-1}$ is downward;
(iii) $\mu d = \nu$ for some downward kernel d.

(i) If $\mu \succeq_1 \nu$, then $\tilde{F}_\mu \tilde{F}_\nu^{-1}$ is downward;
(ii) $\mu \succeq_1 \nu$ if and only if there exists a downward kernel d such that $\mu d = \nu$.

In short, downward kernels relate dominating measures to dominated measures. Next we introduce the standard notion of monotonicity for kernels.

Definition 8. A kernel m on $\mathbf{R}$ is *monotone* if $s > t$ implies $m_s \succeq_1 m_t$, for all $s, t \in \operatorname{dom} m$.

Monotone kernels *preserve* first-order stochastic dominance, as we now show.

Lemma 3 (Monotone kernels)**.**

(i) A kernel m on $\mathbf{R}$ is monotone if and only if $\mu \succeq_1 \nu$ implies $\mu m \succeq_1 \nu m$, for all $\mu, \nu \in \Delta(\operatorname{dom} m)$.

(ii) The composition of monotone kernels is monotone.

A.2 Proof of Proposition 1

Fix a profile (M, g_1, g_2, a_1, a_2) that implements a social choice function f. Define a new profile $(\hat{M}, \hat{g}_1, \hat{g}_2, \hat{a}_1, \hat{a}_2)$ as follows. Set $\hat{M} = \Theta$, $\hat{a}^1 = \mathrm{id}$, and $\hat{g}_1 = a^1 g^1$. For each $\theta \in \Theta$, set $\hat{a}^2_{\theta,\theta} = a^1_\theta a^2_\theta$ and $\hat{g}^2_\theta = a^1_\theta g^2$. The agent's off-path play can be specified arbitrarily. First, we check that this induces the same social choice rule. Evaluating at type $\theta \in \Theta$ gives

$$\begin{aligned}(\hat{a}^1_\theta \otimes \hat{g}^1 \otimes \hat{a}^2_\theta)\hat{g}^2 &= (\delta_\theta \otimes \hat{g}^1 \otimes \hat{a}^2_\theta)\hat{g}^2 \\ &= (\hat{g}_1 \otimes \hat{a}^2_{\theta,\theta})\hat{g}^2_\theta \\ &= (a^1_\theta g^1 \otimes a^1_\theta a^2_\theta)a^1_\theta g^2 \\ &= (a^1_\theta \otimes g^1 \otimes a^2_\theta)g^2.\end{aligned}$$

Now we need to check that there are no feasible deviations. Fix type θ. It suffices to check that for each deviation $(\hat{d}^1, \hat{d}^2)$ in the new mechanism, there is a deviation (d^1, d^2) that induces the same social choice function in the original mechanism. Take $d^1 = \hat{d}^1 a^1$ and $d^2_{\theta,m} = \hat{d}^2_{\theta,\theta}$ for all $\theta \in \Theta$ and $m \in M$.

A.3 Proof of Proposition 2

Fix a direct and truthful profile (M, g^1, g^2, a^1, a^2) that implements a social choice function f. Define a new profile $(\hat{M}, \hat{g}_1, \hat{g}_2, \hat{a}_1, \hat{a}_2)$ as follows. Following the theorem statement, set $\hat{M} = \Theta$, $\hat{g}^1 = g^1$, $\hat{a}^1 = \mathrm{id}$, and $\hat{a}^2_{\theta,\theta,\tau} = p_{\theta,\tau}$ for all $\theta, \theta' \in \Theta$ and $\tau \in T$.

Finally, we define $\hat{g}^2$. Fix θ and τ. The technology constraint implies that $a^2_{\theta,\theta,\tau} \preceq_1 p_{\theta,\tau}$. By Lemma 2, there exists a downward kernel $d_{\theta,\tau}$ such that $a^2_{\theta,\theta,\tau} = p_{\theta,\tau} d_{\theta,\tau}$. Set $\hat{g}^2_{\theta,\tau} = d_{\theta,\tau} g^2_{\theta,\tau}$.

A.4 Proof of Theorem 1

For part 1.

For part 2. First we prove sufficiency of our condition.

Lemma 4. Consider a type-test pair $(\theta, \bar{\tau})$ such that $\bar{\tau}$ is maximal for type $\bar{\theta}$.

(i) For each test $\tau \in T$,
$$p_{\bar{\theta},\bar{\tau}} \tilde{F}_{\bar{\theta},\bar{\tau}} \tilde{F}_{\bar{\theta},\tau}^{-1} = p_{\bar{\theta},\tau}.$$

(ii) For each type $\theta \in \Theta$, score kernel q satisfying $q \preceq_1 p_\theta$, and test $\tau \in T$,
$$q_{\bar{\tau}} \tilde{F}_{\bar{\theta},\bar{\tau}} \tilde{F}_{\bar{\theta},\tau}^{-1} \preceq_1 p_{\theta,\tau}.$$

Proof. Part (i) is immediate from the facts about kernels. For part (ii), it suffices to show
$$p_{\theta,\bar{\tau}} \tilde{F}_{\bar{\theta},\bar{\tau}} \tilde{F}_{\bar{\theta},\tau}^{-1} \preceq_1 p_{\theta,\tau}, \tag{4}$$
for then we get the desired inequality by applying Lemma 3 (i) with the kernel $\tilde{F}_{\bar{\theta},\bar{\tau}} \tilde{F}_{\bar{\theta},\tau}^{-1}$, which is monotone by Lemma 3 (ii).

To prove (4), right-apply the monotone kernel $\tilde{F}_{\bar{\theta},\tau}^{-1}$ to get
$$p_{\theta,\bar{\tau}} \tilde{F}_{\bar{\theta},\bar{\tau}} \tilde{F}_{\bar{\theta},\tau}^{-1} \preceq_1 p_{\theta,\tau} \tilde{F}_{\bar{\theta},\tau} \tilde{F}_{\bar{\theta},\tau}^{-1} \preceq_1 p_{\theta,\tau},$$
where the first inequality uses Lemma 3 (i) and the second inequality uses Lemma 2. □

Fix type $\bar{\theta} \in \Theta$ and test $\bar{\tau} \in T$ such that $\bar{\tau}$ is maximal for type $\bar{\theta}$. Consider an arbitrary incentive-compatible direct mechanism g. Define a new mechanism $\hat{g}$ that coincides with the old mechanism, except we set
$$\hat{g}_{\bar{\theta}}^1 = \delta_{\bar{\tau}}, \qquad \hat{g}_{2\bar{\theta},\bar{\theta}} = \int_T g^1(\bar{\theta}, \mathrm{d}\tau)\, \tilde{F}_{\bar{\theta},\bar{\tau}} \tilde{F}_{\bar{\theta},\tau}^{-1} g_{\bar{\theta},\tau}.$$

Suppose that in the new mechanism, the agent uses strategy $(\delta_{\bar{\theta}}, q)$. From the definition of e^*,
$$\delta_{\bar{\theta}} \otimes e^* \otimes q = \delta_{\bar{\theta}} \otimes \delta_{\bar{\tau}} \otimes q_{\bar{\tau}}.$$
Therefore, decision set A has probability
$$\begin{aligned}
\int_S q(\bar{\tau}, \mathrm{d}s)\, g^*_{\bar{\theta},\bar{\tau}}(s, A) &= \int_S q(\bar{\tau}, \mathrm{d}s) \int_\tau e(\bar{\theta}, \mathrm{d}\tau)\, (\tilde{F}_{\bar{\theta},\bar{\tau}} \tilde{F}_{\bar{\theta},\tau}^{-1} g_{\bar{\theta},\tau})(s, A) \\
&= \int_T e(\bar{\theta}, \mathrm{d}\tau) \int_S q(\bar{\tau}, \mathrm{d}s)\, (\tilde{F}_{\bar{\theta},\bar{\tau}} \tilde{F}_{\bar{\theta},\tau}^{-1} g_{\bar{\theta},\tau})(s, A) \qquad (5) \\
&= \int_T e(\bar{\theta}, \mathrm{d}\tau)\, (q_{\bar{\tau}} \tilde{F}_{\bar{\theta},\bar{\tau}} \tilde{F}_{\bar{\theta},\tau}^{-1} g_{\bar{\theta},\tau}) A,
\end{aligned}$$

where we have used Fubini's theorem and then expressed the inner integrand as a composition of kernels.

To see that the mechanism replicates the decision distribution for type $\bar{\theta}$, put $q = p_{\bar{\theta}}$ in (5) and apply Lemma 4 (i) to get

$$\begin{aligned}\int_T e(\bar{\theta}, \mathrm{d}\tau)\,(p_{\bar{\theta},\tau} g_{\bar{\theta},\tau})A &= \int_T e(\bar{\theta}, \mathrm{d}\tau) \int_S p_{\bar{\theta}}(\tau, \mathrm{d}s)\, g(\bar{\theta}, \tau, s, A) \\ &= (\delta_{\bar{\theta}} \otimes e \otimes p_{\bar{\theta}}) g A.\end{aligned}$$

To see that there are no profitable deviations, we show that this decision distribution can be achieved in the incentive-compatible mechanism (e, g) with a score kernel q' from T to S defined by

$$q'_\tau = q_{\bar{\tau}} \tilde{F}_{\bar{\theta},\bar{\tau}} \tilde{F}^{-1}_{\bar{\theta},\tau}.$$

By Lemma 4 (i), $q' \preceq_1 q$, so it must be feasible if q' is. From (5), the outcome distribution is

$$\begin{aligned}\int_T e(\bar{\theta}, \mathrm{d}\tau)\,(q'_\tau g_{\bar{\theta},\tau})A &= \int_T e(\bar{\theta}, \mathrm{d}\tau) \int_S q'(\bar{\tau}, \mathrm{d}s)\, g(\bar{\theta}, \tau, s, A) \\ &= (\delta_{\bar{\theta}} \otimes e \otimes q') g A.\end{aligned}$$

Next, we prove necessity. Suppose

A.5 Proof of Theorem 2

First we characterize the discerning order in terms of the primitive $p(\theta, \tau)$.

Lemma 5. Test $\tau^\star$ is more discerning for type $\theta^\star$ than test τ if and only if

$$\begin{aligned}&(1 - p(\theta^\star, \tau^\star))(1 - p(\theta, \tau)) \leq (1 - p(\theta^\star, \tau))(1 - p(\theta, \tau^\star)) \\ \text{and } &p(\theta^\star, \tau) p(\theta, \tau^\star) \leq p(\theta^\star, \tau^\star) p(\theta, \tau)\end{aligned}$$

Proof. Denote the cdf of $P_{\theta,\tau} \tilde{F}_{\theta^*,\tau}$ by G. Calculating yields,

$$G(q) = \begin{cases} \frac{1-p(\theta,\tau)}{1-p(\theta^\star,\tau)} q, & 0 \leq q < 1 - p(\theta^\star, \tau) \\ 1 - p(\theta, \tau) + \frac{p(\theta,\tau)}{p(\theta^\star,\tau)}[q - (1 - p(\theta^\star, \tau))], & 1 - p(\theta^\star, \tau) \leq q \leq 1 \end{cases}$$

i.e. the piecewise linear function, whose graph connects $(0, 0)$, $(1-p(\theta^\star, \tau), 1-p(\theta, \tau))$ and $(1, 1)$

We identify any function of this form with a tuple (x, y), such that the graph of the function linearly interpolates the points $(0, 0), (x, y)$ and $(1, 1)$. Graphically a function $G_{(x',y')}$ dominates $G_{(x,y)}$, if (x', y') lies above the curve of $G_{x,y}$, this yields,

$$\forall q:\ G_{(x,y)}(q) \leq G_{(x',y')}(q) \Leftrightarrow x'y \leq xy' \text{ and} (1-x)(1-y') \leq (1-y)(1-x')$$

A test τ^* is more discerning for type θ^* than test τ, if for all θ,

$$\begin{aligned}
&P_{\theta,\tau}\tilde{F}_{\theta^*,\tau} \geq_1 P_{\theta,\tau^*}\tilde{F}_{\theta^*,\tau} \\
\Leftrightarrow &\forall q \in [0,1]:\ G_{(1-p(\theta^*,\tau),1-p(\theta,\tau))}(q) \leq G_{(1-p(\theta^*,\tau^*),1-p(\theta,\tau^*))}(q) \\
\Leftrightarrow &(1-p(\theta^\star,\tau^\star))(1-p(\theta,\tau)) \leq (1-p(\theta^\star,\tau))(1-p(\theta,\tau^\star)) \\
\text{and } &p(\theta^\star,\tau)p(\theta,\tau^\star) \leq p(\theta^\star,\tau^\star)p(\theta,\tau).
\end{aligned}$$

□

Now we turn to the main proof of the theorem. Suppose α is induced by mostly discerning tests, then transitivity of α and intransitivity of $1-\alpha$ follow from Lemma 5.

If on the other hand α is transitive and $1-\alpha$ is intransitive, setting $T = \Theta$ with passage probabilities $p(\theta, \theta') = \alpha(\theta, \theta')$ for all types θ and all tests θ' forms a test environment which has mostly discerning tests (again by Lemma 5 for any type and induces α as reduced form.

A.6 Proof of Proposition 3

Fix a type $\theta_0 \in \Theta$. Let τ and ψ be tests that are both θ_0-mostly discerning. By assumption, the test are informative, so we may pick a type θ such that $p(\theta, \tau) \neq p(\theta_0, \tau)$. By Lemma 5, we have

$$\begin{aligned}
p(\theta_0,\tau)p(\theta,\psi) &= p(\theta,\tau)p(\theta_0,\psi), \\
(1-p(\theta_0,\tau))(1-p(\theta,\psi)) &= (1-p(\theta,\tau))(1-p(\theta_0,\psi)).
\end{aligned}$$

Combining these inequalities gives

$$\begin{aligned}
p(\theta_0,\psi)(p(\theta,\tau)-p(\theta,\tau)) &= p(\theta_0,\tau)(p(\theta,\tau)-p(\theta,\tau)), \\
p(\theta,\psi)(p(\theta,\tau)-p(\theta_0,\tau)) &= p(\theta,\tau)(p(\theta,\tau)-p(\theta_0,\tau)),
\end{aligned}$$

so $p(\theta_0, \psi) = p(\theta_0, \tau)$ and $p(\theta, \psi) = p(\theta, \tau)$. This argument goes through for teach type θ unless $p(\theta, \tau) = p(\theta_0, \tau)$ and $p(\theta, \psi) = (\theta_0, \psi)$. But in this lass case, we have $p(\theta, \psi) = p(\theta_0, \psi) = p(\theta_0, \tau) = p(\theta, \tau)$, so we are done.

A.7 Proof of Theorem 3

First, we introduce notation. For a given quantity function q, there is a one-to-one correspondence between a transfer function t and a utility functions U, given by $U(\theta) = \theta q(\theta) - t(\theta)$. We will interchangeably refer to such a mechanism as (q, t) or (q, U).[12] Let

$$u(\theta, \theta') = \alpha(\theta, \theta')[\theta q(\theta') - t(\theta')].$$

In particular, $U(\theta) = u(\theta, \theta)$.

Lemma 6 (Utility bound)**.** Let q be a bounded quantity function. If (q, U) is a feasible mechanism, then

$$U(\theta) \geq \int_{\underline{\theta}}^{\theta} \alpha(\theta, s) q(s) \, \mathrm{d}s,$$

for each type θ. Moreover, if q is monotone, then there exists a feasible mechanism with quantity q that achieves this bound pointwise.

Now we prove the theorem, taking the lemma as given. There is no loss in restricting attention to bounded quantity functions.[13] Pick a bounded quantity function $q\colon \Theta \to \mathbf{R}_+$. The principal's objective function can be decomposed as the difference between the total surplus and the agent's rents:

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

Plug in the bound from Lemma 6 and switch the order of integration to obtain the following upper bound on the principal's objective:

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

The quantity function $q^\star$ from the theorem statement maximizes the expression in brackets pointwise, and $t^\star$ is the corresponding transfer that

[12] This is a slight abuse of notation, but the symbol t or U will make it clear whether the second argument is a transfer function or a utility function.

[13] There are no explicit bounds on the quantity function q, so first we must justify the restriction to bounded quantity functions. Pick a quantity $\bar{q}$ such that $\bar{\theta}\bar{q} = c(\bar{q})$. Then offering more than $\bar{q}$ will always result in weakly negative profits, so we can remove those offerings from the menu and increase the sender's revenue. Therefore, the is no loss in focusing on quantity functions that are bounded above by $\bar{q}$.

achieves the utility bound. Since $q^\star$ is monotone, this mechanism is feasible and hence optimal.

Now we turn to the proof of the lemma. First we bound the right-derivative of U, wherever it exists. If U is right-differentiable at θ, then by Theorem 1 in Milgrom and Segal (2002),

$$U'(\theta+) \geq \partial_1 u(\theta, \theta) = q(\theta) - \lambda(\theta)U(\theta). \tag{6}$$

To make use of (6), we need to check that U is absolutely continuous. Even though $u(\theta, \theta')$ has a kink, is derivative exists almost everywhere, a slight adaption of Theorem 2 in Milgrom and Segal (2002) show shows that U is absolutely continuous. Therefore, U is differentiable almost everywhere, so

$$U(\theta) = \int_{\underline{\theta}}^{\theta} U'(s+)\,\mathrm{d}s.$$

Now apply (6) and the integral form of Gronwall's inequality to obtain the desired inequality.

It remains to verify incentive compatibility. We need to show that for all types θ and reports θ',

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

or equivalently,

$$U(\theta) - \alpha(\theta, \theta')U(\theta') \geq (\theta - \theta')\alpha(\theta, \theta')q(\theta'). \tag{7}$$

We separate into two cases.

If $\theta > \theta'$, the inequality reduces to

$$\int_{\theta'}^{\theta} \alpha(\theta, s)q(s)\,\mathrm{d}s \geq (\theta - \theta')\alpha(\theta, \theta')q(\theta'). \tag{8}$$

If $\theta < \theta'$, then the left side of (7) is at least

$$-\alpha(\theta, \theta') \int_{\theta}^{\theta'} \alpha(\theta', s)q(s)\,\mathrm{d}s,$$

so a sufficient inequality is

$$\int_{\theta}^{\theta'} \alpha(\theta', s)q(s)\,\mathrm{d}s \leq (\theta' - \theta)q(\theta'). \tag{9}$$

If q is monotone, then (8) and (9) both hold, completing the proof.

References

Auletta, V., P. Penna, G. Persiano, and C. Ventre (2011): "Alternatives to Truthfulness are Hard to Recognize," *Autonomous Agents and Multi-Agent Systems*, 22, 200–216.

Ben-Porath, E., E. Dekel, and B. L. Lipman (2014): "Optimal Allocation with Costly Verification," *American Economic Review*, 104, 3779–3813.

Ben-Porath, E., E. Dekel, and B. L. Lipman (2017): "Mechanisms with Evidence: Commitment and Robustness," Working paper.

Border, K. C. and J. Sobel (1987): "Samurai Accountant: A Theory of Auditing and Plunder," *Review of Economic Studies*, 54, 525–540.

Bull, J. and J. Watson (2004): "Evidence Disclosure and Verifiability," *Journal of Economic Theory*, 118, 1–31.

——— (2007): "Hard Evidence and Mechanism Design," *Games and Economic Behavior*, 58, 75–93.

Caragiannis, I., E. Elkind, M. Szegedy, and L. Yu (2012): "Mechanism design: from partial to probabilistic verification," in *Proceedings of the 13th ACM Conference on Electronic Commerce*, ACM, 266–283.

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

Deneckere, R. and S. Severinov (2008): "Mechanism Design with Partial State Verifiability," *Games and Economic Behavior*, 64, 487–513.

——— (2017): "Screening, Signalling and Costly Misrepresentation," Working paper.

Erlanson, A. and A. Kleiner (2015): "Costly Verification in Collective Decisions," .

Ferraioli, D. and C. Ventre (2018): "Probabilistic Verification for Obviously Strategyproof Mechanisms," *ArXiv e-prints*.

Fotakis, D. and E. Zampetakis (2015): "Truthfulness Flooded Domains and the Power of Verification for Mechanism Design," *ACM Transactions on Economics and Computation*, 3, 20:1–29.

Green, J. R. and J.-J. Laffont (1986): "Partially Verifiable Information and Mechanism Design," *Review of Economic Studies*, 53, 447–456.

Halac, M. and P. Yared (2016): "Commitment vs. flexibility with costly verification," Tech. rep., National Bureau of Economic Research.

Hart, S., I. Kremer, and M. Perry (2017): "Evidence Games: Truth and Commitment," *American Economic Review*, 107, 690–713.

Kartik, N. (2009): "Strategic Communication with Lying Costs," *Review of Economic Studies*, 76, 1359–1395.

Kartik, N., M. Ottaviani, and F. Squintani (2007): "Credulity, Lies, and Costly Talk," *Journal of Economic Theory*, 134, 93–116.

Kartik, N. and O. Tercieux (2012): "Implementation with Evidence," *Theoretical Economics*, 7, 323–355.

Kephart, A. and V. Conitzer (2016): "The Revelation Principle for Mechanism Design with Reporting Costs," in *Proceedings of the 2016 ACM Conference on Economics and Computation*, 85–102.

Koessler, F. and E. Perez-Richet (2017): "Evidence Reading Mechanisms," Working paper.

Lacker, J. M. and J. A. Weinberg (1989): "Optimal Contracts under Costly State Falsification," *Journal of Political Economy*, 97, 1345–1363.

Li, Y. (2017): "Mechanism design with costly verification and limited punishments," .

Lipman, B. L. and D. J. Seppi (1995): "Robust Inference in Communication Games with Partial Provability," *Journal of Economic Theory*, 66, 370–405.

Maggi, G. and A. Rodriguéz-Clare (1995): "Costly Distortion of Information in Agency Problems," *RAND Journal of Economics*, 26, 675–689.

Milgrom, P. and I. Segal (2002): "Envelope theorems for arbitrary choice sets," *Econometrica*, 70, 583–601.

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

Mylovanov, T. and A. Zapechelnyuk (2017): "Optimal Allocation with Ex Post Verification and Limited Penalties," *American Economic Review*, 107, 2666–2694.

Nisan, N. and A. Ronen (2001): "Algorithmic Mechanism Design," *Games and Economic Behavior*, 35, 166–196.

Rabin, M. O. (1980): "Probabilistic algorithm for testing primality," *Journal of number theory*, 12, 128–138.

Rochet, J.-C. (1987): "A Necessary and Sufficient Condition for Rationalizability in a Quasi-Linear Context," *Journal of Mathematical Economics*, 16, 191–200.

Singh, N. and D. Wittman (2001): "Implementation with Partial Verification," *Review of Economic Design*, 6, 63–84.

Solovay, R. and V. Strassen (1977): "A fast Monte-Carlo test for primality," *SIAM journal on Computing*, 6, 84–85.

Strausz, R. (2016): "Mechanism Design with Partially Verifiable Information," Cowles Foundation Discussion Paper No. 2040.

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

Vohra, R. V. (2011): *Mechanism Design: A Linear Programming Approach*, Econometric Society Monographs, Cambridge University Press.

Yu, L. (2011): "Mechanism Design with Partial Verification and Revelation Principle," *Autonomous Agents and Multi-Agent Systems*, 22, 200–216.