Strategic Communication with Minimal Verification*

Gabriel Carroll† **Georgy Egorov**‡

September 2017

Abstract

A receiver wants to learn multidimensional information from a sender, but she has capacity to verify only one dimension. The sender's payoff depends on the belief he induces, via an exogenously given monotone function. We show that by using a randomized verification strategy, the receiver can learn the sender's information fully if the exogenous payoff function is submodular. If it is (strictly) supermodular, then full learning is not possible. In a variant of the model that allows for severe punishments when the sender is found to have lied, we can give a complete characterization of when full learning is possible. Our full learning result does not critically rely on perfect verifiability of one dimension: in an example with noisy verification, the receiver's ex-post perceived distribution of information converges in distribution to the true value as the noise vanishes.

Keywords: Mechanism design, multidimensional information, verifiability, cheap talk

JEL Codes: D82, D83

*We are grateful to Sandeep Baliga, Vincent Crawford, Xavier Gabaix, Johannes Hörner, Navin Kartik, Paul Milgrom, Roger Myerson, Ran Spiegler, and seminar participants at the University of Pittsburgh and conference participants at Tel Aviv University and Transatlantic Theory Workshop in Paris for helpful comments. Authors are listed in random order.

†Stanford University. E-mail: gdc@stanford.edu

‡Kellogg School of Management, Northwestern University, and NBER. E-mail: g-egorov@kellogg.northwestern.edu

1 Introduction

An HR manager is interviewing a job candidate to form an opinion about the candidate's qualities or skills. A prosecutor is interviewing a defendant to decide whether there is a case that she could prosecute. An insurance company employee is evaluating a claim filed by its client to decide if it is legitimate or fraudulent. All these cases can be thought of as an interaction between a sender and a receiver of information, where the former tries to impress the latter, while the latter tries to make a possibly precise inference about the former's private information.

This interaction is unlikely to be pure cheap talk. The HR manager can give the job candidate a test, or can call the college that the candidate lists on his vita to verify truthfulness of the claim. The prosecutor can compare the defendant's statements with evidence obtained otherwise. The insurance company employee can visit the client's property and inspect what was damaged or stolen. However, the verification might be limited: the HR manager might be able to test only a few skills, the prosecutor might be able to corroborate only some of the defendant's claims, and the insurance company might verify only some of the information to ensure speedy processing. To what extent does this limited verification prevent the receiver from fully learning the sender's information?

In this paper, we make a strong assumption on the limits to verification: the sender's type is multidimensional, and the receiver is only able to verify at most one dimension. But the dimension she verifies can depend on the message that the sender sends. What can she do in this context?

The following example previews our ideas.

Example 1. An IT firm is hiring a programmer, and wants to evaluate a job candidate on two dimensions: math skills and coding skills. The candidate knows his skills x and y, but from the firm's perspective, they are i.i.d. uniform on $[0, 1]$. The candidate tries to impress the firm by signaling that the sum of his two skills, $x + y$, is as high as possible, because, for example, this value is linked to the probability of being hired or to the expected salary.

If the firm is not able to verify either dimension, then clearly no useful information about the total value $x + y$ can be credibly transmitted in equilibrium. At the other extreme, if the firm can test both skills, it can learn the candidate's type perfectly. Our question is what can happen if the firm can test just one skill.

Suppose first that the firm chooses in advance which skill to test. If it chooses math, then it learns the value x precisely, but does not get any information about y. Conversely, if it chooses coding, it learns y but gets no update on x.

It is easy to see that the firm can improve by asking the candidate to choose which test he would like to take. The candidate who is better at math ($x > y$) would then ask to take the math test, and the candidate better at coding would ask for the coding test. Then, after giving the math test to the candidate who chose it, the firm not only learns x, but also some information on y, namely, that y is distributed on $[0, x]$, and similarly the firm that ends up giving the coding test to the candidate learns something about his math skills. (Notice that it is incentive-compatible for the candidate to report his best dimension: the firm's posterior expectation of $x + y$ would be $\frac{3}{2}x$ if he asks for the math test and $\frac{3}{2}y$ if he asks for the coding test, so indeed he prefers the former if and only if $x > y$.)

Is there any way the firm can learn even more?

The answer may be a surprise: the firm can learn everything, by using a randomized mechanism. This can be achieved as follows. The firm asks the candidate to report $p = \frac{x}{x+y}$, and then proceeds by giving the candidate the math test with probability p and the coding test with probability $1 - p$. If the candidate plays along, then the firm will indeed achieve full learning: after giving the math test (which is possible only if $p > 0$) and observing x, it would infer y as $y = \frac{1-p}{p}x$; similarly, after giving the coding test and observing y, it would infer x as $x = \frac{p}{1-p}y$. It therefore remains to verify that it is incentive compatible for the candidate to report $p = \frac{x}{x+y}$ truthfully.

A candidate that reports $p = \frac{x}{x+y}$ truthfully makes the firm learn his true $x + y$. A candidate that deviates and reports $\hat{p}$ instead makes the firm believe that $\widehat{x+y} = x + \frac{1-\hat{p}}{\hat{p}}x$ if he gets the math test, and that $\widehat{x+y} = \frac{\hat{p}}{1-\hat{p}}y + y$ if he gets the coding test. Since he gets the former with probability $\hat{p}$ and the latter with probability $1 - \hat{p}$, in expectation he makes the impression

$$\hat{p}\left(x + \frac{1-\hat{p}}{\hat{p}}x\right) + (1-\hat{p})\left(\frac{\hat{p}}{1-\hat{p}}y + y\right) = x + y.$$

This means that the candidate cannot gain by misreporting, and indeed the firm can learn everything if it gives the candidate the freedom to choose the testing probabilities.

In what follows, we study how far the simple logic of this example generalizes. We build a model with two economic agents, a sender (he) and a receiver (she), which we can think of as a job candidate and an interviewer. The sender has multidimensional private information (e.g., his skills) and can send a message to the receiver, who can subsequently verify the value of one of the dimensions. In most of the paper we assume this verification is perfect; in a final section we consider a case with noisy verification. We

think of the receiver's problem as one of mechanism design: she commits to a verification rule so that equilibrium play in the resulting communication game will reveal as much information about the private type as possible. We assume that while the receiver is free to design the verification rule, she has no control over the subsequent (unmodeled) actions that will generate payoffs for the sender. This assumption is natural: e.g., in the case of an interviewer and a job candidate, the interviewer might be obliged to write a truthful (according to her posterior belief) report about the job candidate to her supervisor, so she may design a mechanism that selectively elicits only some information, but she cannot manipulate the candidate's payoffs in any other way. The candidate, in his turn, tries to maximize the overall impression of the receiver (e.g., her posterior belief about the sum of his skills).

The sender's gain from convincing the receiver that his type is a is modeled by an exogenous function $V(a)$ (in the above example, this is the sum of coordinates). We can think of this function as a reduced-form way of modeling the outcome of any subsequent interaction between the sender and receiver. We study how the possibility or impossibility of perfect learning depends on the function $V(a)$. If $V(a)$ is additively separable in the coordinates, the example generalizes in a straightforward way. Moreover, the mechanism is essentially unique. More generally, as long as $V(a)$ is submodular, full learning is possible, whereas if $V(a)$ is strictly supermodular then it is impossible. In the particular case of two dimensions, we can give a complete characterization of when full learning is possible. We then consider a variant model where a sender who has been caught lying can be punished as if his type were zero (e.g. the employer simply refuses to hire the candidate if the test reveals that he lied). In this case, we can strengthen the full learning result for submodular V by giving an explicit description of the mechanism (instead of just an existence proof), and we can also give a full necessary and sufficient condition on V for full learning to be possible. Finally, as a robustness check, we consider an example with noisy verification. While full learning is not achievable with any nonzero amount of noise, we show how it can be achieved in the limit as the level of noise goes to zero.

Our paper contributes to the large literature on strategic information transmission and communication that starts with Crawford and Sobel (1982) and Holmström (1977), and more specifically to transmission of multidimensional information (see Sobel, 2013, for an extensive review of the literature on strategic communication). In the cheap talk framework where no information is verifiable, Chakraborty and Harbaugh (2007) show that some information, in particular, relative statements about the dimensions of interest, may be transmitted. Chakraborty and Harbaugh (2010) further show that in the linear case, information on all but one dimension (the "dimensions of agreement") may be trans-

mitted.[1] In a similar setting but with multiple senders, Battaglini (2002) demonstrates that full learning is possible for a generic set of parameters.[2]

The paper that is the most related to ours is Glazer and Rubinstein (2004), which also studies a receiver ('listener') who is trying to elicit multidimensional information from the sender ('speaker') and is able to verify at most one dimension.[3] In that paper, the receiver uses the information learnt to make a binary decision, e.g. whether to hire the sender or not, and the sender has a constant preference over decisions, e.g. always prefers to be hired. In our terms this corresponds to assuming that V can take two values. The receiver wishes to minimize the probability of a mistake. The authors show that the optimal mechanism exists and furthermore can be characterized as a solution to a linear programming problem defined in the paper, and random mechanisms may be necessary to achieve the optimum. Moreover, the ex-ante optimal mechanism has an indirect implementation for which it is optimal for the receiver to follow through once information is acquired. In contrast to their paper, we consider a broader range of payoffs for the sender, but focus on the possibility of full learning, which is not discussed in Glazer and Rubinstein (2004); in their setting, if full learning were possible, it would of course be optimal.[4]

Our paper is also linked to the growing literature on mechanism design with costly verification, started by Townsend (1979). Recent work in this line includes Kartik and Tercieux (2012), which studies a general problem of implementation in the presence of costly signals. Ben-Porath, Dekel, and Lipman (2014) consider the problem of optimal allocation of an indivisible unit in an auction-type environment, where agents' claims may be verified at a cost. Deb and Stewart (2017) study the problem of adaptive testing where the number and identity of tests depend on the information revealed by the agent.

[1]The reader may notice that our example above has a similar feature: one dimension out of two is transmitted before any verification, and verification fills in the missing dimension. The analogy, however, ends here. First, in the example, the sender's report may actually transmit some information about the sender's overall quality (depending on the distribution of types), so the probability of getting a particular test is by no means a "dimension of agreement." Second, we show our results on full learning in a much more general case than linear.

[2]Other papers highlighting the possibility of full learning with multiple senders include Ambrus and Takahashi (2008) and Meyer, Moreno de Barreda, and Nafziger (2016). Ambrus and Lu (2014) show that almost fully revealing cheap talk is possible when the multiple senders are imperfectly informed; Section 5 of our paper makes a similar robustness case in our setting.

[3]Glazer and Rubinstein (2004) also mention a number of examples, such a judge who can subpoena only one witness and needs to choose; or an investigator who can verify details of only one procedure performed by a doctor who is accused of malpractice. These examples relate to our paper as well.

[4]Other papers studying communication of multidimensional information include Austen-Smith and Fryer (2005) who study a model where signaling communicates two dimensions of an agent's type ('economic' and 'social'), and Polborn and Yi (2006) and Egorov (2015), where each politician chooses a topic to campaign on to impress the voters.

The rest of the paper proceeds as follows. In Section 2, we set up the framework and define the notion of a valid mechanism. Section 3 analyzes the model, and shows that full learning is possible when the sender's reward function is submodular but not when it is supermodular. Then, in Section 4 we consider a variation in which senders who are caught lying can be more severely punished; in this case we can give stronger results, including a complete characterization of when full learning is possible. In Section 5 we show that our results do not critically depend on the assumption that verification is precise, by providing an example where almost full learning is achieved with noisy verification. Section 6 concludes.

2 Setup

There are two agents, whom we call the *sender* and the *receiver*. The sender has multi-dimensional private information, which we call his *type* and denote $a = (a_1, \ldots, a_n) \in A$, where $A = [0, \infty)^n$ is the space of possible types. This type follows a prior distribution $\Phi \in \Delta(A)$.

After the sender and receiver interact, the receiver will be left with some (possibly probabilistic) belief $\mu \in \Delta(A)$ concerning the sender's type. We take as given a function $V : \Delta(A) \to \mathbb{R}$; $V(\mu)$ denotes the payoff that the sender gets if he induces belief μ. In particular, for a type $a \in A$, we write $V(a)$ for the payoff the sender gets if he induces a belief that is a point mass on a.[5] For instance, in the job candidate example, $V(a)$ could represent the salary that the candidate will receive if he convinces the interviewer that his type is a (perhaps this is simply the marginal product for the firm of a worker of type a). In the prosecution example, $V(a)$ would denote the probability that the prosecutor drops the case. More generally, we have in mind a signaling-game-like situation in which, after learning, the receiver takes some action that generates a payoff for the sender; but we have no need to model this action explicitly, so instead we summarize it with the function $V(\cdot)$.

When the sender communicates with the receiver, he faces uncertainty over what belief μ will be induced: in particular, if the receiver plans to verify a randomly chosen dimension, μ may depend on which dimension is verified. We assume that $V(\cdot)$ is a von Neumann-Morgenstern utility function, so that the sender acts to maximize the expectation of $V(\mu)$. We assume throughout that V is weakly increasing: if μ and μ' are two distributions on A, and μ first-order stochastically dominates μ', then $V(\mu) \geq V(\mu')$. We

[5] Although we define $V(\mu)$ for arbitrary beliefs, it will soon become clear that we mostly need to concern ourselves only with degenerate beliefs.

also normalize $V(0) = 0$ (hereinafter, we will use 0 to denote the null vector when it does not cause confusion).

The sender and the receiver can engage in a strategic interaction with the following structure: The sender can transmit a message. The receiver can then verify up to one component of the sender's type, if she wishes. For most of the paper, we assume that verification of dimension i makes the receiver perfectly informed about a_i. (This will be relaxed in Section 5, where the receiver will get a noisy signal s; for now s will denote the perfect signal about a_i obtained by the receiver.) The receiver can commit in advance to the verification strategy, but has no control over the post-verification interaction and thus simply takes as given the function $V(\cdot)$. We will also, in line with the mechanism design tradition, assume that the receiver can choose an equilibrium of the ensuing game.

We assume that the receiver simply wishes to learn as much as possible. In particular, our interest is in whether there exists a way for the receiver to learn the sender's type perfectly in equilibrium.

Formally, the object chosen by the receiver — describing both the game in which the sender and receiver interact, and the equilibrium thereof — is a *mechanism*, a tuple $\mathcal{M} = (M, \sigma, p, \mu)$, where:

- M is a message space;
- $\sigma : A \to \Delta(M)$ is a (possibly mixed) reporting strategy for the sender;
- p is a (possibly mixed) verification strategy for the receiver, specifying probabilities $(p_0(m), \ldots, p_n(m))$ that sum to 1, for each $m \in M$ (here $p_i(m)$ is the probability of verifying dimension i, for $i \geq 1$, and $p_0(m)$ is the probability of no verification);
- μ is a belief system for the receiver, specifying posteriors $\mu(h) \in \Delta(A)$ for each $h \in H$, where $H = (M \times \{0\}) \cup (M \times \{1, \ldots, n\} \times [0, \infty))$ is the set of possible *histories* (for the receiver).

Note that the receiver's beliefs are defined as functions of the history; a history may be of the form $(m, 0)$ (meaning that message m was sent and there was no verification) or (m, i, s) (message m was sent, dimension i was verified and the value observed was s). We assume that once the belief μ is induced, the sender receives a payoff equal to $V(\mu)$.

We say that the mechanism is a *direct mechanism* if the sender just reports his type truthfully: $M = A$, and $\sigma(a) = a$ (deterministically) for each a.

We say that a mechanism $\mathcal{M} = (M, \sigma, p, \mu)$ is *valid* if it satisfies the following three conditions:

- *Incentive compatibility (for sender)*: For each $a \in A$, $\sigma(a)$ has its support contained in the set of $m \in M$ that maximize

$$p_0(m)V(\mu(m,0)) + \sum_{i=1}^{n} p_i(m)V(\mu(m,i,a_i)).$$

- *Bayesian updating*: Let $\zeta(h)$ denote the equilibrium probability measure over the set of *full histories* $\bar{H} = A \times H$, where a full history specifies both the sender's true type and the interaction with the receiver. Then, for any measurable set of full histories $\bar{H}' \subset \bar{H}$ and any measurable set of types $A' \subset A$, $\int_{\bar{H}'} \mu(h)(A')\, d\zeta(h) = \int_{\bar{H}'} \mathbf{1}_{a \in A'}\, d\zeta(h)$.

- *Trusted verification*: For any history $h = (m, i, s)$, the belief $\mu(h)$ puts no probability on types with $a_i \neq s$.

The trusted verification condition effectively serves to restrict off-path beliefs: it says that if the sender sends message m, yet verification reveals that he was a type who should not have sent m in equilibrium, the receiver should infer that the sender deviated, not that the verification went wrong.

The receiver thus chooses a valid mechanism, with the goal of learning as much as possible about a. In particular, we are interested in the existence of mechanisms that allow for *full learning*: for every type a, at every history $h \in H(a|\mathcal{M})$, the belief $\mu(h)$ is degenerate on type a, where we define

$$\begin{aligned} H(a|\mathcal{M}) \;=\; & \{(m,0) : m \in \text{supp}(\sigma(a)) \text{ and } p_0(m) > 0\} \\ & \cup \{(m,i,a_i) : m \in \text{supp}(\sigma(a)) \text{ and } p_i(m) > 0\}, \end{aligned}$$

the set of histories that can arise when the sender has type a.[6]

Because we are interested only in full learning (until Section 5), we can simplify in several ways. First, a version of the revelation principle applies:

Lemma 0. *If there exists a valid mechanism with full learning, then there exists a valid direct mechanism with full learning.*

[6] In some applications, we may think the receiver is content to learn the value of $V(a)$ without learning a itself: e.g. the employer may be interested in knowing the worker's total output, but not how it is achieved. While learning $V(a)$ may appear to be a simpler problem than learning a, in fact it is not: if there is a valid mechanism that allows the receiver to learn $V(a)$, there is also one that achieves full learning. For a formal statement and proof, see Proposition A3 in the Appendix.

The proof is in the Appendix.

With this result in mind, we can focus on direct mechanisms. Since in this case $M = A$ and σ is an identity mapping, the description of a direct mechanism consists only of the verification probabilities and the beliefs. Moreover, beliefs are pinned down at the on-path histories, i.e. those in $H(a|\mathcal{M})$: the belief must be degenerate on type a. At "off-path" histories, in particular (a, i, s) with $s \neq a_i$, the receiver may have any beliefs, as long as they are consistent with the revealed value of s (the trusted verification condition). However, since we are interested in understanding whether it is possible to incentivize truthful reporting, it is sufficient to consider "punishment" beliefs that lead to the lowest possible payoff for the sender at such histories. Since V is increasing, this means that at such an off-path history where s was observed, the receiver should simply believe the sender's type is equal to $\tilde{a}$ with $\tilde{a}_i = s$ and $\tilde{a}_j = 0$ for all $j \neq i$ (and thus confer reward $V(\tilde{a})$ on the sender).

The upshot of this discussion is that, for purposes of studying whether full learning is possible, we can focus on the choice of verification strategy. Full learning is possible if and only if there exists a choice of verification probabilities (for a direct mechanism), $p = (p_0, \ldots, p_n)$ with each $p_i : A \to [0, 1]$ and $\sum_i p_i(a) = 1$ for each a, satisfying the following incentive compatibility condition for all types a and $\hat{a}$:

$$V(a) \geq p_0(\hat{a})V(\hat{a}) + \sum_{i=1}^{n} p_i(\hat{a})w_i(\hat{a}|a), \quad \text{where} \quad w_i(\hat{a}|a) = \begin{cases} V(\hat{a}) & \text{if} \quad \hat{a}_i = a_i, \\ V(a|_i) & \text{if} \quad \hat{a}_i \neq a_i, \end{cases} \tag{1}$$

where $a|_i$ denotes the type whose i-coordinate is a_i and other coordinates are all zero.[7] Indeed, here the left side of the inequality represents the payoff that the sender gets from truthfully reporting type a, which will be $V(a)$ no matter which dimension is verified; and the right side is the expected payoff from reporting $\hat{a}$, given the punishment beliefs.

A few remarks, before moving on:

- The above formulation in terms of direct mechanisms backs up our claim in Footnote 5, that only the values of V at degenerate beliefs matter. So we will henceforth think of V as being defined only on A, instead of on $\Delta(A)$. Then the monotonicity requirement just says that if $a' \leq a$, then $V(a') \leq V(a)$ (hereinafter, we use $\leq$ to denote the componentwise partial order).

- The incentive compatibility conditions are invariant under translating the function $V(\cdot)$ by a constant. Thus it is indeed just a normalization to assume that $V(0) = 0$.

[7] In line with this notation, we will use $1|_i$ to denote the unit vector in the ith direction.

- Although the prior Φ was needed for the general setup (to formulate posteriors), note that the existence or nonexistence of a full-learning mechanism does not depend on the prior at all. We will need it in Section 5, however.

3 Analysis

We start by analyzing the case of additively separable payoff functions; this includes the linear case as in the example from the Introduction. Here we can give a construction that directly generalizes that example. We then study the cases of submodular and supermodular payoff functions. The main results are that full learning is possible when the payoff function is submodular, but not when it is supermodular. The Section finishes with a complete characterization of when full learning is achievable in the special case $n = 2$.

For a rough intuition about why the submodular versus supermodular distinction arises, think about the job candidate with two possible skills, as in Example 1. A candidate who is strong on one skill but weak on the other has a potential incentive to pretend to be strong on both. This can be deterred if the weak skill is verified with sufficiently high probability. But if the skills are complements (supermodular case), the gains from appearing to be strong on both skills rather than just one are high, and there is no way to choose verification probabilities to deter both a (strong math, weak coding) candidate and a (weak math, strong coding) candidate. In contrast, if the skills are substitutes (submodular case), the gains are smaller and this can be done. Another way to put is as follows. In the submodular case (e.g., if the quality equals the maximum of the dimensions), the strongest skill is informative about the candidate's overall quality. In this case, the goals of learning the quality and deterring deviations are aligned, and are achievable by putting a high probability on the stronger dimension. However, in the supermodular case (e.g., if the quality equals the minimum of the dimensions), it is the weakest skill that is informative of the quality, and the goals of learning and deterring deviations are at odds.

3.1 Additively separable case

Suppose $V(a)$ is additively separable in its components, so

$$V(a) = \sum_{i=1}^{n} v_i(a_i), \tag{2}$$

where $v_i : [0, \infty) \to \mathbb{R}$ are increasing functions. Since we assumed $V(0) = 0$, we may pick $v_i(\cdot)$ such that $v_i(0) = 0$ for each i.

Our first result establishes a mechanism that achieves full learning.

Proposition 1. *Suppose that V is additively separable and defined by (2). Then there exists a valid mechanism with full learning. Moreover, this can be accomplished by a direct mechanism using the verification probabilities*

$$p_i(m) = \frac{v_i(m_i)}{V(m)} = \frac{v_i(m_i)}{v_1(m_1) + \cdots + v_n(m_n)} \ (1 \leq i \leq n), \qquad p_0(m) = 0$$

for each m such that $V(m) \neq 0$ (and arbitrary verification probabilities for m such that $V(m) = 0$, in particular $m = (0, \ldots, 0)$).

The argument straightforwardly extends the example in the Introduction.

Proof. We just need to check that incentive compatibility (1) is satisfied. Note that $w_i(\hat{a}|a) = V(\hat{a})$ if $\hat{a}_i = a_i$, and $v_i(a_i)$ otherwise; hence $p_i(\hat{a})w_i(\hat{a}|a) \leq v_i(a_i)$, with equality if $\hat{a}_i = a_i$. Since $p_0(\hat{a})V(\hat{a}) = 0$ for all $\hat{a}$, condition (1) immediately follows. □

The mechanism suggested in Proposition 1 has several remarkable properties. To state them, assume for simplicity that each v_i is strictly increasing, and in particular $V(a) = 0 \Leftrightarrow a = (0, \ldots, 0)$.

- The mechanism can be implemented as an indirect mechanism, as in the Introduction, where the sender chooses a probability distribution $(p_1, \ldots, p_n)$ over dimensions to verify (so M is an $n-1$-dimensional simplex of probabilities). When dimension i is verified and the observed value is s, the receiver infers $v_j(a_j) = \frac{p_j}{p_i} v_i(s)$ for each j, and so infers a completely by inverting each v_j.

- The mechanism also does not actually require the receiver to commit to the verification strategy, as we have assumed. Indeed, if she could freely choose which component to verify, note that once she has heard message m, she expects to end up believing (with probability 1) that the sender is type m (and to give reward $V(m)$) regardless of which component she verifies, so she is indifferent at this stage.

- As mentioned above at the end of Section 2, the mechanism does not depend on the distribution of sender's type Φ. Moreover, it would perform just as fine if the receiver had a wrong belief about Φ. Implementing this mechanism therefore requires the receiver to know the payoff function $V(\cdot)$ and nothing else.

- In the case where all v_i are linear, the indirect implementation highlights that the parties do not need to agree on the "scale" in which the type is measured, i.e. it works even if the sender perceives his type as $(\lambda a_1, \ldots, \lambda a_n)$ rather than $(a_1, \ldots, a_n)$, for an arbitrary positive scalar λ.

Our second result is that the mechanism is essentially unique.

Proposition 2. *Assume that V is additively separable and defined by (2). Let $\mathcal{M} = (M, \sigma, p, \mu)$ be a valid (possibly indirect) mechanism with full learning. Then for any type $a \in A$ with $V(a) > 0$, for any $m \in \text{supp}(\sigma(a))$, we have*

$$p_i(m) = \frac{v_i(a_i)}{v_1(a_1) + \cdots + v_n(a_n)} \ (1 \leq i \leq n), \qquad p_0(m) = 0.$$

Proof. Consider an alternative type a' that agrees with a in all coordinates except in coordinate i, where $a'_i = 0$. Let m be any message in the support of $\sigma(a)$.

The assumption of full learning implies that, if type a' sends message m and coordinate i is not verified, then the resulting belief places probability 1 on type a, and the sender gets reward $V(a)$. Hence, the expected payoff to sending message m is at least $(1-p_i(m))V(a)$. So incentive compatibility for the pair of types a' and a implies

$$V(a') \geq (1 - p_i(m))V(a),$$

which implies

$$p_i(m) \geq \frac{V(a) - V(a')}{V(a)} = \frac{v_i(a_i)}{v_1(a_1) + \cdots + v_n(a_n)}.$$

Since we must also have $\sum_{i=1}^{n} p_i(m) \leq 1$, the inequalities for $p_i(m)$ must hold as equalities, and also $p_0(m) = 0$. □

3.2 Submodular payoff functions

The previous section showed that full learning is possible when V is additively separable. This can be generalized considerably: Full learning is achievable as long as V is submodular.

Proposition 3. *Suppose that V is submodular. Then there exists a valid mechanism that achieves full learning.*

The full proof of Proposition 3 is in the Appendix, but it is worthwhile to give a sketch here. We are looking for a direct mechanism; our goal is to find verification probabilities

$p_i(a)$ for $a \in A$ such that all incentive constraints are satisfied. For any given $a \in A$, we need to choose $p_i(a)$ so that no type $z \neq a$ wants to deviate by reporting type a; in making this choice, we can ignore the incentives of type a to deviate to other types. This contrasts with the typical situation in mechanism design, where the allocation to a particular type affects both that type's own incentives to misreport and other types' incentives to claim to be that type. Here, since we are interested in a mechanism that achieves full learning, the equilibrium payoff for any type a is fixed at $V(a)$, and we are only left with discouraging deviations by other types to a.

With that in mind, we fix $a \in A$ and let $p_0(a) = 0$. Take any vector p of verification probabilities following report a; for any type $z \in A$ we define $G_p(z)$ as the gain in payoff for type z by misreporting as type a rather than reporting truthfully. If we can find some p such that $G_p(z) \leq 0$ for every z, then our mission is accomplished. We also focus just on types z with $z \leq a$; this turns out to be enough (intuitively, smaller types have stronger incentives to misreport).

So assume, to obtain a contradiction, that no matter which p we take, there will always be some z with $G_p(z) > 0$. If so, we consider the set of "maximal deviators" $D_p = \arg\max_{z \leq a} G_p(z)$; this set is nonempty if V is continuous (the full proof makes adjustments that handle discontinuous V as well). Now let E_p be the set of alternative verification probabilities q such that $G_q(z) \leq 0$ for all $z \in D_p$. That is, E_p is the set of probability vectors that would deter deviations by all the types that were most inclined to deviate under p. If it turns out that $p \in E_p$, we would get an immediate contradiction, as the maximal deviations in D_p should at least be profitable under p. The problem is therefore reduced to showing that the set-valued mapping $p \mapsto E_p$ has a fixed point.

It is easy to check that E_p is compact and convex. It is also nonempty: From the submodularity of V, we can show that D_p is a lattice, and its largest element cannot be a (because if $a \in D_p$ then a has the highest incentive to misreport as a, but this incentive is zero, so we are done). Hence there exists some coordinate i in which $z_i < a_i$ strictly for all $z \in D_p$, and then verifying dimension i with probability 1 would successfully discourage all types in D_p from misreporting. Unfortunately, E might not be upper-hemicontinuous, and thus we cannot immediately apply Kakutani's theorem to show existence of a fixed point. However, in the Appendix we "smooth out" E_p by constructing an auxiliary set $Q_p \subset E_p$ for which Kakutani's theorem applies. Any fixed point of mapping Q is also a fixed point of mapping E, which yields the desired contradiction.

While the proof of Proposition 3 is nonconstructive, in the case of two dimensions we can give an explicit description. Specifically, the following is true.

Proposition 4. *Suppose that V is submodular and $n = 2$. Then the following valid direct mechanism achieves full learning: upon receiving message m, the receiver verifies dimensions 1 and 2 with probabilities*

$$p_1(m) = \frac{V(m) - V(m|_2)}{2V(m) - V(m|_1) - V(m|_2)}, \quad p_2(m) = \frac{V(m) - V(m|_1)}{2V(m) - V(m|_1) - V(m|_2)},$$

unless $V(m) = V(m|_1) = V(m|_2)$, in which case these probabilities $p_1(m)$ and $p_2(m)$ may be chosen arbitrarily; in all cases, we take $p_0(m) = 0$.

Proof. We need to show that no type z wants to misreport as type a. If both $z_1 \neq a_1$ and $z_2 \neq a_2$, then the lie will be detected for sure; in particular, if dimension i is verified, the sender's payoff is $V(z|_i) \leq V(z)$, which implies that he would be better off reporting truthfully. So we can assume that z and a differ in only one dimension. Without loss of generality, consider the case $z_1 = a_1$, $z_2 \neq a_2$.

It is clear that the deviation is not profitable if $z_2 > a_2$, because then $z \geq a$ coordinatewise, so $V(z) \geq V(a)$ and there is no gain for type z to misreport as type a. Thus, consider the case $z_2 < a_2$. Then the gain from deviation by type z to a equals

$$\begin{aligned} G(a|z) &= p_1(a)V(a) + p_2(a)V(z|_2) - V(z) \\ &= p_1(a)(V(a) - V(z)) + p_2(a)(V(z|_2) - V(z)) \\ &\leq p_1(a)(V(a) - V(a|_1)) - p_2(a)(V(z) - V(z|_2)) \\ &\leq p_1(a)(V(a) - V(a|_1)) - p_2(a)(V(a) - V(a|_2)) = 0, \end{aligned}$$

where the first inequality follows from monotonicity ($V(z) \geq V(z|_1) = V(a|_1)$) and the second from submodularity. Thus the gain is nonpositive for any a and z, which completes the proof. □

3.3 Supermodular payoff functions

We next observe that for some $V(a)$ there does not exist a mechanism that achieves full learning. Consider the following example.

Example 2. Suppose $n = 2$, and $V(a_1, a_2) = a_1 a_2$. Let us prove that there is no valid mechanism that achieves full learning. Consider the type $a = (1, 1)$, for which $V(a) = 1$. When the sender reports this type, the first dimension must be tested with probability 1: otherwise, the sender of type $(0, 1)$ could report $(1, 1)$ and have a positive probability of being undetected, so would get a positive expected payoff, better than the 0 he gets

by telling the truth. Similarly, for a report of $(1,1)$, the second dimension must also be tested with probability 1. Since only one dimension can be tested, this is a contradiction.

This example generalizes greatly:

Proposition 5. *Suppose that there is a type a such that*

$$V(a) > \sum_{i=1}^{n} V(a|_i). \tag{3}$$

Then there does not exist a valid mechanism that achieves full learning.

Note that if V is strictly supermodular (and $V(0) = 0$ as we have assumed), then the condition in the proposition is satisfied for *any* type a that is positive in every coordinate. So, the proposition covers such functions (but is also much more general).

Proof. Take type a for which the inequality holds. Let p be the vector of probabilities that this type a gets tested on each dimension. For each $i \in \{1, \ldots, n\}$, consider a sender with type $a|_i$. If he chooses his equilibrium action, his payoff will be $V(a|_i)$. If, however, he misreports as type a, then with at least probability $p_0 + p_i$ the deviation will go undetected. In equilibrium, such deviation cannot be profitable, which implies

$$V(a|_i) \geq (p_0 + p_i)V(a).$$

Adding these inequalities for each i, we get a contradiction to (3). This contradiction completes the proof. □

3.4 Concave transformations

Another interesting property is that any mechanism that achieves full learning is robust to concave transformations of the sender's payoff function:

Proposition 6. *Let V be such that full learning is achievable in a valid direct mechanism $\mathcal{M}$. Then, the same mechanism $\mathcal{M}$ also achieves full learning when the payoff function is $V' = U \circ V$, where $U : [0, \infty) \to [0, \infty)$ is any increasing, concave transformation.*

The proof is in the Appendix. Essentially, the result holds because when a mechanism achieves full learning, the sender is certain of his payoff along the equilibrium path, whereas by deviating he gets a lottery over payoffs. Concave transformations make such a lottery even less desirable.

Concave transformations can arise naturally in two ways. First, if V is the monetary payoff that the sender receives (for example, if he is a job candidate who is paid his perceived marginal product), then U can represent risk aversion. Thus, the proposition says that any mechanism that achieves full learning for a risk-neutral sender also works when the sender is risk-averse. Second, V might represent value measured in some abstract units, and U can represent decreasing returns. For example, if the job candidate's "total skill" is $a_1 + \cdots + a_n$, the proposition says that any mechanism that works when the candidate's marginal product equals his total skill also works when there are decreasing returns to total skill.

3.5 Full characterization for $n = 2$

Propositions 1, 3, and 5 above imply that full learning is possible for additively separable or submodular, but not supermodular, payoff functions. In the two-dimensional case, we can provide a necessary and sufficient condition on V for full learning to be achievable.

Proposition 7. *Suppose that $n = 2$. Then full learning is achievable with a valid mechanism if and only if V satisfies the following property: for any two types $x, a \in A$ with $x < a$, we have*

$$\left(V(a) - V(x_1, a_2)\right)\left(V(a) - V(a_1, x_2)\right) \leq \left(V(x_1, a_2) - V(x|_1)\right)\left(V(a_1, x_2) - V(x|_2)\right).$$

The proof is in the Appendix, but the idea behind it is simple, in the light of the ideas discussed earlier. There two types of deviations that the mechanism needs to prevent: from x to a where $x_1 < a_1$ and $x_2 = a_2$, and from x to a where $x_1 = a_1$ and $x_2 < a_2$. To prevent the first type of deviations, we need the probability that the first dimension is checked, $p_1(a)$, to be sufficiently high. To prevent the second type, we similarly need $p_2(a)$ to be sufficiently high. The criterion ensures that these conditions are compatible.

4 Punishment for lying

In the definition of a valid mechanism, we proposed the "trusted verification" requirement, that if the sender's message and the verification device come into conflict, the receiver's belief should respect the result of the verification — and the sender's payoff should reflect this. However, in some applications, it is plausible to think that it is possible to use different payoffs at these off-path histories. For example, in the employment application, we might simply imagine that the company might refuse to hire a candidate who has been

caught lying, even though he still has some skills; in the insurance claim example, the insurance company might not be obligated to honor any claims if it has shown that some claim was false.

For this reason, in this section, we drop the trusted verification requirement from the definition of a valid mechanism; but we still assume that the sender's payoff cannot be made lower than the payoff of the worst possible type, $V(0) = 0$. We call the set of mechanisms defined in this way *valid mechanisms with punishment*. As usual, we can restrict attention to mechanisms that deliver the worst possible punishment if a misreport is detected, which is 0 in our case. Thus (continuing to focus on direct mechanisms), the incentive compatibility condition simplifies to

$$V(a) \geq \left(p_0(\hat{a}) + \sum_{i:\ \hat{a}_i = a_i} p_i(\hat{a}) \right) V(\hat{a}). \tag{4}$$

For this case of valid mechanisms with punishment, we can obtain tighter results, as follows. First, the results from Section 3 (except for the characterization for $n = 2$ in Proposition 7) continue to hold. Second, we can give a full characterization of which functions $V(\cdot)$ allow for full learning: Proposition 9 provides a necessary and sufficient condition for any n. Third, in the case of submodular payoff functions, we can now give an explicit construction of a valid mechanism with punishment that achieves full learning for any n as well; this is shown in Proposition 10.

4.1 Robustness of main results

Proposition 8. *Without the trusted verification requirement (i.e. in the case of valid mechanisms with punishment), Propositions 1–6 are correct as stated.*

Proof. This statement is obvious for Propositions 1, 3, 4, and 6. For the uniqueness result (Proposition 2) and impossibility of full learning for supermodular payoff functions (Proposition 5), examining the proofs reveals that they go through even when punishment with 0 payoff is possible. □

4.2 Full characterization

We now provide a necessary and sufficient condition for full learning to be achievable. Following our earlier $a|_i$ notation, whenever $S \subset \{1, \ldots, n\}$ is a set of indices and $a \in A$, define $a|_S$ as the type that agrees with a on the components $i \in S$, and whose other coordinates are all zero.

Proposition 9. *There exists a valid mechanism with punishment that achieves full learning if and only if V satisfies the following condition. For every $a \in A$, and any collection of nonnegative weights λ_S for each of the 2^n sets $S \subset \{1, \ldots, n\}$ that satisfies $\sum_{S:i \in S} \lambda_S = 1$ for each index $i = 1, \ldots, n$, we have*

$$V(a) \leq \sum_{S \subset \{1,\ldots,n\}} \lambda_S V(a|_S).$$

The proof is in the Appendix.

To see why a full characterization is possible, consider what happens to the incentive condition (4) when we hold fixed the report $\hat{a}$, and also hold fixed the coordinates a_i of the true type for which $a_i = \hat{a}_i$, but vary the other coordinates a_j. Then the right side of (4) is constant, while the left side is increasing in a. Consequently, the constraint is tightest when $a = \hat{a}|_S$ for some set S: if we can deter these types a from reporting $\hat{a}$, then all other types are deterred as well. So full learning is achievable as long as we can choose the verification probabilities for each type a to deter misreporting by the (finitely many) types $a|_S$. The proposition gives a duality-based characterization of when this is possible.

Using this proposition, we can readily check that there are functions $V(a)$ that allow full learning using a valid mechanism with punishment, but not in the original setting where trusted verification limits the possible punishments. Here is an example.

Example 3. Let $n = 2$, and let $V(a)$ be a two-dimensional step function as given by the following table. (By perturbation one could also come up with an example that is continuous and strictly increasing.)

$a_2 \in (1, \infty)$	8	9	12
$a_2 \in (0, 1]$	7	8	9
$a_2 = 0$	0	7	8
	$a_1 = 0$	$a_1 \in (0, 1]$	$a_1 \in (1, \infty)$

This function satisfies the condition of Proposition 9, but not Proposition 7; thus it allows full learning when we allow lies to be punished by zero payoff, but not in the original model.[8] One can also see this directly: to implement when zero punishments are possible, simply specify that a report that is positive on both dimensions should have each dimension tested with probability 1/2; if it is positive only on one dimension, test that dimension with certainty. On the other hand, when zero punishments are not possible,

[8]The condition in Proposition 9 takes a particularly simple form if $n = 2$: the only possible weights are of the form $\lambda_{\{1\}} = \lambda_{\{2\}} = \lambda$ and $\lambda_{\{1,2\}} = 1 - \lambda$, and the condition simplifies to $V(a) \leq V(a|_1) + V(a|_2)$.

consider the report $(2,2)$. Some dimension must be tested with probability at most $1/2$, say dimension 1. Then type $(1,2)$ has the incentive to misreport as $(2,2)$, since even if his lie is caught (which has probability at most $1/2$) he still gets a payoff of 7, and otherwise he gets a payoff of 12, for a total payoff of not less than $19/2 > 9$.

4.3 Explicit construction for submodular functions

Now we give an explicit construction for a mechanism that works when V is submodular. Again following on our $a|_S$ notation, for each $a \in A$ and each $i = 1, \ldots, n$, let $a|_{[i]}$ be the type whose first i components agree with a, and whose remaining $n - i$ components are all zero. Consistently with this, let also $a|_{[0]} = (0, \ldots, 0)$.

Proposition 10. *Suppose that V is submodular. Then the following valid mechanism with punishments achieves full learning: If $V(m) > 0$, dimension i is verified with probability*

$$p_i(m) = \frac{V(m|_{[i]}) - V(m|_{[i-1]})}{V(m)}, \qquad p_0(m) = 0,$$

and if $V(m) = 0$, the probabilities are chosen arbitrarily.

The proof is in the Appendix; it is a straightforward application of a construction used more generally in the proof of Proposition 9.

5 Imperfect verification

So far, we have assumed that if the receiver chooses to verify dimension i of the sender's private information (type), she observes the exact value of a_i. With a direct mechanism in use, this means that a sender can be severely punished if he is caught lying: if he claims (say) type $(3,5)$, but component 1 is verified and found to equal 2.99, he can be punished with the payoff $V(2.99, 0)$. So a natural question is whether our results are robust to noisy verification: If there is a small amount of noise, is it still possible to learn the sender's type almost perfectly? Our answer to this robustness concern is twofold. First, Example 1 presented an indirect version of the mechanism where all histories are on-path; this already shows that our main results are not artifacts of discontinuities in off-path beliefs. Second, below we consider an extension where a small amount of noise is introduced explicitly, and show formally that almost full learning is possible in the limit, by a suitably perturbed version of the original indirect mechanism.

More specifically, assume that if dimension i is verified, the receiver gets a noisy signal $s \in [0, \infty)$, drawn from a full-support distribution $\rho_i(s|a_i)$. The original definition of a valid (indirect) mechanism extends readily to this case, with incentive compatibility appropriately formulated by taking expectations over the possible signals. Note that the trusted verification condition no longer applies, since every signal realization is possible for any type.

In this section, we show how near-full learning can indeed be achieved with small noise, at least in the canonical example. We make some specific distributional assumptions. We define Φ, the distribution of sender's types, as follows. Let $(\nu_1, \ldots, \nu_n)$ be a vector of any real numbers, $(\tau_1, \ldots, \tau_n)$ be a vector of positive numbers (we let $\tau = \sum_i \tau_i$), and let K be a positive constant. Consider an auxiliary distribution on $[0, \infty)^n$, defined by generating a random type z as follows: each z_i is lognormal, with $\log z_i \sim \mathcal{N}\left(\nu_i, \frac{1}{\tau_i}\right)$ (so ν_i is the mean and τ_i is the precision), and the dimensions $\{z_i\}_{i \in \{1,\ldots,n\}}$ are independent. Refer to this distribution of z as Λ.

Let A_K be the cone defined by inequalities

$$A_K = \left\{ a : \frac{a_i}{\sum_{j=1}^n a_j} \geq \frac{\tau_i}{\tau + K} \text{ for each } i \right\}.$$

Notice that for any $K > 0$, this set is nonempty and becomes the entire positive octant A as $K \to \infty$. Now, we define distribution Φ as Λ, restricted on the cone A_K. (The usefulness of this restriction will be evident from the result below, where it will simplify the description of the mechanism by ensuring positiveness of the probabilities involved.)

We now define $\rho_i(s|a_i)$, the conditional distribution of the signal s that the receiver gets if she verifies dimension i of a sender with type a, as lognormal with $\log s \sim \mathcal{N}\left(\log a_i, \frac{1}{\chi}\right)$ (and if $a_i = 0$, then $s = 0$ for sure). This is equivalent to assuming that $s = a_i \eta$, where η is multiplicative noise (independent from a) such that $\log \eta \sim \mathcal{N}\left(0, \frac{1}{\chi}\right)$. Here χ is a parameter governing informativeness of the signal. As $\chi \to \infty$ the signal becomes perfectly informative.

Since the receiver will now typically have non-degenerate posterior beliefs, we need to return to having V be a function of the belief. For simplicity we assume that V depends only on the posterior mean, and is linear: $V(\mu) = \mathbb{E}_{a \sim \mu}\left[\sum_{i=1}^n a_i\right]$.

In this setting, it is easy to see that a direct mechanism would not be able to achieve full learning. Indeed, if the sender reports his type truthfully and any signal s can occur in equilibrium if dimension i is verified, then the Bayesian property implies that the sender's report $\hat{a}$ should always be trusted, which obviously creates room for manipulation. It is

only slightly less obvious that an indirect mechanism cannot achieve full learning either: again, since any signal realization is possible, the message m alone must be sufficient to infer the sender's type in equilibrium, leading to manipulation as before. In this sense, the results that guarantee full learning do not immediately extend to the case with noisy verification. However, below we prove that "almost full" learning is achievable as the variance of noise goes to zero, at least with our distributional assumptions.

To state this result formally, recall that for any mechanism $\mathcal{M} = (M, \sigma, p, \mu)$, $\mu = \mu(h)$ is defined as the posterior distribution of a conditional on history h. Let $\kappa = \kappa(a)$ be the probability measure that "aggregates" $\mu(h)$ over all possible histories that type a can generate. Formally, $\kappa(a)$ on any measurable $A' \subset A_K$ is given by

$$\kappa(a)(A') = \mathbb{E}_{m\sim\sigma(a)}\left[p_0(m)\mu(m,0)(A') + \sum_{i=1}^{n} p_i(m)\int_0^\infty \mu(m,i,s)(A')\,d\rho_i(s|a_i)\right].$$

(In the mechanisms we consider below, the strategy σ will be deterministic, so the expectation operator on the right becomes unnecessary.) We show below that, essentially, $\kappa = \kappa(a)$ converges to a for all $a \in A_K$ as the noise disappears.

Proposition 11. *There exist a set of valid mechanisms $\{\mathcal{M}^\chi\}_{\chi>K}$, where $\mathcal{M}^\chi = (M^\chi, \sigma^\chi, p^\chi, \mu^\chi)$, such that for any $a \in A_K$, the corresponding probability measure $\kappa^\chi(a)$ converges in distribution to an atom on a as $\chi \to \infty$. In other words, for any $a \in A_K$ and any $\varepsilon, \delta > 0$ there is $\chi_{\varepsilon,\delta,a}$ such that for any $\chi > \chi_{\varepsilon,\delta,a}$, the probability $\Pr_{x\sim\kappa^\chi(a)}(\max_i |x_i - a_i| > \varepsilon) < \delta$.*

Furthermore, we can take mechanism $\mathcal{M}^\chi$ such that M^χ is the unit simplex of probabilities restricted to the cone A_K; the reporting strategy $\sigma^\chi(a)$ of any type $a \in A_K$ prescribes him to report his "relative skills" $\left\{\frac{a_i}{\sum_{j=1}^n a_j}\right\}_{i\in\{1,\dots,n\}}$ with probability 1, and the receiver, after getting message m, verifies dimension $i \in \{1,\dots,n\}$ with probability $p_i^\chi(m) = m_i\left(1+\frac{\tau}{\chi}\right) - \frac{\tau_i}{\chi}$, with $p_0^\chi(m) = 0$.

The proof is in the Appendix, but some features of the mechanisms $\{\mathcal{M}^\chi\}_{\chi>K}$ are worth discussing. As in Example 1, the mechanism includes the sender communicating his relative skills to the receiver. Importantly, the total ability $V(a) = \sum_{i=1}^n a_i$ is not communicated, and thus the receiver uses the signal s to make conclusions about the "scale." The receiver tests dimension i with probabilities $p_i^\chi(m)$, which are all positive on A_K, provided that $\chi > K$. As before, a higher sender's report on dimension i increases the likelihood of it being tested, which is intuitive as then the potential gain from deviation is higher. In addition, and this is a new feature, dimension i is more likely to be tested if τ_i is low, i.e., if the variance $\frac{1}{\tau_i}$ of the prior distribution of a_i is high. This feature is

intuitive: with higher variance of a_i, the sender's message has a significant influence on the receiver's posterior, and it is important to discourage the sender from exaggerating that dimension, which is achieved by testing it with a high probability. In contrast, if the variance of a_i is low, then different types largely generate the same distribution of signals when dimension i is tested, which means that testing this dimension is of little use in providing incentives for truthful reporting. In the extreme, with zero variance of a_i, its value is perfectly known to the receiver, and testing that dimension clearly is not useful.[9]

We do not claim that the mechanisms constructed, $\{\mathcal{M}^\chi\}$, are optimal in any sense on their own; what is important is that these are valid mechanisms that achieve convergence of the posterior distributions. In other words, we showed that it is possible to approximate the mechanism highlighted in the Introduction such that full learning is achieved in the limit as verification becomes more and more precise, at least under some distributional assumptions. This demonstrates that the possibility of eliciting multidimensional information with verification of only one dimension is not an artifact of perfect verification and discontinuous punishment.

6 Conclusion

We considered the problem of strategic transmission of multidimensional information between a sender and a receiver, where the receiver is able to verify at most one dimension. If the receiver chooses this dimension without any input from the sender, she learns just that dimension, at least if dimensions are uncorrelated. An obvious improvement is to ask the sender which dimension to test; in this case, the receiver perfectly learns that dimension, and the sender's choice reveals some information about the other dimensions as well. The main contribution of our paper is showing that if we take this logic just one step further and allow the sender to choose among randomized tests, the receiver may learn the sender's type fully, for a wide range of the sender's objective functions. Importantly for us, full learning does not arise as an artifact of discontinuities resulting from perfect verification; we show that with small noise, full learning may be achieved in the limit, at least in our example where we obtained a closed-form solution.

While the paper's main contribution is theoretical, we believe it has practical takeaways. In the additively separable case, our mechanism has an intuitive implementation, with the sender offering probabilities of being tested in each dimension and the receiver

[9] These particular mechanisms $\{\mathcal{M}^\chi\}$ remain valid and achieve full learning in the limit for some other payoff functions, in particular, for $V(\mu) = (\mathbb{E}_{a\sim\mu}[\sum_{i=1}^n a_i])^\gamma$ for $\gamma \in (0,1)$. In other words, almost full learning is possible in this case even if the sender is risk-averse rather than risk-neutral (this complements Proposition 6). This fact is proved in the Appendix (Proposition A2).

giving the tests with exactly these probabilities. Such a mechanism is, in our view, sufficiently natural to be of practical use. For example, it is quite common for an interviewer to ask the job candidate to describe a project (or, in an academic context, a paper) that he listed on the vita, with the understanding that the candidate will proceed with the best one. But the candidate may instead offer the interviewer to make the selection, or at least suggest a couple to choose from, or he may even suggest a few and try to nudge the interviewer towards one or the other. Clearly, this communicates additional information about the candidate's willingness to talk about each project, which is very much in line with the spirit of the proposed mechanism.

The paper's results suggest many interesting directions for further inquiry. A natural question is how much information may be transmitted if full learning is impossible, for example, if the sender's payoff is supermodular, e.g. a Leontief function of his skills. Alternatively, suppose that even offering one test is costly (as in e.g. Ben-Porath, Dekel, and Lipman, 2014); then a direct mechanism would create commitment problems for the receiver, who would not want to verify ex post, but the indirect mechanism, where probabilities are communicated but not the scale, would not (at least for small cost). Actually, in this case, if full learning is achievable, it might not be optimal, since the receiver could economize by not testing over a small range of types close to zero; a natural question is what an optimal mechanism looks like. As another possible application, consider a professor who wants to test her students on multiple topics. In this example, running our proposed mechanism would consist of asking students to report their relative skills and then administering a test with just one (randomly determined) question. This might not be desirable, either because any single question reveals too noisy a signal, or because the students may not know their relative skills perfectly. Here, the natural solution is to offer several problems instead of one, which in turn poses the problem of the optimal number of questions an exam should have, and how to choose their topics for each student in an optimal way (see also Deb and Stewart, 2017, who study a similar question with a one-dimensional type space).

In all these extensions, a departure from the goal of full learning prevents the results of our paper from being directly applicable, but some insights and intuitions we have obtained might be nevertheless useful. Studying the properties of optimal learning mechanisms in cases where full learning is not achievable or desirable could be an important future direction.

References

[1] Ambrus, Attila, and Shih En Lu (2014), "Almost fully revealing cheap talk with imperfectly informed senders," *Games and Economic Behavior*, 88: 174–189.

[2] Ambrus, Attila, and Satoru Takahashi (2008), "Multi-sender cheap talk with restricted state spaces," *Theoretical Economics*, 3(1): 1-27.

[3] Austen-Smith, David, and Roland G. Fryer Jr., "An Economic Analysis of 'Acting White,'" *Quarterly Journal of Economics*, 120(2): 551–583.

[4] Battaglini, Marco (2002), "Multiple Referrals and Multidimensional Cheap Talk," *Econometrica*, 70: 1379–1401.

[5] Ben-Porath, Elchanan, Eddie Dekel, and Barton L. Lipman (2014), "Optimal allocation with costly verification," *American Economic Review*, 104(12): 3779–3813.

[6] Chakraborty, Archishman, and Rick Harbaugh (2007), "Comparative Cheap Talk," *Journal of Economic Theory*, 132(1): 70–94.

[7] Chakraborty, Archishman, and Rick Harbaugh (2010) "Persuasion by Cheap Talk," *American Economic Review*, 100(5): 2361–2382.

[8] Deb, Rahul, and Colin Stewart (2017), "Optimal Adaptive Testing: Informativeness and Incentives," unpublished paper.

[9] Egorov, Georgy (2015), "Single-issue Campaigns and Multidimensional Politics," NBER working paper No. w21265.

[10] Glazer, Jacob, and Ariel Rubinstein (2004), "On Optimal Rules of Persuasion," *Econometrica*, 72(6): 1715–1736.

[11] Holmström, Bengt (1977), "On Incentives and Control in Organizations," Ph.D. Thesis, Stanford University.

[12] Kartik, Navin, and Olivier Tercieux (2012), "Implementation with Evidence," *Theoretical Economics*, 7(2): 323–355.

[13] Meyer, Margaret, Inés Moreno de Barreda, and Julia Nafziger (2016), "Robustness of Full Revelation in Multisender Cheap Talk," unpublished paper.

[14] Polborn, Mattias K., and David T. Yi (2006), "Informative Positive and Negative Campaigning," *Quarterly Journal of Political Science*, 1(4): 351–71.

[15] Sobel, Joel (2013), "Giving and Receiving Advice," In Advances in Economics and Econometrics, edited by Daron Acemoglu, Manuel Arellano, and Eddie Dekel. Vol. 1. (Cambridge: Cambridge University Press): 305–341.

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

Appendix

The Appendix contains the proofs of all results that are not proved in the main text.

Proof of Lemma 0. Let $\mathcal{M} = (M, \sigma, p, \mu)$ be a valid mechanism that achieves full learning. We wish to construct p', μ' that (together with the message space $M' = A$ and the truthful reporting strategy $\sigma'(a) = a$) form a valid direct mechanism $\mathcal{M}'$ that achieves full learning. Let $p'_i(a) = \mathbb{E}_{m\sim\sigma(a)}[p_i(m)]$, the expected probability with which dimension i is verified for type a in the original mechanism. To define the beliefs, we proceed as suggested in the text: at histories of the form $h = (a, 0)$ or (a, i, a_i), $\mu'(h)$ puts probability 1 on type a; at other histories (a, i, s), it puts probability 1 on type $(s_i, 0_{-i})$.

It is immediate from the definition that the mechanism satisfies full learning. To see that the mechanism is valid, we check the conditions one by one. For incentive compatibility, notice that if type a reports truthfully he gets a payoff of $V(a)$, whereas by reporting $\hat{a}$ he gets a payoff

$$\begin{aligned} p'_0(\hat{a})V(\hat{a}) + \sum_{i=1}^{n} p'_i(\hat{a})V(\mu'(\hat{a}, i, a_i)) &= \mathbb{E}_{m\sim\sigma(\hat{a})}\left[p_0(m)V(\hat{a}) + \sum_{i=1}^{n} p_i(m)V(\mu'(\hat{a}, i, a_i))\right] \\ &\leq \mathbb{E}_{m\sim\sigma(\hat{a})}\left[p_0(m)V(\mu(m, 0)) + \sum_{i=1}^{n} p_i(m)V(\mu(m, i, a_i))\right] \\ &\leq V(a). \end{aligned}$$

Here the first inequality follows from the fact that for each $m \in \text{supp}(\sigma(\hat{a}))$, if $p_0(m) > 0$ then $V(\mu(m, 0)) = V(\hat{a})$ by full learning in the original mechanism; and likewise, for $i \geq 1$, if $p_i(m) > 0$ then either $\hat{a}_i = a_i$ implying $V(\mu(m, i, a_i)) = V(\hat{a}) = V(\mu'(\hat{a}, i, a_i))$ by full learning, or $\hat{a}_i \neq a_i$ and $V(\mu'(\hat{a}, i, a_i)) = V(a|_i) \leq V(\mu(m, i, a_i))$ by trusted verification and monotonicity of V. The second inequality comes from incentive compatibility of the original mechanism.

Now, Bayesian updating for the new mechanism follows from the definition of beliefs, since in equilibrium, with ex ante probability 1, the receiver puts probability 1 on the true type. Lastly, trusted verification holds by construction. □

Proof of Proposition 3. The proof is non-constructive. For each type a, we show that there exist corresponding verification probabilities that deter any other type from misreporting as type a. By doing this for every a, we form an incentive compatible mechanism.

So fix a type a henceforth. Consider any particular verification probabilities $p =$

$(p_1, \ldots, p_n)$ that sum to 1. (We focus on probabilities that definitely verify one dimension, i.e. the probability of no test is zero.) Notice that the function

$$U_p(a|z) = \sum_{i=1}^{n} p_i(z) w_i(a|z), \quad \text{where} \quad w_i(a|z) = \begin{cases} V(a) & \text{if } z_i = a_i \\ V(z|_i) & \text{if } z_i \neq a_i \end{cases}$$

is additively separable in the components of z. Therefore, the gain to type z from misreporting as a,

$$G_p(z) = U_p(a|z) - V(z),$$

is supermodular in z.

Notice first that we can reduce the problem to showing existence of p such that $G_p(z) \leq 0$ for all $z \leq a$. Indeed, suppose that this is true, but there is some $x \not\leq a$ with $G_p(x) > 0$. Then supermodularity of $G_p(\cdot)$ implies that $G_p(a \wedge x) + G_p(a \vee x) \geq G_p(a) + G_p(x) > 0$, since $G_p(a) = 0$ (here, $\wedge, \vee$ are the componentwise min and max operations). But $a \wedge x \leq a$, which by assertion satisfies $G_p(a \wedge x) \leq 0$; and $G_p(a \vee x) \leq 0$ because $a \vee x \geq a$ implies $V(a \vee x) \geq V(a)$, so type $a \vee x$ cannot gain from the deviation. Contradiction.

Hereinafter, we consider $z \in B = \{z \in A : z \leq a\}$, and use Δ to denote the $(n-1)$-dimensional unit simplex. Suppose, to obtain a contradiction, that for every $p \in \Delta$ there is $z \in B$ such that $G_p(z) > 0$.

For each $p \in \Delta$, let $l_p = \sup_{z \in \Delta} G_p(z)$. We then have $l_p > 0$ for all p, and since $G_p(z)$ is a continuous function of p for any fixed z (moreover, it is Lipschitz continuous with coefficient $V(a)$), then l_p is also a continuous function of p. Now, for any p, let

$$D_p = \left\{ z \in B : G_p(z) > \frac{n+1}{n+2} l_p \right\};$$

in other words, D_p is the set of z such that the gain from deviation to a is sufficiently close to the supremum. By definition of l_p, $D_p \neq \varnothing$ for all p.

For each $i \in \{1, \ldots, n\}$, define

$$R_i = \{p \in \Delta : \exists z \in D_p : z_i = a_i\}.$$

Let us show that $R_i \neq \varnothing$ for any i. To do that, we show that $1|_i \in R_i$ (recall that $1|_i$ means putting probability 1 on component i). Indeed, suppose $1|_i \notin R_i$, then for all $z \in D_{1|_i}$, $z_i < a_i$, and by definition of $G_p(z)$, we have $G_{1|_i}(z) \leq 0$. However, this is impossible for $z \in D_{1|_i}$ by definition of D_p; this contradiction shows that indeed $1|_i \in R_i$.

Introduce the following notation. Let $\|\cdot\|$ denote the sup-norm on $\mathbb{R}^n$, and let $d(x, Y)$

be the distance from point x to nonempty set Y:

$$d(x,Y) = \inf_{y\in Y} \|x-y\|.$$

Now for any $\varepsilon \geq 0$ and nonempty $Y \subset \Delta$, let $N(Y,\varepsilon)$ be the closed ε-neighborhood of set Y, i.e.,

$$N(Y,\varepsilon) = \{p \in \Delta : d(p,Y) \leq \varepsilon\}.$$

Consistently with this definition, $N(Y,0) = \overline{Y}$, the closure of Y (which equals Y if Y is closed).

Let us now show that $\bigcap_{i=1}^n \overline{R_i} = \varnothing$. Suppose not, then there is some $p \in \bigcap_{i=1}^n \overline{R_i}$. Take $\varepsilon \in \left(0, \frac{1}{n(n+1)} \frac{l_p}{V(a)+1}\right]$ such that for any $r \in N(\{p\},\varepsilon)$, $l_r \geq \frac{n(n+2)}{(n+1)^2} l_p$; this is possible because l_p is continuous (and the coefficient is smaller than 1). Since $p \in \bigcap_{i=1}^n \overline{R_i}$, for each $i \in \{1,\ldots,n\}$ there is $p^{(i)} \in N(\{p\},\varepsilon) \cap R_i$; by definition of R_i we can then take $z^{(i)} \in D_{p^{(i)}}$ such that $z_i^{(i)} = a_i$. By definition of $D_{p^{(i)}}$, we have $G_{p^{(i)}}\left(z^{(i)}\right) > \frac{n+1}{n+2} l_{p^{(i)}} \geq \frac{n}{n+1} l_p$. By Lipschitz continuity of $G_p\left(z^{(i)}\right)$ as a function of p (with coefficient $V(a)$), we have

$$\begin{aligned} G_p\left(z^{(i)}\right) &\geq G_{p^{(i)}}\left(z^{(i)}\right) - V(a)\left\|p - p^{(i)}\right\| \\ &> \frac{n}{n+1} l_p - V(a) \frac{1}{n(n+1)} \frac{l_p}{V(a)+1} \\ &> \left(\frac{n}{n+1} - \frac{1}{n(n+1)}\right) l_p = \frac{n-1}{n} l_p. \end{aligned}$$

Denote, for any $k \in \{1,\ldots,n\}$, $y^{(k)} = \bigvee_{i=1}^k z^{(i)}$; in particular, $y^{(1)} = z^{(1)}$. Let us now show, by induction, that $G_p\left(y^{(k)}\right) > \frac{n-k}{n} l_p$. Indeed, the base $k=1$ is already established. Suppose that $G_p\left(y^{(k-1)}\right) > \frac{n-(k-1)}{n} l_p$, then we have by supermodularity

$$\begin{aligned} G_p\left(y^{(k)}\right) &= G_p\left(y^{(k-1)} \vee z^{(k)}\right) \\ &\geq G_p\left(y^{(k-1)}\right) + G_p\left(z^{(k)}\right) - G_p\left(y^{(k-1)} \wedge z^{(k)}\right) \\ &> \frac{n-(k-1)}{n} l_p + \frac{n-1}{n} l_p - l_p = \frac{n-k}{n} l_p, \end{aligned}$$

where we used $G_p\left(y^{(k-1)} \wedge z^{(k)}\right) \leq l_p$ by definition of l_p. This proves the induction step. Now, taking $k=n$, we have $G_p\left(y^{(n)}\right) > 0$. However, $y^{(n)} = a$, and we get a contradiction, since $G_p(a) = 0$. This contradiction shows that p cannot exist, so $\bigcap_{i=1}^n \overline{R_i} = \varnothing$.

Now for every $p \in \Delta$, define $E_p = \{q \in \Delta : G_q(x) \leq 0 \text{ for all } x \in D_p\}$. In other words, E_p is the set of probabilities that make deviation to a unprofitable for all types $x \in D_p$. If we can prove existence of p such that $p \in E_p$ (i.e., a fixed point of mapping $p \mapsto E_p$),

then we will reach a contradiction that proves the result. Notice that for every $p \in \Delta$, E_p is closed and convex, because it is the intersection of closed convex sets given by linear inequalities. Also, for every $p \in \Delta$, E_p is nonempty, because $p \notin R_i$ for some R_i (indeed, we showed that $\bigcap_{i=1}^n \overline{R_i} = \varnothing$, so $\bigcap_{i=1}^n R_i = \varnothing$ as well), in which case vector $1|_i \in E_p$. If the correspondence E_p were upper-hemicontinuous, we would immediately get existence of a fixed point by Kakutani's theorem. Unfortunately, this might not be true.

Define

$$h = \inf_{p \in \Delta} \max_{i \in \{1,\ldots,n\}} d(p, R_i);$$

for each p the maximum is finite and well-defined, because each $R_i \neq \varnothing$. Let us show that $h > 0$. Since the infimum is taken over a compact set and the function $d(p, R_i)$ is continuous in p, it is achieved for some $p \in \Delta$. If $h = 0$, then $d(p, R_i) = 0$ for all i, and thus $p \in \overline{R_i}$ for all R_i. But we showed that $\bigcap_{i=1}^n \overline{R_i} = \varnothing$, which yields a contradiction that proves that $h > 0$. This implies, in particular, that for every point $p \in \Delta$, there is $i \in \{1, \ldots, n\}$ such that within the $\frac{h}{2}$-neighborhood of p there are no points belonging to R_i.

For each $p \in \Delta$, introduce the set Q_p given by:

$$Q_p = \bigcap_{q \in \Delta} N\left(E_q, \frac{2}{h}\|p - q\|\right).$$

We establish the following properties.

First, for every p, $Q_p \subset E_p$, because for $q = p$, $N\left(E_q, \frac{2}{h}\|p - q\|\right) = N(E_p, 0) = \overline{E_p} = E_p$, since E_p is closed.

Second, for every p, Q_p is convex, because it is the intersection of convex sets ($N(Y, \varepsilon)$ is convex for any ε if Y is convex, and E_q is convex for each q).

Third, let $Q \subset \Delta \times \Delta$ be the graph of mapping $p \mapsto Q_p$, i.e.,

$$Q = \{(p, r) \in \Delta \times \Delta : r \in Q_p\};$$

then Q is closed. To see this, notice that

$$\begin{aligned} Q &= \bigcup_{p \in \Delta} \bigcap_{q \in \Delta} \left\{(p, r) : r \in N\left(E_q, \frac{2}{h}\|p - q\|\right)\right\} \\ &= \bigcap_{q \in \Delta} \bigcup_{p \in \Delta} \left\{(p, r) : r \in N\left(E_q, \frac{2}{h}\|p - q\|\right)\right\}. \end{aligned}$$

But for each q, the mapping $p \mapsto N\left(E_q, \frac{2}{h}\|p - q\|\right)$ has a closed graph (this is a continuous

set-valued mapping), and thus Q is closed as an intersection of closed sets.

Fourth, for every $p \in \Delta$, Q_p is nonempty. Indeed, from the definition of h it follows that there is $i \in \{1, \ldots, n\}$ such that $q \in N\left(\{p\}, \frac{h}{2}\right)$ implies $q \notin R_i$, and in particular $p \notin R_i$. Let us show that the vector $1|_i \in Q_p$. To do this, we need to show that for every $q \in \Delta$,

$$1|_i \in N\left(E_q, \frac{2}{h}\|p - q\|\right).$$

If $q \in N\left(\{p\}, \frac{h}{2}\right)$, we have $q \notin R_i$, which implies $1|_i \in E_q$, which establishes the required inclusion for such q. In the complementary case, $q \notin N\left(\{p\}, \frac{h}{2}\right)$, we have $\|p - q\| > \frac{h}{2}$, and thus $N\left(E_q, \frac{2}{h}\|p - q\|\right) = \Delta$ (since E_q is nonempty and the maximum distance between two points in Δ is 1). So the required inclusion is satisfied in this case as well. Since it holds for every q, this proves that $1|_i \in Q_p$, so $Q_p \neq \varnothing$ for any $p \in \Delta$.

Now, the second, third, and fourth properties show that the mapping $p \mapsto Q_p$ satisfies the requirements of Kakutani's fixed-point theorem. Therefore, there is $p \in \Delta$ such that $p \in Q_p$. The first property now implies that this $p \in E_p$. Therefore, the mapping $p \mapsto E_p$ has a fixed point. We have that for all $x \in D_p$, $G_p(x) \leq 0$, which contradicts the definition of D_p. This contradiction completes the proof. □

Proof of Proposition 6. We just need to check that if condition (1) is satisfied for the function V, then it is also satisfied for $V' = U \circ V$. We have, for any $a, \hat{a}$,

$$\begin{aligned} U(V(a)) &\geq U\left(p_0(\hat{a})V(\hat{a}) + \sum_{i=1}^{n} p_i(\hat{a})w_i(\hat{a}|a)\right) \\ &\geq p_0(\hat{a})U(V(\hat{a})) + \sum_{i=1}^{n} p_i(\hat{a})U(w_i(\hat{a}|a)) \end{aligned}$$

where the first inequality is because U is increasing and the second is because U is concave. The result follows. □

Proof of Proposition 7. As in the proof of Proposition 4, it is enough to consider deviations to a by types (x_1, a_2) and (a_1, x_2) with $x < a$.

To show necessity: Denote $p_1 = p_1(a)$, so at a, dimension 1 is checked with probability p_1. Then write conditions stating that from (a_1, x_2) and (x_1, a_2) it is not profitable to deviate to a:

$$p_1 V(a) + (1 - p_1) V(0, x_2) \leq V(a_1, x_2); \tag{A1}$$

$$p_1 V(x_1, 0) + (1 - p_1) V(a) \leq V(x_1, a_2). \tag{A2}$$

The first implies $p_1 \leq \frac{V(a_1,x_2)-V(0,x_2)}{V(a)-V(0,x_2)}$ (if the denominator is 0, then by monotonicity $V(a) = V(a_1, x_2) = V(0, x_2)$ and the numerator is also 0). The second implies $p_1 \geq \frac{V(a)-V(x_1,a_2)}{V(a)-V(x_1,0)}$ (again, if the denominator is 0, then $V(a) = V(x_1, a_2) = V(x_1, 0)$ by monotonicity and the numerator is also 0). We thus have $\frac{V(a)-V(x_1,a_2)}{V(a)-V(x_1,0)} \leq \frac{V(a_1,x_2)-V(0,x_2)}{V(a)-V(0,x_2)}$. Cross-multiplying gives

$$(V(a) - V(x_1, a_2))(V(a) - V(0, x_2)) \leq (V(a) - V(x_1, 0))(V(a_1, x_2) - V(0, x_2)),$$

which also holds in either of the zero-denominator cases (since both sides are then zero). Adding $(V(a) - V(x_1, a_2))(V(0, x_2) - V(a_1, x_2))$ to both sides gives the condition in the proposition.

To show sufficiency: Suppose the condition holds for all $x, a \in A$ such that $x < a$. Fix a, and let us find probability p_1 such that if a report of a leads to verification probabilities $p_1(a) = p_1$, $p_2(a) = 1 - p_1$, this deters all deviations to a. First, notice that if $V(a) = V(x_1, 0)$ for some $x < a$, then $p_1 = 1$ will work (monotonicity implies $V(a_1, x_2) = V(a)$ for all $x_2 < a_2$, so none of these types gains from deviating, and types (x_1, a_2) will be caught with certainty). Similarly, if $V(a) - V(0, x_2)$ for some $x < a$, then $p_1 = 0$ will work. Thus, we may assume that for any $x < a$, $V(a) > V(x_1, 0)$ and $V(a) > V(0, x_2)$. Again rearranging the terms, the given inequality implies

$$\frac{V(a) - V(x_1, a_2)}{V(a) - V(x_1, 0)} \leq \frac{V(a_1, x_2) - V(0, x_2)}{V(a) - V(0, x_2)}.$$

Since the left-hand side depends on x_1 only and right-hand side depends on x_2 only, we have

$$\sup_{x_1} \frac{V(a) - V(x_1, a_2)}{V(a) - V(x_1, 0)} \leq \inf_{x_2} \frac{V(a_1, x_2) - V(0, x_2)}{V(a) - V(0, x_2)}.$$

Now if we take $p_1 \in \left[\sup_{x_1} \frac{V(a)-V(x_1,a_2)}{V(a)-V(x_1,0)}, \inf_{x_2} \frac{V(a_1,x_2)-V(0,x_2)}{V(a)-V(0,x_2)}\right]$, we will have that for any x_1 and any x_2,

$$\frac{V(a) - V(x_1, a_2)}{V(a) - V(x_1, 0)} \leq p_1 \leq \frac{V(a_1, x_2) - V(0, x_2)}{V(a) - V(0, x_2)}.$$

Rearranging each of these inequalities, we get the conditions (A1)–(A2), saying that from (a_1, x_2) and (x_1, a_2) it is not profitable to deviate to a. □

Proof of Proposition 9. First we show necessity. Take the weights λ_S as given; we can assume $\lambda_\varnothing = 0$, since the value of $\lambda_\varnothing$ has no effect either on the validity of the collection of weights or on the inequality to be proven. Type $a|_S$ can, by imitating type

a, get at least $\left(p_0(a) + \sum_{i \in S} p_i(a)\right) V(a)$. Hence, incentive compatibility implies

$$\left(p_0(a) + \sum_{i \in S} p_i(a)\right) V(a) \leq V(a|_S).$$

Now multiply by λ_S, and then sum over all S. On the left side, the coefficient $p_0(a)$ appears with total weight $\sum_S \lambda_S \geq 1$, and for each $i = 1, \ldots, n$, $p_i(a)$ appears with total weight $\sum_{S: i \in S} \lambda_S = 1$. Hence, we get

$$\left(p_0(a) + \sum_{i=1}^{n} p_i(a)\right) V(a) \leq \sum_S \lambda_S V(a|_S).$$

The left side is just $V(a)$, showing that the asserted condition holds.

Now we prove sufficiency. For each type a, we need to construct the appropriate verification probabilities $p_i(a)$ to discourage deviations to a. If $V(a) = 0$ we can choose these probabilities arbitrarily, as clearly no type would deviate to such a. Now assume $V(a) > 0$.

We claim that there exist nonnegative numbers $r_1, \ldots, r_n$ such that $r_1 + \cdots + r_n = V(a)$ and, for each subset $S \subset \{1, \ldots, n\}$, $\sum_{i \in S} r_i \leq V(a|_S)$.

Suppose not. Then, applying a theorem of the alternative, we get the existence of nonnegative numbers λ_S, for each $S \subset \{1, \ldots, n\}$, such that $\sum_{S: i \in S} \lambda_S \geq 1$ for each i and $\sum_S \lambda_S V(a|_S) < V(a)$.

This is almost a contradiction to our assumed condition on V, except that for each index i, the total weight on sets containing i is ≥ 1, rather than exactly 1 as required. However, if the inequality is strict, then we can take some of the weight on a set S containing i and transfer it to set $S \backslash \{i\}$. This decreases the total weight on sets containing i, without changing the total weight on sets containing j, for any $j \neq i$. Iterating this, we can eventually get the total weight on sets containing i to be exactly 1 for each i. Moreover, each such operation can only decrease the value of $\sum_S \lambda_S V(a|_S)$, since V is monotone and we are transferring weight from larger to smaller sets. Hence the final weights will satisfy $\sum_{i \in S} \lambda_S = 1$ for each index i, and will still satisfy $\sum_S \lambda_S V(a|_S) < V(a)$, thus contradicting the assumption.

This implies the desired numbers $r_1, \ldots, r_n$ exist. Define the verification probabilities by $p_i(a) = r_i / V(a)$. We just need to check incentive compatibility condition (4).

Suppose the sender has type a, but reports $\hat{a}$. Let S be the set of coordinates i for which $\hat{a}_i = a_i$. Then

$$\sum_{i \in S} p_i(\hat{a}) = \frac{\sum_{i \in S} r_i}{V(\hat{a})} \leq \frac{V(a|_S)}{V(\hat{a})} \leq \frac{V(a)}{V(\hat{a})},$$

which is exactly what (4) requires. □

Proof of Proposition 10. Let us prove that the proposed mechanism is incentive compatible. If the sender has type a and reports truthfully, he evidently gets $V(a)$. If he falsely reports $\hat{a}$, then he gets the reward $V(\hat{a})$ only if the verified dimension i is such that $\hat{a}_i = a_i$; let S be the set of such indices i. Using the notation $a|_S$ as in the text, the sender's expected payoff from misreporting is

$$\begin{aligned}\sum_{i\in S} \frac{V(\hat{a}|_{[i]}) - V(\hat{a}|_{[i-1]})}{V(\hat{a})}\, V(\hat{a}) &= \sum_{i\in S}(V(\hat{a}|_{[i]}) - V(\hat{a}|_{[i-1]})) \\ &\le \sum_{i\in S}(V((a|_S)|_{[i]}) - V((a|_S)|_{[i-1]})) \\ &= V(a|_S) \\ &\le V(a).\end{aligned}$$

Here the first inequality is by submodularity, and the second is because V is increasing. So, there is no incentive to lie. □

The following auxiliary result will be used in the proof of Proposition 11.

Lemma A1. *Let $q_1, \dots, q_n$, $m_1, \dots, m_n$ and $t_1, \dots, t_n$ be positive numbers such that $t_j < m_j$ for all j, and $\sum_j q_j = \sum_j m_j = 1$. Put $t = \sum_j t_j$. Then,*

$$\left(\sum_j (m_j - t_j)\left(\frac{q_j}{m_j}\right)^{1-t}\right)\prod_j \left(\frac{q_j}{m_j}\right)^{t_j} \le 1 - t.$$

Proof. Note that for each i, we have

$$(m_i - t_i)\left(\frac{q_i}{m_i}\right)^{1-t}\prod_j \left(\frac{q_j}{m_j}\right)^{t_j} \le (m_i - t_i)\left[(1-t)\frac{q_i}{m_i} + \sum_j t_j \frac{q_j}{m_j}\right]$$

by the AM-GM inequality. Sum over $i = 1, \dots, n$. On the right-hand side, each q_i/m_i appears with coefficient

$$(m_i - t_i)(1-t) + \sum_j (m_j - t_j)t_i = (m_i - t_i)(1-t) + (1-t)t_i = m_i(1-t).$$

So the right-hand side simplifies to $\sum_i m_i(1-t)(q_i/m_i) = (1-t)\sum_i q_i = 1 - t$, and the lemma follows. □

Proof of Proposition 11. We have defined M^χ and p^χ in the text. We have also defined $\sigma^\chi(a)$ for types $a \in A_K$. This allows us to define beliefs $\mu^\chi(h) \in \Delta(A_K)$ by Bayesian updating. Below, we will give an explicit formula for μ^χ. This then completes the definition of the mechanism, except for one detail: the formal definition of a mechanism requires specifying $\sigma^\chi(a)$ for every type $a \in A$, not just for types $a \in A_K$. However, we can subsequently assign to each type $a \notin A_K$ whatever message maximizes its expected payoff (the maximum exists by continuity arguments); since these types collectively have probability zero, this assumption does not affect the Bayesian updating. The mechanism is then completely defined. We will then show that the resulting mechanism is incentive-compatible, i.e., a valid mechanism; and we will establish the convergence property.

Let $q_i = \frac{a_i}{\sum_{j=1}^n a_j} = \frac{a_i}{V(a)}$ be the relative skills of the sender. Suppose that the receiver got message m and treats it as truthfully reflecting the relative skills of the sender, $m = q$. Conditional on this information, the posterior distribution of $V(a)$ is lognormal, so that $(\log V(\tilde{a}) \mid m) \sim \mathcal{N}\left(\sum_{i=1}^n \frac{\tau_i}{\tau}(\nu_i - \log m_i), \frac{1}{\tau}\right)$. (Hereinafter we write $\tilde{a}$ for the unknown type that is a random variable from the receiver's point of view, and a for the true type.) This follows from the following calculation: the conditional density of $(\log V(\tilde{a}) \mid m)$ at point z is equal to:

$$
\begin{aligned}
& \frac{\prod_{i=1}^n \sqrt{\frac{\tau_i}{2\pi}} \exp\left(-\frac{1}{2}\sum_{i=1}^n \tau_i \left(z + \log m_i - \nu_i\right)^2\right)}{\int_{-\infty}^{+\infty} \prod_{i=1}^n \sqrt{\frac{\tau_i}{2\pi}} \exp\left(-\frac{1}{2}\sum_{i=1}^n \tau_i \left(\lambda + \log m_i - \nu_i\right)^2\right) d\lambda} \\
= & \frac{\exp\left(-\frac{1}{2}\sum_{i=1}^n \tau_i \left(z + \log m_i - \nu_i\right)^2\right)}{\int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2}\sum_{i=1}^n \tau_i \left(\lambda + \log m_i - \nu_i\right)^2\right) d\lambda} \\
= & \frac{\exp\left(-\frac{\tau}{2}\left(\left(z - \sum_{i=1}^n \frac{\tau_i(\nu_i - \log m_i)}{\tau}\right)^2 + \sum_{i=1}^n \frac{\tau_i(\nu_i - \log m_i)^2}{\tau} - \left(\sum_{i=1}^n \frac{\tau_i(\nu_i - \log m_i)}{\tau}\right)^2\right)\right)}{\int_{-\infty}^{+\infty} \exp\left(-\frac{\tau}{2}\left(\left(\lambda - \sum_{i=1}^n \frac{\tau_i(\nu_i - \log m_i)}{\tau}\right)^2 + \sum_{i=1}^n \frac{\tau_i(\nu_i - \log m_i)^2}{\tau} - \left(\sum_{i=1}^n \frac{\tau_i(\nu_i - \log m_i)}{\tau}\right)^2\right)\right) d\lambda} \\
= & \frac{\sqrt{\frac{\tau}{2\pi}} \exp\left(-\frac{1}{2}\tau\left(z - \sum_{i=1}^n \frac{\tau_i}{\tau}(\nu_i - \log m_i)\right)^2\right)}{\sqrt{\frac{\tau}{2\pi}} \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2}\tau\left(\lambda - \sum_{i=1}^n \frac{\tau_i}{\tau}(\nu_i - \log m_i)\right)^2\right) d\lambda} \\
= & \sqrt{\frac{\tau}{2\pi}} \exp\left(-\frac{1}{2}\tau\left(z - \sum_{i=1}^n \frac{\tau_i}{\tau}(\nu_i - \log m_i)\right)\right).
\end{aligned}
$$

Now suppose that the receiver tested dimension i and got signal $s = a_i\eta = V(a)\, m_i\eta$. Conditional on m being equal to q as assumed by the receiver so far, s has lognormal distribution, with $\log s = (\log V(a) \mid m) + \log m_i + \log \eta$. Thus, $\log s - \log m_i$ is a signal of the unknown value $(\log V(\tilde{a}) \mid m)$ with precision χ. Thus, we have that the posterior of $V(\tilde{a})$ is lognormal, with

$$(\log V(\tilde{a}) \mid m, i, s) \sim \mathcal{N}\left(\frac{\tau}{\tau+\chi}\sum_{j=1}^{n}\frac{\tau_j}{\tau}(\nu_j - \log m_j) + \frac{\chi}{\tau+\chi}(\log s - \log m_i), \frac{1}{\tau+\chi}\right).$$

This pins down the belief $\mu^\chi(m, i, s)$: it is a (one-dimensional) lognormal distribution on the ray of types whose relative skills agree with m; the parameters of this lognormal are as indicated above. This completes the description of the mechanism, as indicated in the first paragraph above.

Notice that this formula for beliefs implies the convergence part of the Proposition. Indeed, for a truthful report by the sender of a given type a, $m_i = q_i = \frac{a_i}{V(a)}$ and $\log s - \log m_i = \log V(a) + \log \eta$. For a fixed η (equivalently, fixed s), $(\log V(\tilde{a}) \mid m, i, s)$ may be thought of as a sum of a variable distributed as $\mathcal{N}\left(\frac{\tau}{\tau+\chi}\sum_{j=1}^{n}\frac{\tau_j}{\tau}(\nu_j - \log m_j) + \frac{\chi}{\tau+\chi}\log V(a), \frac{1}{\tau+\chi}\right)$ and a constant $\frac{\chi}{\tau+\chi}\log\eta$. Therefore, if we take the expectation over realizations of s (equivalently, η), we get

$$\mathbb{E}_s[\log V(\tilde{a}) \mid m, i, s] \sim \mathcal{N}\left(\frac{\tau}{\tau+\chi}\sum_{j=1}^{n}\frac{\tau_j}{\tau}(\nu_j - \log m_j) + \frac{\chi}{\tau+\chi}\log V(a), \frac{\tau+2\chi}{(\tau+\chi)^2}\right);$$

we used that $\frac{\chi}{\tau+\chi}\log\eta$ is normal with expectation 0 and variance $\left(\frac{\chi}{\tau+\chi}\right)^2\frac{1}{\chi} = \frac{\chi}{(\tau+\chi)^2}$. For each given i, this distribution converges to an atom on $\log V(a)$ as $\chi \to \infty$; furthermore, note that the distribution actually is the same for all i. This proves that following truthful report $m = q$, the expected posterior over $\log V(\tilde{a})$ (averaged over both i and s) converges to an atom in $\log V(a)$. Since there is no uncertainty about the relative skills (they are given by m), the convergence of $\kappa(a)$ follows.

It remains to prove that the constructed mechanism is incentive compatible for the sender. For a given realization of i and s, the sender who sent message m (not necessarily truthfully!) expects to get a payoff equal to

$$\mathbb{E}[V(\tilde{a}) \mid m, i, s] = \exp\left(\frac{\tau}{\tau+\chi}\sum_{j=1}^{n}\frac{\tau_j}{\tau}(\nu_j - \log m_j) + \frac{\chi}{\tau+\chi}(\log s - \log m_i) + \frac{1}{2}\frac{1}{\tau+\chi}\right),$$

since this is the expectation of exponent of $(\log V(\tilde{a}) \mid m, i, s)$, which is normally distributed. Now, continuing to write a for the true type, and taking expectation over possible realizations of s (or, equivalently, over η), we get

$$\exp\left(\frac{1}{\tau+\chi}\sum_{j=1}^{n}\tau_j\nu_j + \frac{1}{2}\frac{2\chi+\tau}{(\chi+\tau)^2}\right)\left(\frac{a_i}{m_i}\right)^{\frac{\chi}{\tau+\chi}}\prod_{j=1}^{n}m_j^{-\frac{\tau_j}{\tau+\chi}}.$$

Indeed, we have

$$\begin{aligned}
&\mathbb{E}_\eta \exp\left(\frac{\tau}{\tau+\chi}\sum_{j=1}^{n}\frac{\tau_j}{\tau}(\nu_j-\log m_j) + \frac{\chi}{\tau+\chi}(\log a_i+\log\eta-\log m_i) + \frac{1}{2}\frac{1}{\tau+\chi}\right)\\
=\;& \exp\left(\frac{\tau}{\tau+\chi}\sum_{j=1}^{n}\frac{\tau_j}{\tau}(\nu_j-\log m_j) + \frac{\chi}{\tau+\chi}(\log a_i-\log m_i) + \frac{1}{2}\frac{1}{\tau+\chi}\right)\mathbb{E}_\eta\exp\left(\frac{\chi}{\tau+\chi}\log\eta\right)\\
=\;& \exp\left(\frac{\tau}{\tau+\chi}\sum_{j=1}^{n}\frac{\tau_j}{\tau}(\nu_j-\log m_j) + \frac{\chi}{\tau+\chi}(\log a_i-\log m_i) + \frac{1}{2}\frac{1}{\tau+\chi} + \frac{1}{2}\frac{\chi}{(\tau+\chi)^2}\right)\\
=\;& \exp\left(\frac{\tau}{\tau+\chi}\sum_{j=1}^{n}\frac{\tau_j}{\tau}(\nu_j-\log m_j) + \frac{\chi}{\tau+\chi}(\log a_i-\log m_i) + \frac{1}{2}\frac{\tau+2\chi}{(\tau+\chi)^2}\right)\\
=\;& \exp\left(\frac{1}{\tau+\chi}\sum_{j=1}^{n}\tau_j\nu_j + \frac{1}{2}\frac{\tau+2\chi}{(\tau+\chi)^2}\right)\left(\frac{a_i}{m_i}\right)^{\frac{\chi}{\tau+\chi}}\prod_{j=1}^{n}m_j^{-\frac{\tau_j}{\tau+\chi}}.
\end{aligned}$$

Therefore, his expected payoff from sending message m equals

$$\begin{aligned}
&\sum_{i=1}^{n}p_i(m)\exp\left(\frac{1}{\tau+\chi}\sum_{j=1}^{n}\tau_j\nu_j + \frac{1}{2}\frac{\tau+2\chi}{(\tau+\chi)^2}\right)\left(\frac{a_i}{m_i}\right)^{\frac{\chi}{\tau+\chi}}\prod_{j=1}^{n}m_j^{-\frac{\tau_j}{\tau+\chi}}\\
=\;& C\times\sum_{i=1}^{n}\left(m_i\left(1+\frac{\tau}{\chi}\right)-\frac{\tau_i}{\chi}\right)\left(\frac{a_i}{m_i}\right)^{\frac{\chi}{\tau+\chi}}\prod_{j=1}^{n}m_j^{-\frac{\tau_j}{\tau+\chi}}\\
=\;& C\times H(m),
\end{aligned}$$

where

$$C=\exp\left(\frac{1}{\tau+\chi}\sum_{j=1}^{n}\tau_j\nu_j+\frac{1}{2}\frac{\tau+2\chi}{(\tau+\chi)^2}\right),$$

$$H(m)=\left(\prod_{j=1}^{n}m_j^{-\frac{\tau_j}{\tau+\chi}}\right)\left(\sum_{j=1}^{n}\left(m_j\left(1+\frac{\tau}{\chi}\right)-\frac{\tau_j}{\chi}\right)\left(\frac{a_j}{m_j}\right)^{\frac{\chi}{\tau+\chi}}\right).$$

Now, we need to prove that it is indeed optimal to send message $m=q$, with $q_i=$

$\frac{a_i}{\sum_{j=1}^n a_j}$. Since the C factor is a constant, we need to prove that

$$q \in \arg\max_{m \in M} H(m).$$

We apply Lemma A1 to $\{q_j\}$ and $\{m_j\}$, taking $t_j = \tau_j/(\tau+\chi)$ (so $t = \tau/(\tau+\chi)$). Take the resulting inequality and multiply both sides by $\frac{\tau+\chi}{\chi}\left(\sum_j a_j\right)^{\frac{\chi}{\tau+\chi}}\left(\prod_j q_j^{-\frac{\tau_j}{\tau+\chi}}\right)$. Then the left side equals $H(m)$, and the right side equals $\left(\sum_j a_j\right)\left(\prod_j a_j^{-\frac{\tau_j}{\tau+\chi}}\right) = H(q)$. This shows that $H(m) \leq H(q)$ for all m, so it is optimal to report the true relative skills, i.e. the mechanism is incentive-compatible. □

The next result formalizes and proves the claim in Footnote 9.

Proposition A2. *Suppose $V(\mu) = (\mathbb{E}_{a\sim\mu}[\sum_{i=1}^n a_i])^\gamma$ for some $\gamma \in (0,1)$. Then Proposition 11 is correct as stated.*

Proof. The detailed proof of this result follows the same steps as the proof of Proposition 11 and is omitted. The argument below highlights one additional step needed to complete the proof.

As in the proof of Proposition 11, we show that the posterior distribution of $\sum_{i=1}^n \tilde{a}_i = V(\tilde{a})^{1/\gamma}$ is given by

$$\left(\log V(\tilde{a})^{1/\gamma} \mid m, i, s\right) \sim \mathcal{N}\left(\frac{\tau}{\tau+\chi}\sum_{j=1}^n \frac{\tau_j}{\tau}(\nu_j - \log m_j) + \frac{\chi}{\tau+\chi}(\log s - \log m_i), \frac{1}{\tau+\chi}\right).$$

We then use similar steps to reduce incentive-compatibility to showing that the true vector q of relative skills satisfies

$$q \in \arg\max_{m \in M} H_\gamma(m),$$

where

$$H_\gamma(m) = \left(\prod_{j=1}^n m_j^{-\frac{\gamma\tau_j}{\tau+\chi}}\right)\left(\sum_{j=1}^n \left(m_j\left(1+\frac{\tau}{\chi}\right) - \frac{\tau_j}{\chi}\right)\left(\frac{a_j}{m_j}\right)^{\frac{\gamma\chi}{\tau+\chi}}\right).$$

To show this, we take notice that $H_\gamma(m) \leq (H(m))^\gamma$ for all m; indeed,

$$\left(\sum_{j=1}^n \left(m_j\left(1+\frac{\tau}{\chi}\right) - \frac{\tau_j}{\chi}\right)\left(\frac{a_j}{m_j}\right)^{\frac{\chi}{\tau+\chi}}\right)^\gamma \geq \sum_{j=1}^n \left(m_j\left(1+\frac{\tau}{\chi}\right) - \frac{\tau_j}{\chi}\right)\left(\frac{a_j}{m_j}\right)^{\frac{\gamma\chi}{\tau+\chi}}$$

by concavity of the power function for $\gamma \in (0,1)$. At the same time, $H_\gamma(q) = H(q)$, because $\frac{a_j}{q_j} = (V(a))^{1/\gamma}$, which is a constant. Thus, the inequality $H(m) \leq H(q)$, which

was established in the proof of Proposition 11, implies

$$H_\gamma(m) \leq (H(m))^\gamma \leq (H(q))^\gamma = H_\gamma(q),$$

thus establishing incentive compatibility. □

Finally, we return to the setting of perfect verification, and we formalize the claim in Footnote 6, that full learning of $V(a)$ implies full learning of a is possible. For this we must return to the original formulation of the model, where posterior beliefs are non-degenerate, and V is defined on $\Delta(A)$. We need an extra assumption: Say that V *respects constant values* if, for every constant c, if μ is any distribution on A such that $V(a) = c$ for all a in the support of μ, then $V(\mu) = c$ as well.

Say that an (indirect) mechanism achieves *full learning of* $V(a)$ if, for every type a, every history $h \in H(a|\mathcal{M})$, and every $a' \in \text{supp}(\mu(h))$, we have $V(a') = V(a)$.

Proposition A3. *Assume that V respects constant values. If there exists an indirect mechanism that achieves full learning of $V(a)$, then there exists a direct mechanism with full learning of a.*

Proof. Let $\mathcal{M} = (M, \sigma, p, \mu)$ be the mechanism that achieves full learning of $V(a)$. We now repeat the proof of Lemma 0. The same proof goes through, except for two adjustments: in the two steps that originally applied full learning for the original mechanism, we now apply full learning of $V(a)$ together with respecting constant values; and the fact that type a receives equilibrium payoff $V(a)$ in the original mechanism also uses these two properties. □