

Strategic Communication with Minimal Verification*

Gabriel Carroll[†]  Georgy Egorov[‡]



December 2018

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 in many cases. We characterize exactly when it is possible to do so. In particular, when the exogenous payoff function is submodular, we can explicitly describe a full-learning mechanism; when it is (strictly) supermodular, full learning is not possible. We consider variants, including the possibility of using an indirect mechanism with no off-path histories, and a version with noisy verification. We also show that constructions based on our approach remain useful even when full learning is not possible.

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

JEL Codes: D82, D83

*We are grateful to Sandeep Baliga, Vincent Crawford, Tommaso Denti, Xavier Gabaix, Johannes Hörner, Navin Kartik, Elliot Lipnowski, Paul Milgrom, Roger Myerson, Larry Samuelson, Ran Spiegler, Nisan Stiennon, and seminar participants at the University of Chicago, Northwestern University, the University of Pittsburgh, Princeton University, ETH/UZH, CEU, CERGE-EI, LMU and conference participants at Tel Aviv University, Cornell University, and Transatlantic Theory Workshop in Paris for helpful comments. Carroll also thanks the Research School of Economics at the Australian National University for their hospitality during a sabbatical.  denotes random author order (Ray  Robson, 2018).

[†]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 infer the former's private information as precisely as possible.

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.

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 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. (For most of the paper we assume this verification is perfect, although we briefly consider a variation 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. The sender, 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 Example 1, 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 $V(a)$; we can give a complete characterization of the functions $V(a)$ for which full learning is possible. In particular, when $V(a)$ is submodular, full learning is possible, whereas if $V(a)$ is strictly supermodular then it is impossible. (In the boundary case where V is additively separable, as in Example 1 above, the mechanism is essentially unique.) Our general argument uses direct-revelation mechanisms, but when $V(a)$ is submodular and satisfies some additional regularity conditions, we can also construct an indirect mechanism in which the sender chooses probabilities of testing each dimension, generalizing Example 1. We then give some robustness checks, including giving an example where the verified dimension is observed with noise, in which case exact learning cannot be achieved but we show how it can be achieved in the limit as the level of noise goes to zero. Finally, we return to perfect verification but consider specifications of $V(a)$ such that full learning is impossible, and show that the ideas from our initial analysis can still be leveraged to learn a substantial amount, and in some cases lead to mechanisms that are optimal in an appropriate sense. In the conclusion we discuss the practical takeaways and interpretation of the formal results.

Our assumption that the receiver has no control over payoffs (for given beliefs) is natural for many settings: e.g., in the case of an interviewer and a job candidate, the interviewer might be obliged to write a truthful report of what she learned to her supervisor, so she may exercise control over what she chooses to learn, but she cannot manipulate the candidate's payoffs in any other way. Our other central assumption, that the receiver can verify exactly one dimension, is of course more stylized. We adopt this assumption to achieve the starkest results, showing that a minimal amount of verification allows full learning in Example 1; by maintaining the assumption throughout, we can ask how far the example can be pushed, in a way that allows for a crisp answer (Proposition 2). It also allows us to best connect to existing literature, as discussed below.

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, even when the sender’s preferences are independent of his type, information on all but one dimension (the “dimensions of agreement”) may be transmitted; this result has some resemblance to our example above, where one might view the verification as filling in the missing dimension. Lipnowski and Ravid (2017) consider a more general, abstract formulation and characterize optimal equilibrium outcomes for the sender. Battaglini (2002) studies cheap talk with multiple senders; his model shares with ours the possibility of full learning.¹

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. 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.² 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 characterize the optimal mechanism as a solution to a particular linear programming problem, show that it takes a fairly simple form, and show that random mechanisms may be necessary to achieve the optimum. In contrast to their paper, we consider a broader range of payoffs for the sender, but focus primarily 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.³

Azar and Micali (2018) also study a problem in which an agent has access to a high-dimensional vector, and the principal wishes to know the value of some function of the vector, without having the whole vector communicated. They show a result with some resemblance to ours: their principal can incentivize approximate revelation of the true value while verifying just one component. They allow the principal to design the incentives freely, in contrast to our exogenously given $V(a)$.

Our paper is also related in spirit to other literature on communication with verifica-

¹Other papers addressing full or nearly-full learning with multiple senders include Ambrus and Takahashi (2008), Meyer, Moreno de Barreda, and Nafziger (2016), and Ambrus and Lu (2014).

²Glazer and Rubinstein (2004) also mention a number of further examples of applications, which could apply to our paper as well.

³Other papers studying communication of multidimensional information include Austen-Smith and Fryer (2005), Polborn and Yi (2006), and Egorov (2015).

tion. Dziuda and Salas (2018) study a cheap-talk model in which the receiver may learn that the sender lied, but without learning what the truth was; in their model, discovery of lies is random and exogenous, unlike ours where verification is the object of design. Deb and Stewart (2018) study an adaptive testing problem where there is a limit on the number of tests that may be performed, as in our model. There is also a growing branch of the mechanism design literature with costly verification, started by Townsend (1979), and more recently including Kartik and Tercieux (2012), Ben-Porath, Dekel, and Lipman (2014), and Erlanson and Kleiner (2017).

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, characterizing when full learning is possible. Section 4 considers robustness to several variations, including adding a condition on off-path beliefs that limits the possibility of punishing deviations by the sender, and also presents our example with noisy verification. Section 5 takes up the question of designing mechanisms in cases where full learning is impossible. 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 $a = (a_1, \dots, a_n) \in A$, 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) posterior belief $\mu \in \Delta(A)$ concerning the sender's type. We take as given a function $V : \Delta(A) \rightarrow \mathbb{R}$; $V(\mu)$ 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 that the sender gets if he induces a belief that is a point mass on a . For instance, in the job candidate example, $V(\mu)$ could represent the salary that the candidate will receive if the interviewer's posterior belief is μ (perhaps this is simply the posterior expectation of his marginal product for the firm). In the prosecution example, $V(\mu)$ 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 $\mu, \mu' \in \Delta(A)$, and μ first-order stochastically dominates μ' , then $V(\mu) \geq V(\mu')$. We also normalize $V(0) = 0$ (hereinafter, we 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 one component of the sender's type. We will assume that if the receiver chooses to verify dimension i , she then learns the value of a_i perfectly. 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.

Our primary interest is whether the receiver can fully learn the sender's type in equilibrium. This allows us to avoid having to specify a particular numerical objective for the receiver to maximize. However, in Sections 4 and 5 we will consider situations where full learning is not possible, and will discuss relevant criteria for the receiver as needed.

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;
- $\sigma : A \rightarrow \Delta(M)$ is a (possibly mixed) reporting strategy for the sender;
- p is a (possibly mixed) verification strategy for the receiver, specifying probabilities $(p_1(m), \dots, p_n(m))$ that sum to 1, for each $m \in M$ (here $p_i(m)$ is the probability of verifying dimension i);⁴
- μ is a belief system for the receiver, specifying posteriors $\mu(h) \in \Delta(A)$ for each $h \in H$, where

$$H = \{(m, i, s) \mid m \in M, p_i(m) > 0, s \in [0, \infty)\}$$

is the set of (*receiver*) *histories* that are possible given verification strategy p .

⁴Notice that this formulation *requires* the receiver to check one dimension. We could also allow for some probability $p_0(m)$ of not checking anything, at the cost of some notational inconvenience. For our main (full-learning) results, this would not help the receiver: for any full-learning mechanism that places positive probability on no verification, we could move this probability mass onto verifying dimension 1 without weakening the incentives for truthful reporting.

Note that the receiver's beliefs are defined as functions of the history; a history (m, i, s) means that 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 the mechanism is *valid* if the sender's strategy and beliefs constitute an equilibrium (more specifically, a weak PBE). That is, validity requires the following:

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

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

- *Bayesian updating*: Let $\bar{H} = A \times H$ denote the set of *full histories*, specifying both the sender's true type and the interaction with the receiver. The prior Φ and the strategies σ, p together induce a probability distribution $\bar{\zeta}$ over \bar{H} . Let ζ be the marginal distribution over H . Then, we require that for any measurable set of receiver histories $H' \subset H$ and any measurable set of types $A' \subset A$,

$$\int_{H'} \mu(h)(A') d\zeta(h) = \bar{\zeta}(A' \times H').$$

Note that we have not required incentive compatibility for the receiver's verification. This reflects the assumption that the receiver commits in advance to the verification strategy. Our definition of H in the specification of beliefs also reflects this assumption: we do not require beliefs to be defined at "histories" (m, i, s) that would be reachable only if the receiver failed to follow the verification strategy.

Bayesian updating serves to pin down beliefs at on-path histories, but imposes no constraints on beliefs at off-path histories. It will sometimes be convenient to focus on mechanisms satisfying the following:

- *Punishment beliefs*: For any receiver history $h = (m, i, s)$ that is outside the support of ζ , the belief $\mu(h)$ is a point mass on type 0.

Indeed, the usual argument shows that any outcome that can be supported by some mechanism can in particular be supported by a mechanism with punishment beliefs. (Recall that by monotonicity, the punishment belief is indeed the one with the lowest possible value of V .)

Notice, however, that punishment beliefs effectively mean that at off-path histories, the receiver does not place full faith in the accuracy of the verification technology, since if $s \neq 0$, the verification shows that the sender is not actually the zero type. We could alternatively impose the following condition:

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

For our main analysis, we will not impose this condition. The model without trusted verification has several interpretations: We could view the model as a limiting case where the receiver has infinitesimal uncertainty about the correctness of the verification technology. We could also treat it as a shorthand for a situation in which the receiver can commit to give the worst payoff $V(0)$ when the sender is known to have deviated (for example, in the employment application, we might simply imagine that the company refuses to hire a candidate who has been caught lying; in the insurance claim example, the insurance company might not be obligated to honor any claims if it has shown that one claim was false). Finally, we can also associate it with an alternative model in which the receiver can only perform verifications of the form “is a_i equal to s ?” for a specific value of s , rather than “what is a_i ?” In such a model, an off-path negative answer would generally not preclude the punishment belief that places all weight on type 0. (For brevity, we avoid writing out this alternative model in full.) In any case, in Section 4 we will consider imposing trusted verification and will show that our main conclusions are robust to it.

We are particularly interested in mechanisms that allow full learning of the sender’s type.⁵

- *Full learning*: For every type a , at every history $h \in H(a \mid \mathcal{M})$, the belief $\mu(h)$ is a point mass on type a . Here, we define

$$H(a \mid \mathcal{M}) = \{(m, i, a_i) \in H : m \in \text{supp}(\sigma(a))\},$$

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

Before moving on, we add one observation: The function $V(\cdot)$ appears only in the incentive compatibility condition, and this condition is invariant under translating the

⁵In 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 A2 in the Appendix.

whole function $V(\cdot)$ by a constant. Thus it is indeed just a normalization to assume that $V(0) = 0$.

3 Main Analysis

3.1 Initial observations

Our main question in this section is: For what payoff functions $V(\cdot)$ does there exist a mechanism that achieves full learning?

We begin with a version of the revelation principle. This shows that we can restrict attention to direct mechanisms, and also can assume punishment beliefs as described above.

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

The proofs of this and other results that are not given in the text are in the Appendix.

By focusing on direct mechanisms that furthermore satisfy punishment beliefs, we see that we need only specify the verification strategy p , since the message space, sender strategy and beliefs are pinned down. Specifically, full learning is possible if and only if there exists a choice of verification probabilities $p = (p_1, \dots, p_n)$, with each $p_i : A \rightarrow [0, 1]$ and $\sum_i p_i(a) = 1$ for all a , satisfying the incentive compatibility condition for all types a and \hat{a} :

$$V(a) \geq \left(\sum_{i: \hat{a}_i = a_i} p_i(\hat{a}) \right) V(\hat{a}). \quad (1)$$

Indeed, here the left side 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.

3.2 Additively separable case

We begin by reconsidering Example 1 from the Introduction; actually a slight formal generalization of this example. We show that the verification probabilities that allow full learning not only exist for any n , but are essentially unique.

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

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

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

Proposition 1. *Suppose that V is additively separable and defined by (2). Then, full learning is achieved by the valid direct mechanism using the verification probabilities*

$$p_i(a) = \frac{v_i(a_i)}{V(a)} = \frac{v_i(a_i)}{v_1(a_1) + \cdots + v_n(a_n)} \quad (1 \leq i \leq n)$$

for each a such that $V(a) \neq 0$ (and arbitrary verification probabilities for a such that $V(a) = 0$). Furthermore, these probabilities are unique: If $\mathcal{M} = (M, \sigma, p, \mu)$ is 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

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

Proof. For existence, we just need to check that incentive compatibility (1) is satisfied. For any \hat{a} , the right-hand side of (1) equals $\sum_{i:\hat{a}_i=a_i} v_i(\hat{a}_i)$. This is clearly at most $\sum_{i=1}^n v_i(a_i) = V(a)$, as needed.

To prove uniqueness, consider any type a , and the 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 $\sum_{i=1}^n p_i(m) = 1$, these inequalities must hold as equalities. \square

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

- The mechanism can be implemented as an indirect mechanism, as in the Introduc-

tion, where the sender chooses a probability distribution (q_1, \dots, q_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 $v_j(a_j) = \frac{q_j}{q_i} v_i(s)$ 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.
- 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 $(\lambda a_1, \dots, \lambda a_n)$ rather than (a_1, \dots, a_n) , for an arbitrary positive scalar λ .

As it turns out, all these properties (with the exception of the last one) hold quite a bit more generally.

3.3 General characterization

We now provide a necessary and sufficient condition for full learning to be achievable. For this we need a bit of notation. Whenever $S \subset \{1, \dots, n\}$ 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. Also, when S has a single element i , we will write $a|_i$ rather than $a|_{\{i\}}$.

Proposition 2. *There exists a valid mechanism 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 $S \subset \{1, \dots, n\}$ that satisfies $\sum_{S:i \in S} \lambda_S = 1$ for each index $i = 1, \dots, n$, we have*

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

To see why the characterization takes this form, consider what happens to the incentive condition (1) 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 (1) 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.

The condition in Proposition 2 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 $V(a) \leq V(a|_1) + V(a|_2)$. Indeed, in this case, the argument from the previous paragraph implies that taking $p_1(a) = V(a|_1)/V(a)$ and $p_2(a) = 1 - p_1(a)$ will suffice. The following result describes the complete set of verification probabilities.

Proposition 3. *If $n = 2$, then there is a valid mechanism that achieves full learning if and only if $V(a) \leq V(a|_1) + V(a|_2)$ for each a . A direct mechanism is valid if and only if verification probabilities satisfy, for any a :*

$$p_1(a) \geq 1 - \frac{V(a|_2)}{V(a)}, \quad p_2(a) \geq 1 - \frac{V(a|_1)}{V(a)}.$$

3.4 Submodular and supermodular functions

Here we give a couple of illustrative applications of Proposition 2.

First, suppose that the payoff function V is submodular.⁶ In this case, the condition in Proposition 2 holds, and in fact the proof of that proposition leads to a simple explicit construction for a direct mechanism with full learning, which we state as a separate result.

To state the result formally, following on our $a|_S$ notation, for each $a \in A$ and each $i = 1, \dots, n$, 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 $a|_{[0]} = (0, \dots, 0)$.

Proposition 4. *Suppose that V is submodular. Then the following valid direct mechanism achieves full learning: If $V(m) > 0$, dimension i is verified with probability*

$$p_i(a) = \frac{V(a|_{[i]}) - V(a|_{[i-1]})}{V(a)},$$

⁶The function V is *submodular* if $V(a \vee a') + V(a \wedge a') \leq V(a) + V(a')$ for all a, a' , where \vee denotes componentwise max and \wedge denotes componentwise min. V is *strictly submodular* if the inequality holds strictly whenever $\{a \vee a', a \wedge a'\} \neq \{a, a'\}$. V is *supermodular* (resp. strictly supermodular) if $-V$ is submodular (resp. strictly submodular). Additively separable functions are both sub- and supermodular.

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

On the other hand, if V is (strictly) supermodular, full learning is not achievable. For example, suppose $V(a_1, a_2) = \min\{a_1, a_2\}$. To deter deviations to reporting type $(1, 1)$, the first dimension must be tested with probability 1 (otherwise type $(0, 1)$ would misreport), but likewise the second dimension must be tested with probability 1, and we cannot do both. By the exact same reasoning, $V(a_1, a_2) = a_1 a_2$ would not allow full learning either.

In fact, we can give a broader impossibility result:

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

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

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. Note that the condition in Proposition 2 is violated, by taking $\lambda_{\{i\}} = 1$ for each i , and $\lambda_S = 0$ for all non-singleton sets. \square

For a simple 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. Whereas if the skills are substitutes (submodular case), the gains are smaller and this can be done.

For some specific, stark examples, consider first the function $V(a_1, a_2) = \max\{a_1, a_2\}$, which is submodular; then full learning is achievable simply by always testing whichever coordinate is reported higher — which is in line with Proposition 3. In contrast, in the case of function $V(a_1, a_2) = \min\{a_1, a_2\}$, which is supermodular; we show below (Proposition 16) that in a certain precisely defined sense, the receiver can do no better than learning one dimension.

3.5 Indirect mechanisms

The results presented in the preceding subsections provide a general characterization of when full learning is achievable. The construction employed a direct mechanism that, in particular, required punishing the sender with the worst possible belief in case verification failed.

In Example 1, however, we used an indirect mechanism, in which the sender effectively just reports the probability vector q by which he should be tested, and the one verified coordinate is then used to infer all other coordinates. This has a few advantages. First, essentially all histories are on-path, so we do not need to worry about the choice of off-path beliefs. Second, direct mechanisms with punishment beliefs are fragile in the sense that, if the sender's belief about his own type is off by an ε amount, the receiver ends up with a posterior that is very far from the truth; indirect mechanisms avoid this fragility, as long as V is continuous. (We will address a related topic in more detail in Subsection 4.4.) A third advantage is that, if the receiver could actually choose not to verify anything, and verifying came at a small cost $\varepsilon > 0$, then in a direct mechanism, the receiver ex-post would not have the incentive to actually carry out the verification, whereas in an indirect mechanism she could, since she is still uncertain about the type after hearing the message.

Hence, we might naturally wonder whether the indirect mechanism can be readily generalized to other $V(\cdot)$. As it turns out, it generalizes quite broadly: we can give a construction for any V that is submodular and satisfies some regularity conditions; although our construction is not quite as explicit as the one in Proposition 4 above.

Specifically, suppose that V is submodular and continuously differentiable, and write V_i for the derivative with respect to coordinate i . Suppose further that all partial derivatives V_i are bounded in an interval $[k, K]$, where $0 < k < K < \infty$. These assumptions will be maintained for the remainder of this subsection. Note that the set of submodular functions satisfying these regularity conditions is dense in the set of all increasing continuous submodular functions, in the topology of uniform convergence on compact sets.

For any vector $q = (q_1, \dots, q_n)$ of probabilities summing to 1, define a parametric curve $a(q, t) = (a_1(q, t), \dots, a_n(q, t))$ for $t \geq 0$ by the differential equations

$$\frac{\partial a_i}{\partial t} = \frac{q_i}{V_i(a(q, t))} \quad (4)$$

and the initial condition $a(q, 0) = q$. In the indirect mechanism, we simply have the agent report the probability vector q for the curve his type lies on, and the receiver verifies each dimension i with the corresponding probability q_i .

Of course, for this mechanism to be well-defined, we need to know that every possible type does indeed lie on some such curve.

Lemma 6. *For every type $a \in A$, there exist q and t such that $a(q, t) = a$.*

Note that in fact, a lies on the curve defined by q if and only if $a = a(q, V(a))$. This follows from the fact that $\frac{d}{dt}V(a(q, t)) = \sum_i \frac{\partial a_i}{\partial t} \cdot V_i(a(q, t)) = \sum_i q_i = 1$, hence $V(a(q, t)) = t$ for all t .

We have not ruled out the possibility that the type a lies on more than one such curve.⁷ (And of course this is true for $a = 0$, which lies on every curve.) In this case, we will have type a mix according to an arbitrary full-support distribution over the relevant set of curves.

If i is a coordinate such that $q_i > 0$, notice that (4), together with our bounds on derivatives, ensures that $a_i(q, t)$ is strictly increasing and goes to ∞ as $t \rightarrow \infty$. Continuity then implies that for every $s \geq 0$, there exists a unique t such that $a_i(q, t) = s$. Type $a(q, t)$ can generate the history (q, i, s) by reporting as prescribed above; thus every history in H is on-path.

In summary, the mechanism is described as follows:

- The message space consists of all probability vectors $q = (q_1, \dots, q_n)$, with $q_i \geq 0$ and $\sum_i q_i = 1$.
- For the reporting strategy, each type a uses an (arbitrary) full-support distribution over the set of q such that a lies on the curve defined by q . (This set is nonempty, by Lemma 6, and closed since it is given by the equation $a = a(q, V(a))$.)
- Given message q , the receiver verifies each coordinate i with probability q_i .
- At any history $(q, i, s) \in H$, the receiver's belief puts probability 1 on $a(q, t)$ where t is the unique value satisfying $a_i(q, t) = s$.

Proposition 7. *Suppose that V is submodular and continuously differentiable, and all partial derivatives V_i are bounded in the interval $[k, K]$. Then the indirect mechanism described above is a valid mechanism that achieves full learning.*

The proof involves a judicious use of submodularity to compare the payoff that a type not on the $a(q, \cdot)$ curve would get by reporting probability vector q against this type's equilibrium payoff, and to show that there is no gain from deviating.

⁷For $n = 2$, it is easy to show that the curves do not intersect except at $a = 0$.

4 Robustness

4.1 Trusted verification

As mentioned in Section 2, it is natural to consider imposing the trusted verification condition as a restriction on beliefs when the sender is found to have misreported. How much do our results change under this restriction?

First, our major qualitative conclusions remain unchanged. In particular, full learning is still possible whenever $V(\cdot)$ is submodular, although the explicit mechanism from Proposition 4 no longer works,⁸ and indeed, we do not know of a similarly simple explicit formula for a mechanism that works in general. Actually, when V satisfies the regularity assumptions of Proposition 7, that proposition already shows that full learning is possible; notice that trusted verification is automatically satisfied since every history $h = (m, i, s)$ is on-path. But even for submodular functions that fail those regularity assumptions, full learning is possible:

Proposition 8. *Suppose that V is submodular. Then there exists a valid direct mechanism that achieves full learning and satisfies trusted verification.*

The proof of Proposition 8 is nonconstructive. As before, the idea is to find verification probabilities $p(a)$ for any fixed $a \in A$ that deter any other type $z \neq a$ from deviating by reporting type a , and we now use the Kakutani fixed-point theorem to show that such probabilities exist. More specifically, for any verification probability vector p , we consider the set of types $z \leq a$ that would gain the most from misreporting as a . We then let E_p be the set of all alternative verification probability vectors that would successfully deter these types from deviating. This set is quickly shown to be nonempty using the submodularity of V . It turns out that the correspondence $p \mapsto E_p$ is not upper-hemicontinuous, so we cannot apply the Kakutani fixed-point theorem immediately, but we can “smooth out” E_p appropriately to yield a correspondence for which the theorem does apply. Taking p to be a fixed point, then, all types that would gain the most from deviating under p are deterred from deviating, which is exactly what we need.

Our other main conclusion from Section 3 was that strictly supermodular V does not allow full learning (Proposition 5). Clearly this conclusion also still holds up when we restrict mechanisms by requiring trusted verification.

⁸For example, take $n = 2$ and $V(a_1, a_2) = \max\{a_1, a_2\}$. Then under the verification strategy from Proposition 4, type $(1, 1)$ can strictly gain from reporting as type $(1, 2)$ if the receiver’s beliefs upon detecting the lie are constrained by trusted verification.

With trusted verification, we do not know of a complete characterization of the functions $V(\cdot)$ for which full learning is possible. However, we do have such a characterization for the two-dimensional case:

Proposition 9. *Suppose that $n = 2$. Then full learning is achievable with a valid mechanism satisfying trusted verification if and only if V satisfies the following property: for any two types $x, a \in A$ with $x < a$, we have*

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

The proof in fact gives an explicit construction of a full-learning mechanism when the condition is satisfied. One can also use this result to construct functions for which full learning is possible without the trusted verification requirement, but not possible with it, thus showing that the restriction on beliefs does have bite.

4.2 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 10. *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 $U : [0, \infty) \rightarrow [0, \infty)$ is any increasing, concave transformation.*

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 $a_1 + \dots + a_n$, 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.

4.3 Payoffs depending on sender's type

We have so far assumed that the sender's payoff $V(a)$ (or, more generally, $V(\mu)$) depends only on the receiver's posterior belief, but not on the sender's true type. In some natural cases, however, this assumption ought to be relaxed. To allow for these possibilities, let us denote the payoff of a sender of type x if the receiver's posterior is that he is of type a by $V(a; x)$. We have the following result.

Proposition 11. *Suppose that $V(a; a)$ satisfies the condition in Proposition 2 and $V(a; x) \leq \max\{V(a; a), V(x; x)\}$ for all a and x , and furthermore, $V(0; x) = 0$ for every x . Then there is a valid direct mechanism that achieves full learning.*

Proposition 11 says that the previous results generalize as long as type x pretending to be type a cannot be better off than both types x and a telling the truth. This would be the case, for example, if there is a cost of lying: in this case, $V(a; x) \leq V(a; a)$ (and in particular, $V(0; x) \leq V(0; 0) = 0$) and the condition holds. But the assumption is more general than that: suppose, for example, that for the deception to continue, a lower type must exert extra effort, while a higher type can afford to slack. For example, let $V(a; x) = \sum_{i=1}^n a_i - \sum_{i=1}^n \kappa(a_i - x_i)$, so the cost of effort is proportionate to the difference between the receiver's belief that the sender needs to maintain and the true dimension, with coefficient $\kappa \in (0, 1)$ (and suppose that the worst type 0 is never hired, so $V(0; x) = 0$ for all x). Then

$$\begin{aligned} V(a; x) &= \sum_{i=1}^n (1 - \kappa) a_i + \sum_{i=1}^n \kappa x_i = (1 - \kappa) V(a; a) + \kappa V(x; x) \\ &\leq \max\{V(a; a), V(x; x)\}, \end{aligned}$$

so the condition is satisfied. This Proposition also shows the limits of the argument: for example, if $\kappa > 1$, then for a high type the temptation to pretend to be a low type and save on effort would be too high; in that case, clearly, full learning would not be feasible.

4.4 Noisy 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 . Suppose, however, that if dimension i is verified, the receiver gets a noisy signal $s \in [0, \infty)$, drawn from a full-support distribution $\rho_i(s|a_i)$; this formulation is natural in many settings. 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. But how robust are our results on full learning?

In such a setting, it is easy to see that full learning is not possible. Indeed, if any signal s can occur in equilibrium if dimension i is verified, then the Bayesian property implies that the receiver only uses the sender's message to infer his type, which obviously creates room for manipulation. Thus, the right question to ask is how close we can get to full learning. We cannot use direct mechanisms (since any such mechanism would necessarily imply full learning), so we have to focus on indirect ones. Below, for specificity we return to our canonical example with $V(a) = \sum_{i=1}^n a_i$, and show that an appropriate modification of the “relative skills” mechanism from that example can still perform well; in fact, it can approximate full learning as the level of noise goes to zero. For this we make some specific distributional assumptions.

We define Φ , the distribution of sender's types, as follows. Let (ν_1, \dots, ν_n) be a vector of any real numbers, (τ_1, \dots, τ_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 $\{z_i\}_{i \in \{1, \dots, n\}}$ are independent. Refer to this distribution of z as Λ . Now 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 $K \rightarrow \infty$. Now, we define distribution Φ as Λ , restricted to the cone A_K . (This restriction will help simplify the description of the mechanism in the result below, 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 $\chi \rightarrow \infty$ 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 $V(\mu)$ is just the posterior mean of $V(a)$, and the latter is a concave power function of the “total ability”: $V(\mu) = \mathbb{E}_{a \sim \mu} [(\sum_i a_i)^\gamma]$ for $\gamma \in (0, 1]$. (Thus we slightly generalize the setting of Example 1, where $\gamma = 1$.)

For a final bit of notation, recall that for any mechanism $\mathcal{M} = (M, \sigma, p, \mu)$, $\mu(h)$ is

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 may generate in equilibrium (see the Appendix for formalities). We show that, for an appropriately constructed family of mechanisms (indexed by χ), the corresponding distributions $\kappa^\chi(a)$ converge to a for all $a \in A_K$ as the noise disappears:

Proposition 12. *There exist a set of mechanisms $\{\mathcal{M}^\chi\}_{\chi>K}$, where $\mathcal{M}^\chi = (M^\chi, \sigma^\chi, p^\chi, \mu^\chi)$, satisfying the following: 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 $p_i^\chi(m) = m_i \left(1 + \frac{\tau_i}{\chi}\right) - \frac{\tau_i}{\chi}$. Mechanism \mathcal{M}^χ is valid when the signal precision is χ .*

Under these mechanisms, for any $a \in A_K$, the corresponding probability measure $\kappa^\chi(a)$ converges in distribution to an atom on a as $\chi \rightarrow \infty$. 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 $\Pr_{x \sim \kappa^\chi(a)}(\max_i |x_i - a_i| > \varepsilon) < \delta$.

Notice that we do not claim that the mechanisms constructed, $\{\mathcal{M}^\chi\}$, are optimal in any sense; what is important is that these are valid mechanisms that achieve convergence of the posterior distributions. One notable new feature is that in these mechanisms, dimension i is less likely to be tested if τ_i is high, i.e., if the variance $\frac{1}{\tau_i}$ of the prior distribution of a_i is low. The reason is intuitive: in this case, the receiver knows a_i quite well without testing, and thus testing that dimension is not as useful.

5 Imperfect Learning

5.1 Motivating example

Our inquiry so far was focused on the cases where full learning is possible, and Proposition 2 gave a complete characterization. But for some important classes of functions full learning is not achievable, e.g. supermodular ones. It is natural to ask how much the receiver can learn using a valid mechanism.

This is not an easy question with a simple universal answer, as one would see below. The first complication, that we have avoided so far, is that the receiver’s posterior beliefs (and thus the sender’s incentives and thereby the solution to the receiver’s problem) would depend on the distribution of sender’s types. The second is that we would have to define

a specific objective function for the receiver: for example, some receivers might want to be guaranteed to learn the sender's type almost perfectly, while others might prefer full learning for some types of senders at the expense of little learning for others. Mechanisms that achieved full learning were robust both to the distribution of senders and to the receiver's objective as long as she valued more information; in this Section this will not be the case.

To make progress, we assume $n = 2$ for now. Also, for the remainder of this Section, we will define $V(a)$ for individual types a , and will assume that V is extended to non-degenerate beliefs by $V(\mu) = \mathbb{E}_{a \sim \mu}[V(a)]$.⁹

We find it helpful to start by identifying the obstacles to existence of valid mechanisms with full learning. Recall that according to Proposition 3, full learning is achievable for $n = 2$ if and only if for every $a = (a_1, a_2)$, $V(a_1, 0) + V(0, a_2) \geq V(a)$. Intuitively, if this does not hold, then types $(a_1, 0)$ and $(0, a_2)$ have too strong incentive to pretend to be type a , and deterring them with just one test is impossible. This is an important insight, and we will use it to construct valid mechanisms with only one test by improving the payoffs of types of the form $(a_1, 0)$ and $(0, a_2)$ by pooling them with types with higher values of V . The following example previews our results and is important for what follows.

Example 2. Define, for each value of $t \in [0, 1]$, function

$$Z_t(a) = (1 - t)(a_1 + a_2) + t \min\{a_1, a_2\} \quad (5)$$

(it may also be written as $a_1 + a_2 - t \max\{a_1, a_2\}$). Let Φ be the uniform distribution on $[0, L] \times [0, L]$ for some finite $L > 0$, and let ϕ be the corresponding density function. Function $Z_t(a)$ is strictly supermodular unless $t = 0$, and thus full learning is not achievable when $V \equiv Z_t$.

However, consider the following mechanism. Types a with $a_2 < ta_1$ send message $(a_1, *)$ (where $*$ is a special symbol; we use it to distinguish one-dimensional messages that identify the first coordinate from those that identify the second coordinate). Upon receiving such a message, the receiver verifies dimension 1 for sure. Similarly, types a with $a_1 < ta_2$ send message $(*, a_2)$; upon receiving such a message, the receiver verifies dimension 2 for sure. Lastly, any type a with $\min\{a_1, a_2\} \geq t \max\{a_1, a_2\}$ reveals his type truthfully, whereas the receiver who got message (a_1, a_2) verifies the two dimensions with probabilities $p_1(a) = \frac{a_1}{a_1 + a_2}$ and $p_2(a) = \frac{a_2}{a_1 + a_2}$ (assuming $a \neq 0$; that type may have any verification probabilities). In this mechanism, full learning is achieved for measure $1 - t$ of types.

⁹If V is unbounded, we may have $V(\mu) = \infty$ for some μ , but such beliefs will not actually arise in our arguments.

Let us show that this mechanism (with corresponding beliefs¹⁰) is valid. In equilibrium, the types that pool to send message $(a_1, *)$ (that is, types a with $a_2 \in [0, ta_1]$) get payoff of $a_1 \left(1 - \frac{t}{2}\right)$. Such a type cannot benefit by sending $(x, *)$ with $x \neq a_1$ (he will get caught for sure and get payoff 0) or $(*, y)$ for $y \neq a_2$ (for the same reason). If he reports a_2 truthfully (sends message $(*, a_2)$), he would be pooled with types who send this message in equilibrium, and get a payoff of $a_2 \left(1 - \frac{t}{2}\right)$, but this is not profitable, since $a_1 \geq a_2$. Now suppose the sender decides to mimic a type that reveals fully in equilibrium; this can only be worthwhile if he reports truthfully in one of the coordinates. Suppose he reports (a_1, y) ; then it must be that $y \geq ta_1$ (otherwise such message is not allowed). He passes with probability $\frac{a_1}{a_1+y}$ and gets caught otherwise, so his expected payoff is

$$\frac{a_1}{a_1+y} (a_1 + y - t \max\{a_1, y\}) = a_1 \left(1 - t \frac{\max\{a_1, y\}}{a_1+y}\right) \leq a_1 \left(1 - \frac{t}{2}\right),$$

where we used $\max\{a_1, y\} \geq \frac{a_1+y}{2}$. If he reports (x, a_2) instead (in which case $x \geq ta_2$ for similar reasons), then his expected payoff is

$$\frac{a_2}{x+a_2} (x + a_2 - t \max\{x, a_2\}) \leq a_2 \left(1 - \frac{t}{2}\right) < a_1 \left(1 - \frac{t}{2}\right).$$

In either case, the deviation is not profitable. Types that report their second dimension by sending message $(*, a_2)$ in equilibrium do not have a profitable deviation either.

The last thing to check is that types that are supposed to reveal fully do not want to pretend to be some other type. Consider type a ; clearly (by monotonicity) he would get less by sending one dimension only, $(a_1, *)$ or $(*, a_2)$. Suppose he benefits from deviating to (a_1, y) that also reveals fully. Then type $(a_1, 0)$ should benefit even more from reporting (a_1, y) , as he has the same probabilities of passing and getting caught, but lower equilibrium payoff. However, we already showed that such type does not have a profitable deviation. This shows that the described mechanism is valid.

In Example 2, one can think of t as a “measure of supermodularity”, with additively separable function $Z_0(a) = a_1 + a_2$ for $t = 0$. Then the interpretation of this example is that if a supermodular function is sufficiently close to an additively separable one, then full learning is achievable for a sufficiently large set of types (this robustness result is true

¹⁰To be precise, Bayesian updating pins down beliefs only a.e., not everywhere, since conditional expectations are defined only up to measure-zero events. However, here and for all mechanisms considered throughout this section, there is a clear “natural” choice of beliefs at all on-path histories (distributed over the relevant line segment with density induced by ϕ), so we assume without further comment that these beliefs are employed.

more generally, see Proposition 14). On the other hand, for supermodular functions that are not “close” to additively separable ones, less learning may be achieved.

We next show that Example 2 is considerably more general.

5.2 Valid mechanisms with partial learning

Fix a payoff function V and distribution function Φ that has strictly positive density ϕ on the support $\Omega = [0, L] \times [0, L]$ for some $L > 0$ ($L = \infty$ is admissible, too). We define

$$\eta = \inf \{ \eta' \geq 0 : \forall (a_1, a_2) \in \Omega : V(a_1, \eta' a_1) + V(\eta' a_2, a_2) \geq V(a_1, a_2) \}; \quad (6)$$

$$\xi = \inf (\{ \xi' \geq 0 : \forall a_1 \in (0, L) : \mathbb{E}_{0 \leq a_2 \leq \xi' a_1} V(a_1, a_2) > V(a_1, \eta a_1) \} \cup \{1\}), \quad (7)$$

where $\mathbb{E}_{0 \leq a_2 \leq z} V(a_1, a_2) = (\int_0^z V(a_1, a_2) \phi(a_1, a_2) da_2) / (\int_0^z \phi(a_1, a_2) da_2)$. These values are well-defined: the set in (6) is nonempty as it contains $\eta' = 1$ by monotonicity of V , which also implies $\eta \leq 1$, and the set in (7) is nonempty because it is defined so as to contain 1. Notice that we must have $\xi \geq \eta$ (this follows from $\eta \leq 1$ and the fact that $\mathbb{E}_{0 \leq a_2 \leq \xi' a_1} V(a_1, a_2) \leq V(a_1, \eta a_1)$ for $\xi' < \eta$ by monotonicity, so the set in (7) only contains elements with $\xi' \geq \eta$). As an example, if V is submodular or additively separable, then $\eta = 0$, and $\xi = 0$ also as long as V is strictly increasing (this excludes boundary cases such as $V(a_1, a_2) = \max\{a_1, a_2\}$). If $V = Z_t$ (defined in (5)) with $t > 0$, then $\eta = \frac{t}{2}$; for $V = a_1 a_2$, $\eta = \frac{1}{2}$. Moreover, in both these cases, if Φ is uniform, then $\xi = \min\{2\eta, 1\}$.

Say that payoff function V and distribution function Φ satisfy the *monotone expectation property* if $\mathbb{E}_{0 \leq a_2 \leq z} V(a_1, a_2)$ is nondecreasing in a_1 for any z . In other words, if an individual’s second dimension is known not to exceed some z , he is never worse off if he is revealed to have a high first dimension rather than a low one. For example, this property is automatically satisfied if Φ has the two coordinates distributed independently.

The following result generalizes Example 2 beyond the case of $V = Z_t$ and Φ uniform.

Proposition 13. *Suppose that V is symmetric and continuous and Φ is symmetric and has a positive density $\phi(a)$, and the monotone expectation property is satisfied. There exists a valid mechanism with the following properties. Types (a_1, a_2) such that $a_2 \leq \xi a_1$ report $(a_1, *)$, and dimension 1 gets verified with probability 1. Types (a_1, a_2) such that $a_1 < \xi a_2$ report $(*, a_2)$, and dimension 2 gets verified with probability 1. All other types (a_1, a_2) reveal fully, and verification probabilities following report (a_1, a_2) satisfy $p_1(a_1, a_2) \geq 1 - \frac{V(\eta a_2, a_2)}{V(a_1, a_2)}$ and $p_2(a_1, a_2) \geq 1 - \frac{V(a_1, \eta a_1)}{V(a_1, a_2)}$.*

We now proceed with formulating a limit result, stating that if V is “sufficiently close” to a submodular (in particular, additively separable) function, then almost full learning

(in the sense of full learning for an arbitrarily large set of sender's types) is achievable.

Proposition 14. *Fix distribution Φ on $[0, \infty) \times [0, \infty)$, with a symmetric continuous density ϕ . Let $\bar{V}(a_1, a_2)$ be a submodular symmetric continuous function that is strictly increasing in a_1 at $a_1 = 0$, and satisfies the monotone expectation property with respect to Φ . Take any symmetric increasing function $V(a_1, a_2)$ and consider a family of payoff functions given by $\tilde{V}_t(a_1, a_2) = (1 - t)\bar{V}(a_1, a_2) + tV(a_1, a_2)$ for $t \in [0, 1]$. Then for any $\delta > 0$ there is $T = T(\delta) \in (0, 1)$ such that for all $t \in (0, T)$, there is a mechanism that achieves full learning for a measure of types at least $1 - \delta$.*

5.3 Optimal mechanisms

In Proposition 14, we proved a limit result, stating that almost full learning is achievable but without any claim to optimality of that particular mechanism. We did not discuss the question of optimality of the mechanism constructed in Example 2 or its extension, Proposition 13, either. We now partially address this gap by showing that, in certain cases and within a natural class of mechanisms, this construction is indeed optimal. We define this class for an arbitrary number of dimensions n , although we will return to $n = 2$ for our optimality results.

Specifically, we will call mechanism \mathcal{M} *semi-direct* if the sender is supposed to truthfully report some (nonempty) subvector of his type, and the receiver verifies some dimension that was reported. More precisely, \mathcal{M} is semi-direct if:

- M is a subset of the space $([0, \infty) \cup \{*\})^n \setminus \{(*, \dots, *)\}$.
- The reporting strategy σ is deterministic. Moreover, for each type a and each component i , either $\sigma_i(a) = a_i$ or $\sigma_i(a) = *$.
- For each message $m \in M$ and each i such that $m_i = *$, $p_i(m) = 0$.

The mechanism in Example 2 is an example of a semi-direct mechanism.¹¹ We view the definition as a relatively tractable and interpretable generalization of this example.

An additional strength of semi-direct mechanisms is the following. Recall that in Section 2, we briefly mentioned an alternative model in which the receiver can perform

¹¹According to our definition, no information is allowed to be transmitted along the dimensions that are not revealed. We could imagine a variant in which the sender is offered multiple ways to express a non-revealed (and non-verified) coordinate, but such a variant does not seem to expand the possibilities: for example, if types $(1, y)$ for $y \in [0, 1]$ sent message $(1, *)$, and those with $y \in [2, 3]$ sent message $(1, \diamond)$, then the former would always be able to pass as the latter. This means that if both messages were used in equilibrium, the corresponding sets of types would receive the same payoffs; but then we may as well pool these types.

verifications by asking yes-no questions of the form “is a_i equal to s ?” rather than asking “what is the value of a_i ?” Any valid semi-direct mechanism would work equally well in this alternative model, whereas most other indirect mechanisms would not. For instance, Example 2 has an indirect implementation in the original model in which the sender calculates $p = \frac{a_1}{a_1+a_2}$, reports p if $\frac{t}{1+t} \leq p \leq \frac{1}{1+t}$, and reports “low” or “high” if $p < \frac{t}{1+t}$ or $p > \frac{1}{1+t}$ respectively; but this indirect implementation would fail in the alternative model.

We will call a semi-direct mechanism \mathcal{M} *connected* if, for each message m that is used in equilibrium, the set of types a for which $\sigma(a) = m$ is a connected subset of A . Again, the mechanism in Example 2 satisfies this property.

Lastly, given a valid mechanism \mathcal{M} , define $U(a)$ to be the payoff received by type a in equilibrium. (For notational brevity, we suppress the fact that $U(a)$ depends on the mechanism.) In particular, for types a that are supposed to reveal fully, $U(a) = V(a)$.

We now specify the receiver’s objective as minimizing

$$\mathcal{W}(\mathcal{M}) = \mathbb{E} [|V(a) - U(a)|].$$

We are now ready to formulate the following proposition.

Proposition 15. *Assume $n = 2$. Take $t \in (0, \sqrt{\frac{1}{2}})$ and let $V(a) = Z_t(a)$ and Φ be the uniform distribution on the square $[0, L] \times [0, L]$ for some $L > 0$. Then the mechanism \mathcal{M}_t constructed in Example 2 minimizes $\mathcal{W}(\mathcal{M})$ within the class of connected semi-direct mechanisms (and this minimum equals $\mathcal{W}(\mathcal{M}_t) = \frac{1}{6}t^2L$).*

Proposition 15 suggests that building on the intuition from the main model in Section 3, we can construct optimal mechanisms for supermodular functions, at least under some assumptions and within a certain class. We should note that for supermodular functions that are not close to linearly additive, the optimal mechanism may look very differently. For example, the mechanism from Example 2 for $t = 1$ would correspond to the sender revealing the value of the largest dimension; indeed this is a valid mechanism that achieves $\mathcal{W}(\mathcal{M}_1) = \frac{1}{6}L$. However, the optimal mechanism looks different, as the next result shows.

Proposition 16. *Let $V(a_1, a_2) = \min\{a_1, a_2\}$ and let Φ be uniform distribution on the square $[0, L] \times [0, L]$ for some $L > 0$. Then there are exactly two mechanisms (up to differences on measure-zero sets) that minimize $\mathcal{W}(\mathcal{M})$ within the class of connected semi-direct mechanisms. One involves all sender types sending messages $(a_1, *)$, and the other involves all types sending messages $(*, a_2)$. In both mechanisms, $\mathcal{W}(\mathcal{M}) = \frac{2}{15}L$.*

We finish this subsection by noting that the optimality results here are sensitive to distributions and objective functions, as well as to the class of mechanisms that we allow.

For example, for $t \in \left(0, \sqrt{\frac{1}{2}}\right)$, one can do better than Proposition 15 in terms of $\mathcal{W}(\mathcal{M})$ using semi-direct but not necessarily connected mechanisms. For $V(a_1, a_2) = \min\{a_1, a_2\}$, the optimal mechanism from Proposition 16 does not achieve full learning for any type, so if the objective is instead to maximize the share of types for which full learning is achieved, one can do better. These facts (whose proofs are omitted for space, but are available from the authors) highlight that while the methods of this paper are helpful to construct valid mechanisms, the question of optimality is far from trivial and requires further inquiry.

5.4 Testing multiple dimensions

In the above discussion, we have studied situations where the type cannot be fully learned by verifying just one dimension, and asked how much can be learned. An alternative extension of our model to such situations would instead ask how many verifications are needed if we insist on achieving full learning. We consider this version very briefly. We return to the setting of general n (since $n = 2$ is not interesting if more than one verification is allowed).

Example 3. Suppose that the type is n -dimensional, and $V(x)$ is given by the k -th highest coordinate of x , where k is constant, $1 \leq k \leq n$.

If the receiver can test k dimensions, then full learning is possible. Just take a direct mechanism, where the sender reports his type, and the receiver deterministically tests the highest k reported coordinates (with ties broken lexicographically). This is incentive-compatible: if the sender lies, the k -th highest of the measured coordinates is no higher than the k -th highest of the true coordinates, and all other coordinates are believed to be lower than the k -th highest measured value, so the lie cannot be profitable.

If the receiver can test at most $k - 1$ dimensions, then full learning is not possible. Just consider the type $x = (1, 1, \dots, 1, 0, \dots, 0)$ with k 1's. This type has $V(x) = 1$. For any type y obtained by replacing the 1 in some position i by a 0, we have $V(y) = 0$, so y can gain by misreporting as x unless coordinate i is tested with probability 1. Hence, if the sender reports x , each of the first k dimensions needs to be tested with probability 1, otherwise full learning is not achieved.

More generally, one can ask the question about the minimal number of dimensions to test in order to achieve full learning. The next proposition gives a lower bound for the number of dimensions.

Proposition 17. *For every type a let $k(a)$ be the solution to the following linear programming problem:*

$$\begin{aligned} & \min p_1 + \dots + p_n \\ \text{s.t. } & \forall S \subset \{1, \dots, n\}, S \neq \emptyset : V(a|_S) \geq \left(1 - \sum_{j \notin S} p_j\right) V(a). \end{aligned}$$

Then if k is such that $k < k(a)$ for some a , there is no valid mechanism with k or fewer verifications that achieves full learning.

Proposition 17 generalizes the insight that deviations are deterred by high probabilities of getting caught, and if these probabilities cannot be “packaged” in k tests, then k tests cannot achieve full learning. Unlike Proposition 2, there is no sufficiency result; an exact characterization of the number of tests needed would require more understanding of how to optimality correlate the tests across dimensions.

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 for randomizations over tests, the receiver may learn the sender’s type fully, for a wide range of the sender’s objective functions. While the main focus of the theoretical results has been on direct mechanisms, we also showed that in the main leading cases, full learning is possible using an indirect mechanism in which the sender just chooses the probability of testing each dimension.

While the paper’s main contribution is theoretical, we believe it has practical take-aways. In our view, the indirect mechanism, where the sender suggests probabilities to verify each dimension, is not so far from the structure of interactions that may occur in practice. 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 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. Similarly, the idea of drawing inference from choice of tests has recently gained attention in the insurance literature, see Crocker and Zhu (2018).

The paper's results suggest many interesting directions for further inquiry. One question that we have only begun to address is how to identify optimal mechanisms (or even how to formulate the optimality problem in a tractable way) when full learning is impossible. For another, 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, 2018, who study a similar question with a one-dimensional type space).

We have seen that the basic insights from our full-learning setting have some application even when full learning is not achievable. We may hope that these insights eventually contribute to addressing some of these other directions as well.

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] Azar, Pablo, and Silvio Micali (2018), “Computational Principal-Agent Problems,” *Theoretical Economics*, 13(2): 553–578.
- [5] Battaglini, Marco (2002), “Multiple Referrals and Multidimensional Cheap Talk,” *Econometrica*, 70: 1379–1401.
- [6] Ben-Porath, Elchanan, Eddie Dekel, and Barton L. Lipman (2014), “Optimal allocation with costly verification,” *American Economic Review*, 104(12): 3779–3813.
- [7] Chakraborty, Archishman, and Rick Harbaugh (2007), “Comparative Cheap Talk,” *Journal of Economic Theory*, 132(1): 70–94.
- [8] Chakraborty, Archishman, and Rick Harbaugh (2010) “Persuasion by Cheap Talk,” *American Economic Review*, 100(5): 2361–2382.
- [9] Crocker, Keith, and Nan Zhu (2018), “The Efficiency of Voluntary Risk Classification in Insurance Markets,” unpublished paper.
- [10] Deb, Rahul, and Colin Stewart (2018), “Optimal Adaptive Testing: Informativeness and Incentives,” *Theoretical Economics*, 13(3): 1233–1274.
- [11] Dziuda, Wioletta, and Christian Salas (2018), “Communication with Detectable Deceit,” unpublished paper.
- [12] Egorov, Georgy (2015), “Single-issue Campaigns and Multidimensional Politics,” NBER working paper No. w21265.
- [13] Erlanson, Albin, and Andreas Kleiner (2017), “Costly Verification in Collective Decisions,” unpublished paper.
- [14] Glazer, Jacob, and Ariel Rubinstein (2004), “On Optimal Rules of Persuasion,” *Econometrica*, 72(6): 1715–1736.

- [15] Holmström, Bengt (1977), “On Incentives and Control in Organizations,” Ph.D. Thesis, Stanford University.
- [16] Jamison, Robert E., and William H. Ruckle (1976), “Factoring Absolutely Convergent Series,” *Mathematische Annalen*, 224 (2): 143–148.
- [17] Kartik, Navin, and Olivier Tercieux (2012), “Implementation with Evidence,” *Theoretical Economics*, 7(2): 323–355.
- [18] Lipnowski, Elliot, and Doron Ravid (2017), “Cheap Talk with Transparent Motives,” unpublished paper.
- [19] Meyer, Margaret, Inés Moreno de Barreda, and Julia Nafziger (2016), “Robustness of Full Revelation in Multisender Cheap Talk,” unpublished paper.
- [20] Polborn, Mattias K., and David T. Yi (2006), “Informative Positive and Negative Campaigning,” *Quarterly Journal of Political Science*, 1(4): 351–371.
- [21] Ray, Debraj & Arthur Robson (2018), “Certified Random: A New Order for Coauthorship,” *American Economic Review*, 108(2): 489–520.
- [22] 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.
- [23] Townsend, Robert (1979), “Optimal Contracts and Competitive Markets with Costly State Verification,” *Journal of Economic Theory*, 21(2): 265–293.

The Appendix contains the proofs of all results that are not proved in the main text. It is organized in sections, corresponding to sections of the main text.

A Proofs for Section 3

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. Beliefs μ' are uniquely determined by the criteria of full learning and punishment beliefs: at histories (a, i, a_i) that can be generated by truthful reporting, the belief is degenerate on type a ; at other histories (a, i, s) , it puts probability 1 on type 0.

It is immediate from the construction that the mechanism satisfies full learning and punishment beliefs. To see that the mechanism is valid, we check the two conditions. 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

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

Here the first inequality follows from the fact that for each $m \in \text{supp}(\sigma(\hat{a}))$ and each i , 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 in the original mechanism, or $\hat{a}_i \neq a_i$ and $V(\mu'(\hat{a}, i, a_i)) = V(0) = 0$ by construction. The second inequality comes from incentive compatibility of the original mechanism.

Finally, Bayesian updating is immediate, since in equilibrium, with ex ante probability 1, the receiver puts probability 1 on the true type, which equals the report. \square

Proof of Proposition 2. First we show necessity. Take the weights λ_S as given; we can assume $\lambda_\emptyset = 0$, since the value of λ_\emptyset 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 $(\sum_{i \in S} p_i(a)) V(a)$. Hence, incentive compatibility implies

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

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

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

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, \dots, r_n such that $r_1 + \dots + r_n = V(a)$ and, for each subset $S \subset \{1, \dots, n\}$, $\sum_{i \in S} r_i \leq V(a|_S)$.

Suppose not. Then, applying a theorem of the alternative, we get the existence of nonnegative numbers λ_S , for each $S \subset \{1, \dots, n\}$, 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 $S \setminus \{i\}$. 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, \dots, r_n exist. Define the verification probabilities by $p_i(a) = r_i/V(a)$. We just need to check incentive compatibility condition (1).

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 (1) requires. \square

Proof of Proposition 3. The fact that the given condition is necessary and sufficient to achieve full learning follows from the discussion immediately preceding the proposition statement. Similarly, this discussion shows that a direct mechanism is valid if and only if types $a|_1$ and $a|_2$ are both deterred from reporting as type a , for each a . The relevant incentive constraints are $V(a|_1) \geq p_1(a)V(a) = (1-p_2(a))V(a)$ and $V(a|_2) \geq p_2(a)V(a) = (1-p_1(a))V(a)$, which are equivalent to the conditions given in the proposition statement. \square

Proof of Proposition 4. 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

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

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

Proof of Lemma 6. It is not hard to see that $a(q, t)$ is continuous in q . Now, consider any $t > 0$. As noted in the main text, $V(a(q, t)) = t$ for any probability vector q . Let Δ_n denote the probability simplex $\{(q_1, \dots, q_n) \mid q_i \geq 0 \text{ and } \sum_i q_i = 1\}$. Define a function $G : A \setminus \{0\} \rightarrow \Delta_n$ by rescaling: $G_i(a_1, \dots, a_n) = a_i / (a_1 + \dots + a_n)$. Now define $F : \Delta_n \rightarrow \Delta_n$ by $F(q) = G(a(q, t))$. This is a continuous map from the simplex Δ_n to itself. Moreover, for each coordinate i , it sends the face $q_i = 0$ of the simplex to itself, since $q_i = 0$ implies $\partial a_i(q, t) / \partial t = 0$ and therefore $a_i(q, t) = 0$ for each t . A result in topology (e.g. Jamison and Ruckle 1976, Lemma 2.1) then implies that F is surjective.

Now to prove the lemma, consider any type $a \neq 0$ (the lemma statement is trivial for $a = 0$) and put $t = V(a)$ in the above. So by surjectivity, there exists some q such that $G(a(q, t)) = F(q) = G(a)$, or equivalently, $a(q, t) = \lambda a$ for some $\lambda > 0$. Moreover, as noted in the main text, $V(a(q, t)) = t$. Thus, combining, we get $V(a) = t = V(a(q, t)) = V(\lambda a)$. However, since V is strictly increasing, this equality can only hold if $\lambda = 1$. Thus we have

shown existence of q and t satisfying $a(q, t) = a$, proving the lemma. \square

We now give a proof of Proposition 7. After the general proof, we give a more explicit version for the case $n = 2$ to illustrate more clearly the role of the submodularity condition. The key to the proof of Proposition 7 is the lemma below:

Lemma A1. *Let $W : [0, \infty)^n \rightarrow \mathbb{R}$ be submodular and continuously differentiable. Write W_i for the derivative with respect to coordinate i . Suppose, moreover, that $W_i(t, t, \dots, t) = 0$ for all t and each i .*

Then, for any $t_1, \dots, t_n \in [0, \infty)$ and for each coordinate index i ,

$$W(t_i, t_i, \dots, t_i) \leq W(t_1, t_2, \dots, t_n).$$

Proof. It suffices to prove the lemma under the assumption that $t_1 \leq t_2 \leq \dots \leq t_n$; the general statement will then follow by permuting coordinates. Also, it suffices to prove the lemma for $i = n$, and the statement for any other i will follow. This is because $W(t, t, \dots, t)$ is constant as a function of t (since its total derivative with respect to t is $\sum_i W_i(t, t, \dots, t) = 0$).

Define a sequence of n -vectors by

$$\begin{aligned} v_1 &= (t_1, t_2, t_3, \dots, t_{n-1}, t_n) \\ v_2 &= (t_2, t_2, t_3, \dots, t_{n-1}, t_n) \\ v_3 &= (t_3, t_3, t_3, \dots, t_{n-1}, t_n) \\ &\vdots \\ v_n &= (t_n, t_n, t_n, \dots, t_n, t_n). \end{aligned}$$

Now, for each i with $1 \leq i < n$,

$$\begin{aligned} W(v_{i+1}) - W(v_i) &= \int_{t_i}^{t_{i+1}} \left[\frac{d}{dt} W(\underbrace{t, t, \dots, t}_i, t_{i+1}, \dots, t_n) \right] dt \\ &= \int_{t_i}^{t_{i+1}} \left[\sum_{j=1}^i W_j(\underbrace{t, t, \dots, t}_i, t_{i+1}, \dots, t_n) \right] dt \\ &\leq \int_{t_i}^{t_{i+1}} \left[\sum_{j=1}^i W_j(t, t, \dots, t) \right] dt \\ &= 0. \end{aligned}$$

Here, the inequality holds because submodularity implies that each term W_j increases when the k -th argument (for $k > i$) is decreased from t_k to $t \leq t_{i+1}$.

Consequently,

$$W(v_n) \leq W(v_{n-1}) \leq \dots \leq W(v_2) \leq W(v_1),$$

which is exactly what we wanted. \square

Proof of Proposition 7. Full learning and Bayesian updating are immediate, so we just need to check that incentive compatibility is satisfied. That is, for any probability vector $q = (q_1, \dots, q_n)$, we check that no type a would gain by reporting q instead of following his intended reporting strategy.

Now, for any nonnegative numbers t_1, \dots, t_n , write

$$W(t_1, t_2, \dots, t_n) = V(a_1(q, t_1), a_2(q, t_2), \dots, a_n(q, t_n)) - \sum_{i=1}^n q_i t_i.$$

This function is submodular in (t_1, \dots, t_n) : the $V(\dots)$ term is a submodular function because it is obtained from the submodular function V by a monotone reparameterization of each coordinate; and the remaining terms are additively separable. Moreover, W is continuously differentiable, with derivatives

$$\frac{\partial W}{\partial t_i} = \left[V_i(a_1(q, t_1), \dots, a_n(q, t_n)) \cdot \frac{\partial a_i}{\partial t} \Big|_{(q, t_i)} \right] - q_i.$$

In particular, when all t_i are equal to the same value t , we get

$$\frac{\partial W}{\partial t_i} \Big|_{(t, t, \dots, t)} = V_i(a(q, t)) \cdot \frac{\partial a_i}{\partial t} \Big|_{(q, t)} - q_i = 0.$$

Hence Lemma A1 applies to W . For each i , apply the lemma, then multiply both sides by q_i , and sum over i . We get

$$\sum_{i=1}^n q_i W(t_i, \dots, t_i) \leq W(t_1, \dots, t_n). \quad (\text{A1})$$

Noting that $W(t_i, \dots, t_i) = V(a(q, t_i)) - t_i$, and $W(t_1, \dots, t_n) = V(a_1(q, t_1), \dots, a_n(q, t_n)) - \sum_i q_i t_i$, we can add $\sum_i q_i t_i$ to both sides of (A1) to obtain

$$\sum_{i=1}^n q_i V(a(q, t_i)) \leq V(a_1(q, t_1), \dots, a_n(q, t_n)). \quad (\text{A2})$$

This holds for all t_1, \dots, t_n .

Finally, suppose a type a sends message q . Let S be the set of coordinates i such that $q_i > 0$. For each $i \in S$, let t_i be the value such that $a_i(q, t_i) = a_i$ (we observed in the text that such a value exists and is unique). Then, if dimension i is verified, the sender will be believed to be type $a(q, t_i)$, and so will get payoff $V(a(q, t_i))$. For any $i \notin S$, dimension i will not be verified; we may take t_i arbitrary. Then, the left side of (A2) equals the expected payoff that the sender gets by sending message q . Meanwhile, the right side of (A2) equals $V(a|_S) \leq V(a)$. Hence, the deviation gives a payoff of at most $V(a)$, the payoff to following the prescribed strategy. \square

For more intuition about the role of submodularity in the argument, consider the special case of $n = 2$. Then reporting a probability vector is equivalent to reporting a single number $q = q_1$. Suppose that a particular type (a_1, a_2) deviates to reporting q that is higher than what he is supposed to report in equilibrium, so that the true type lies below and to the right of the q -curve for the reported q . If dimension 1 is tested, the sender is believed to be (a'_1, a_2) lying on the q -curve with $a'_1 < a_1$, thereby getting lower payoff than by telling the truth; if dimension 2 is tested, he is believed to be (a_1, a'_2) with $a'_2 > a_2$, for a gain in payoff. The magnitude of the gain, $V(a_1, a'_2) - V(a_1, a_2)$, can be written as the integral of V_2 on the line segment from (a_1, a_2) to (a_1, a'_2) . Submodularity means that V_2 is decreasing in the a_1 -coordinate, so this gain is bounded above by a corresponding weighted integral of V_2 on the q -curve from (a'_1, a_2) to (a_1, a'_2) . Similarly, the net gain if dimension 1 is tested (which is negative) can be written as the integral of V_1 on a horizontal line segment, which (again by submodularity) is bounded above by a weighted integral of V_1 on this same segment of the q -curve. When we add these two derivatives, and use the differential equation defining the q -curve, they cancel out exactly; thus, the overall net gain from misreporting is at most zero.

More precisely, for any probability $q = q_1 \in (0, 1)$, define function $y \equiv f_q(x)$ as the solution to the differential equation

$$\frac{dy}{dx} = \frac{1 - q}{q} \frac{V_1(x, y)}{V_2(x, y)}$$

with the initial condition $y(0) = 0$; notice that this defines the same curve as (4). Denote its inverse by $g_q(\cdot)$. Now consider sender of type (a_1, a_2) ; if he were truthful, he would report q such that $a_2 = f_q(a_1)$. Suppose he reports probability $q \in (0, 1)$ such that $a_2 < f_q(a_1)$ instead (the opposite case is analogous). If tested on dimension 1, he will be believed to be type $(a_1, f_q(a_1))$, and if tested on dimension 2, he will be believed to be

type $(g_q(a_2), a_2)$. Thus, in the first case, he would get, in expectation, a gain in utility

$$\begin{aligned} q(V(a_1, f_q(a_1)) - V(a_1, a_2)) &= q \int_{a_2}^{f_q(a_1)} V_2(a_1, y) dy \\ &\leq q \int_{a_2}^{f_q(a_1)} V_2(g_q(y), y) dy, \end{aligned}$$

where we used submodularity of V and the fact that $g_q(y) \leq a_1$ for $y \in [a_2, f_q(a_1)]$. In the second case, the expected net change in utility is

$$\begin{aligned} (1-q)(V(g_q(a_2), a_2) - V(a_1, a_2)) &= -(1-q) \int_{g_q(a_2)}^{a_1} V_1(x, a_2) dx \\ &\leq -(1-q) \int_{g_q(a_2)}^{a_1} V_1(x, f_q(x)) dx, \end{aligned}$$

where we again used submodularity of V and the fact that $f_q(x) > a_2$ for all x . However, using substitution $y = f_q(x)$ and the definition of function $f_q(\cdot)$, we get that these two upper bounds are equal in absolute value and opposite in sign, so the deviation cannot be profitable.

Finally, we formalize the claim in Footnote 5, 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 A2. *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 step 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. \square

B Proofs for Section 4

Proof of Proposition 8. As a preliminary, we should give the appropriate formulation of the incentive constraint (analogous to (1)) for direct mechanisms in this model. If type z reports a and is tested on dimension i for which $a_i \neq z_i$, trusted verification implies that he necessarily receives a payoff of at least $V(z|i)$ (and indeed, this can be done using the belief that places probability 1 on this type). Thus, a verification strategy $p(a)$ can be part of a valid direct mechanism with full learning if and only if

$$V(z) \geq \sum_{i=1}^n p_i(a) w_i(a|z), \quad \text{where} \quad w_i(a|z) = \begin{cases} V(a) & \text{if } z_i = a_i \\ V(z|i) & \text{if } z_i \neq a_i \end{cases} \quad (\text{B1})$$

for all a and z .

Now we proceed to prove existence of the desired direct mechanism. The approach is non-constructive. For each type a , we show that there exist corresponding verification probabilities $p_i(a)$ that satisfy (B1) for all z . By doing this for every a , we form an incentive compatible mechanism.

So fix a type a henceforth. Consider any particular verification probabilities $p = (p_1, \dots, p_n)$ that sum to 1. Notice that the function

$$U_p(a|z) = \sum_{i=1}^n p_i(a) w_i(a|z)$$

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 $l_p = \sup_{z \in B} G_p(z)$. 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 $V(a)$), 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 , $D_p \neq \emptyset$ for all p .

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

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

Let us show that $R_i \neq \emptyset$ for any i . To do that, we show that $1|_i \in R_i$ (here $1|_i$ 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 $\bigcap_{i=1}^n \overline{R_i} = \emptyset$. 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 $i \in \{1, \dots, n\}$ 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 $G_{p^{(i)}}(z^{(i)}) > \frac{n+1}{n+2} l_{p^{(i)}} \geq \frac{n}{n+1} l_p$. By

Lipschitz continuity of $G_p(z^{(i)})$ as a function of p (with coefficient $V(a)$), we have

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

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

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

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

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 $\bigcap_{i=1}^n \overline{R}_i = \emptyset$, so $\bigcap_{i=1}^n R_i = \emptyset$ 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

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

for each p the maximum is finite and well-defined, because each $R_i \neq \emptyset$. 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 $\bigcap_{i=1}^n \overline{R}_i = \emptyset$, which yields a contradiction that proves that $h > 0$. This implies, in particular, that for every point $p \in \Delta$, there is $i \in \{1, \dots, n\}$ 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 $q = p$, $N \left(E_q, \frac{2}{h} \|p - q\| \right) = N \left(E_p, 0 \right) = \overline{E_p} = E_p$, 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 $i \in \{1, \dots, n\}$ 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 $Q_p \neq \emptyset$ 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. \square

Proof of Proposition 9. To show necessity: Suppose such a mechanism exists. Fix a type $a = (a_1, a_2)$. Suppose that when a follows his equilibrium strategy,¹² dimensions 1 and 2 are checked with probability p_1 and p_2 respectively. Now consider any $x_1 < a_1$ and $x_2 < a_2$. If type (a_1, x_2) follows the strategy of type a , with probability p_1 he is believed to be a (due to full learning) and receives payoff $V(a)$; with probability $p_2 = 1 - p_1$ he is believed to be at least $(0, x_2)$ (due to trusted verification). If he instead follows his equilibrium strategy then, by full learning, his payoff is $V(a_1, x_2)$. So incentive compatibility requires

$$p_1 V(a) + (1 - p_1) V(0, x_2) \leq V(a_1, x_2). \quad (\text{B2})$$

Similarly, the incentive of type (x_1, a_2) gives

$$p_1 V(x_1, 0) + (1 - p_1) V(a) \leq V(x_1, a_2). \quad (\text{B3})$$

The first equation implies $p_1 \leq \frac{V(a_1, x_2) - V(0, x_2)}{V(a) - V(0, x_2)}$ (note 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 likewise implies $p_1 \geq \frac{V(a) - V(x_1, a_2)}{V(a) - V(x_1, 0)}$. 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 $x < a$. We will construct a direct mechanism that achieves full learning: as argued in the proof of Proposition 8, it suffices to find verification probabilities satisfying (B1).

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 . Note that deviation by types x with $x_1 \neq a_1$ and $x_2 \neq a_2$ is automatically deterred, since the deviation will always be detected and the sender will be believed to be either $(x_1, 0)$ or $(0, x_2)$, both of which are worse than truth-telling. Moreover, types (x_1, a_2) with $x_1 > a_1$ cannot benefit

¹²In fact, one can formulate a revelation principle (analogous to Lemma 0) under the restriction of trusted verification. For brevity, we omit a formal statement.

from deviating to a since the truth-telling payoff is $V(x_1, a_2) \geq V(a)$; likewise for types (a_1, x_2) with $x_2 > a_2$. So we need only worry about deviations by types (x_1, a_2) with $x_1 < a_1$ or (a_1, x_2) with $x_2 < a_2$.

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 inequality in the proposition statement 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 brings us back to conditions (B2)–(B3), which coincide with the incentive constraints (B1) for types (a_1, x_2) and (x_1, a_2) . So deviations to a by these types are deterred. \square

Proof of Proposition 10. We just need to check that if condition (1) is satisfied for the function V , then it is also satisfied for $V' = U \circ V$. For any a, \hat{a} , put $\lambda = \sum_{i: \hat{a}_i = a_i} p_i(\hat{a})$; thus $\lambda \in [0, 1]$. Condition (1) says that $V(a) \geq \lambda V(\hat{a})$. Then,

$$U(V(a)) \geq U(\lambda V(\hat{a})) \geq \lambda U(V(\hat{a}))$$

where the first inequality is because U is increasing and the second is because U is concave (and must map 0 to 0 in order for $U \circ V$ to be an allowable payoff function). Thus, (1) holds for $U \circ V$. \square

Proof of Proposition 11. Let $\tilde{V}(a) = V(a; a)$. By Proposition 2, there is a valid direct mechanism that achieves full learning for payoff function $\tilde{V}(a)$; denote such a mechanism by \mathcal{M} . Let us show that this same mechanism would remain incentive compatible if the payoffs of type x if he is believed to be type a were given by $V(a; x)$.

Suppose not, so that some type x prefers to deviate and report type $a \neq x$. This

immediately implies $V(a; x) > V(x; x)$, for otherwise this deviation would not have any upside. Since we assumed that $V(a; x) \leq \max\{V(a; a), V(x; x)\}$, it must be that $V(a; x) \leq V(a; a)$. Given that \mathcal{M} is a valid mechanism for the payoff function $\tilde{V}(a)$, it must be that

$$\tilde{V}(x) \geq \left(\sum_{i: a_i=x_i} p_i(a) \right) \tilde{V}(a) + \left(\sum_{i: a_i \neq x_i} p_i(a) \right) \tilde{V}(0).$$

(the last term $\tilde{V}(0)$ is zero, but we spelled it out). Since $V(a; x) \leq V(a; a) = \tilde{V}(a)$ and $V(0; x) = \tilde{V}(0)$, this implies

$$V(x; x) \geq \left(\sum_{i: a_i=x_i} p_i(a) \right) V(a; x) + \left(\sum_{i: a_i \neq x_i} p_i(a) \right) V(0; x);$$

in other words, the deviation to reporting a is not profitable. This proves that \mathcal{M} is a valid mechanism. \square

The following auxiliary result will be used in the proof of Proposition 12.

Lemma B1. *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,*

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

Proof. Note that for each i , we have

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

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

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

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. \square

Proof of Proposition 12. We start by formalizing the problem: given a mechanism and a noise parameter χ , define the posterior measure $\kappa(a)$ on any measurable $A' \subset A_K$

by

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

where $\rho_i(s|a_i)$ is lognormal with $\log s \sim \mathcal{N}\left(\log a_i, \frac{1}{\chi}\right)$. We will show that, when the mechanism is chosen to depend on χ as in the proposition statement, then $\kappa^\chi(a)$ converges to a for all $a \in A_K$ as $\chi \rightarrow \infty$, and the mechanism is valid.

Let us first present the proof assuming $\gamma = 1$. We will consider $\gamma \in (0, 1)$ in the end of the proof.

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 $(\log V(\tilde{a}) | m) \sim \mathcal{N}\left(\sum_{i=1}^n \frac{\tau_i}{\tau} (\nu_i - \log m_i), \frac{1}{\tau}\right)$. (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 $(\log V(\tilde{a}) | m)$

at point z is equal to:

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

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 $\log s = (\log V(a) | m) + \log m_i + \log \eta$. Thus, $\log s - \log m_i$ is a signal of the unknown value $(\log V(\tilde{a}) | m)$ with precision χ . Thus, we have that the posterior of $V(\tilde{a})$ is lognormal, with

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

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 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), $(\log V(\tilde{a}) | m, i, s)$ 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 $\chi \rightarrow \infty$; 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

$$\begin{aligned}
C &= \exp \left(\frac{1}{\tau + \chi} \sum_{j=1}^n \tau_j \nu_j + \frac{1}{2} \frac{\tau + 2\chi}{(\tau + \chi)^2} \right), \\
H(m) &= \left(\prod_{j=1}^n m_j^{-\frac{\tau_j}{\tau + \chi}} \right) \left(\sum_{j=1}^n \left(m_j \left(1 + \frac{\tau}{\chi} \right) - \frac{\tau_j}{\chi} \right) \left(\frac{a_j}{m_j} \right)^{\frac{\chi}{\tau + \chi}} \right).
\end{aligned}$$

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 B1 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. This completes the proof for $\gamma = 1$.

Now suppose that $\gamma \in (0, 1)$. The mechanism is exactly the same as before. We just need to check that it remains incentive-compatible. The proof follows the same steps as the previous case. The following argument highlights one additional step needed to complete the proof.

As in the previous case, we can 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, $H_\gamma(q) = H(q)^\gamma$, 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 earlier in the proof, implies

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

thus establishing incentive compatibility. This completes the proof. \square

C Proofs for Section 5

Proof of Proposition 13. We consider the following cases. If $\xi = 0$, then $\eta = 0$ as well. The mechanism described therefore achieves full learning, and the fact that this mechanism is valid follows immediately from Proposition 3.

If $\xi = 1$, then for any a either $a_2 \leq \xi a_1$ or $a_1 < \xi a_2$, so the set of types that report truthfully is empty. Consider any type (a_1, a_2) with $a_2 \leq a_1$. If he reports $(x, *)$ with $x \neq a_1$, then he will be caught for sure; the same is true if he reports $(*, y)$ with $y \neq a_2$. The remaining deviation to consider is deviating to reporting his true a_2 , that is, sending message $(*, a_2)$. In this case, he will get payoff $\mathbb{E}_{0 \leq x \leq \xi a_2} V(x, a_2)$. However, by the symmetry of V and Φ , we have $\mathbb{E}_{0 \leq x \leq \xi a_2} V(x, a_2) \leq \mathbb{E}_{0 \leq x \leq \xi a_2} V(a_2, x) \leq \mathbb{E}_{0 \leq x \leq \xi a_2} V(a_1, x) \leq \mathbb{E}_{0 \leq x \leq \xi a_1} V(a_1, x)$, where the last two inequalities are due to monotonicity of V and the monotone expectation property. Thus, this deviation is not profitable. The proof that type (a_1, a_2) with $a_1 < a_2$ does not have a profitable deviation is similar.

Consider the remaining case $\xi \in (0, 1)$. Notice that for each (a_1, a_2) that is supposed to report truthfully (satisfies $a_2 > \xi a_1$ and $a_1 \geq \xi a_2$), we have $1 - \frac{V(a_1, \eta a_1)}{V(a_1, a_2)} \geq 0$ and $1 - \frac{V(\eta a_2, a_2)}{V(a_1, a_2)} \geq 0$ by the observation that $\xi \geq \eta$. Furthermore, $1 - \frac{V(a_1, \eta a_1)}{V(a_1, a_2)} + 1 - \frac{V(\eta a_2, a_2)}{V(a_1, a_2)} = 1 - \frac{V(a_1, \eta a_1) + V(\eta a_2, a_2) - V(a_1, a_2)}{V(a_1, a_2)} \leq 1$, because the last term is positive by definition of η in (6). Therefore, probabilities $p_1(a_1, a_2)$ and $p_2(a_1, a_2)$ with the required properties exist.

Take type (a_1, a_2) with $a_2 \leq \xi a_1$. The proof that he would not be better off by reporting $(x, *)$ or $(*, y)$ for any x and y is identical to the previous case. Suppose that he deviates and reports (a_1, y) ; he is only able to do so if $y > \xi a_1$ and $\xi y \leq a_1$; the first inequality implies $y > a_2$. His payoff from deviating is at most $(1 - p_2(a_1, y)) V(a_1, y) \leq V(a_1, \eta a_1)$. At the same time, his equilibrium payoff equals $\mathbb{E}_{0 \leq z \leq \xi a_1} V(a_1, z)$, and by definition of ξ in (7) there exists ξ' arbitrarily close to ξ such that $\mathbb{E}_{0 \leq z \leq \xi' a_1} V(a_1, z) \geq V(a_1, \eta a_1)$. Since the left-hand side is continuous in ξ' , we have $\mathbb{E}_{0 \leq z \leq \xi a_1} V(a_1, z) \geq V(a_1, \eta a_1)$, which implies that this deviation is not profitable. Now suppose that he deviates and reports type (x, a_2) with $a_2 > \xi x$ and $\xi a_2 \leq x$; now the first inequality implies $x < a_1$. The payoff from deviating is at most $(1 - p_1(x, a_2)) V(x, a_2) \leq V(\eta a_2, a_2)$. However, we have $V(\eta a_2, a_2) = V(a_2, \eta a_2) \leq \mathbb{E}_{0 \leq z \leq \xi a_2} V(a_2, z) \leq \mathbb{E}_{0 \leq z \leq \xi a_2} V(a_1, z) \leq \mathbb{E}_{0 \leq z \leq \xi a_1} V(a_1, z)$, which is his equilibrium payoff. Thus, this deviation cannot be profitable either.

If we take type (a_1, a_2) with $a_1 < \xi a_2$, then we similarly get that this type does not have a profitable deviation. It remains to consider deviations by type (a_1, a_2) that reports truthfully in equilibrium (so $a_2 > \xi a_1$ and $a_1 \geq \xi a_2$). If he deviates by sending message $(a_1, *)$, his payoff will equal $\mathbb{E}_{0 \leq z \leq \xi a_1} V(a_1, z)$, which is less than $V(a_1, a_2)$ since $a_2 > \xi a_1$. This cannot be profitable; we get a similar contradiction if he deviates by sending $(*, a_2)$.

Notice that if he sends $(x, *)$ for $x \neq a_1$ or $(*, y)$ for $y \neq a_2$, he will be caught for sure, and this is not profitable. Now suppose that he deviates by mimicking some type that also reveals itself in equilibrium. To avoid getting caught with probability 1, he must send either message (a_1, y) for some y or (x, a_2) for some x . In the former case, his payoff from this deviation is the same as that of type $(a_1, 0)$, but that type has a lower equilibrium payoff. Consequently, if the deviation is profitable for type (a_1, a_2) , it must be profitable for $(a_1, 0)$, but we already proved that he does not have profitable deviations. We would get a similar contradiction if deviating to (x, a_2) was profitable. We have thus proved that no type has a profitable deviation, so there exists a valid mechanism with the required properties. \square

Proof of Proposition 14. Let us pick $q, Q \in (0, \infty)$ such that the mass of sender types in $[q, Q] \times [q, Q]$ is at least $1 - \delta$. Now consider the function $R(a_1) = \mathbb{E}_{0 \leq a_2 \leq q} \bar{V}(a_1, a_2) - \bar{V}(a_1, 0)$. Since \bar{V} is continuous and Φ is atomless with full support, $R(a_1)$ is a continuous function on $[q, Q]$. Furthermore, since $\bar{V}(a_1, a_2)$ is locally increasing in a_2 , it is always positive. Let $\nu > 0$ be its minimum; since \bar{V} is uniformly continuous on $[q, Q]$, there is $\varepsilon > 0$ such that for all $a_1 \in [q, Q]$ and $a_2 \leq \varepsilon$, $\bar{V}(a_1, a_2) - \bar{V}(a_1, 0) \leq \frac{\nu}{2}$. Thus, ε has the property that $\mathbb{E}_{0 \leq a_2 \leq q} \bar{V}(a_1, a_2) \geq \bar{V}(a_1, \varepsilon) + \frac{\nu}{2}$ for all $a_1 \in [q, Q]$.

Since \bar{V} is submodular and strictly increasing in a_2 at $a_2 = 0$, we have $\bar{V}(a_1, \varepsilon) + \bar{V}(\varepsilon, a_2) > \bar{V}(a_1, 0) + \bar{V}(0, a_2) \geq \bar{V}(a_1, a_2)$. Let $S(a_1, a_2) = \bar{V}(a_1, \varepsilon) + \bar{V}(\varepsilon, a_2) - \bar{V}(a_1, a_2)$; this function is continuous in (a_1, a_2) and positive, and thus its minimum on compact $[q, Q] \times [q, Q]$ must be positive; denote it by $\chi > 0$. Lastly, let $L = V(Q, Q)$, and take $T = \min\{\frac{\nu}{2L}, \frac{\chi}{2L}\}$.

We claim that the following is a valid mechanism for function $\tilde{V}_t(a_1, a_2)$ for $t < T$. If either (a) $a_1 \notin [q, Q]$ and $a_2 \leq a_1$, or (b) $a_1 \in [q, Q]$ and $a_2 < q$, then the sender reports $(a_1, *)$, and the first dimension is verified with probability 1. If either (a) $a_2 \notin [q, Q]$ and $a_1 < a_2$, or (b) $a_2 \in [q, Q]$ and $a_1 < q$, then the sender reports $(*, a_2)$, and the second dimension is verified with probability 1. Otherwise, the sender reports (a_1, a_2) , and the verification probabilities satisfy $p_1(a_1, a_2) \geq 1 - \frac{\tilde{V}_t(\varepsilon, a_2)}{\tilde{V}_t(a_1, a_2)}$ and $p_2(a_1, a_2) \geq 1 - \frac{\tilde{V}_t(a_1, \varepsilon)}{\tilde{V}_t(a_1, a_2)}$. In this mechanism, for $a_1 \in [q, Q]$, we have $\mathbb{E}_{0 \leq a_2 \leq q} \tilde{V}_t(a_1, a_2) \geq \tilde{V}_t(a_1, \varepsilon)$; this follows from

$$\begin{aligned} & \mathbb{E}_{0 \leq a_2 \leq q} \tilde{V}_t(a_1, a_2) - \tilde{V}_t(a_1, \varepsilon) \\ &= (1-t) (\mathbb{E}_{0 \leq a_2 \leq q} \bar{V}(a_1, a_2) - \bar{V}(a_1, \varepsilon)) + t (\mathbb{E}_{0 \leq a_2 \leq q} V(a_1, a_2) - V(a_1, \varepsilon)) \\ &\geq (1-t) \frac{\nu}{2} - tL \geq \frac{\nu}{2} - TL \geq \frac{\nu}{2} - \frac{\nu}{2L}L = 0. \end{aligned}$$

We also have $\tilde{V}_t(a_1, \varepsilon) + \tilde{V}_t(\varepsilon, a_2) \geq \tilde{V}_t(a_1, a_2)$ for $a_1, a_2 \in [q, Q]$; this follows from

$$\begin{aligned} & \tilde{V}_t(a_1, \varepsilon) + \tilde{V}_t(\varepsilon, a_2) - \tilde{V}_t(a_1, a_2) \\ &= (1-t) (\bar{V}(a_1, \varepsilon) + \bar{V}(\varepsilon, a_2) - \bar{V}(a_1, a_2)) + t(V(a_1, \varepsilon) + V(\varepsilon, a_2) - V(a_1, a_2)) \\ &\geq (1-t)\chi - 2tL \geq \chi - 2TL \geq \chi - 2\frac{\chi}{2L}L = 0. \end{aligned}$$

This implies that nonnegative $p_1(a_1, a_2) \geq 1 - \frac{\tilde{V}_t(\varepsilon, a_2)}{\tilde{V}_t(a_1, a_2)}$ and $p_2(a_1, a_2) \geq 1 - \frac{\tilde{V}_t(a_1, \varepsilon)}{\tilde{V}_t(a_1, a_2)}$ exist. The proof that the mechanism is valid, i.e. that there is no profitable deviation, is similar to that in Proposition 13 given the inequalities above. Lastly, the mechanism achieves full learning on $[q, Q] \times [q, Q]$. \square

To prove Propositions 15 and 16, we need some notation and a few lemmas. In what follows, $t \in (0, 1]$. Suppose that $V(a_1, a_2) = Z_t(a_1, a_2)$, Φ is uniform on $\Omega = [0, L] \times [0, L]$, and \mathcal{M} is a connected semi-direct mechanism. Let $\Omega_1(a_1)$ be the set of types reporting $(a_1, *)$ and $\Omega_2(a_2)$ be the set of types reporting $(*, a_2)$ according to \mathcal{M} . Let

$$\Omega_0 = \Omega \setminus \left(\left(\bigcup_{a_1 \in [0, L]} \Omega_1(a_1) \right) \cup \left(\bigcup_{a_2 \in [0, L]} \Omega_2(a_2) \right) \right)$$

be the set of types who report their type truthfully. Define $\omega_1(a_1) = \sup(\{a_2 : (a_1, a_2) \in \Omega_1(a_1)\} \cup \{0\})$ and $\omega_2(a_2) = \sup(\{a_1 : (a_1, a_2) \in \Omega_2(a_2)\} \cup \{0\})$; these values are well-defined and lie in $[0, L]$.

Lemma C1. *For any $a_1, a_2, x, y \in [0, L]$, $U(a_1, y) + U(x, a_2) \geq U(a_1, a_2)$.*

Proof. Let p_1 and p_2 be the probabilities of verification of the two dimensions for type (a_1, a_2) . Then incentive compatibility implies $U(a_1, y) \geq p_1 U(a_1, a_2)$ and $U(x, a_2) \geq p_2 U(a_1, a_2)$. We have

$$U(a_1, y) + U(x, a_2) \geq (p_1 + p_2)U(a_1, a_2) = U(a_1, a_2).$$

\square

Lemma C2. *If $t < 1$, it is impossible that for some $a_1, a_2 \in [0, L]$, $(a_1, 0) \in \Omega_2(0)$ and $(0, a_2) \in \Omega_1(0)$.*

Proof. Suppose that this is possible for some a_1, a_2 . Then there is some type $(x, 0) \in \Omega_2(0)$ such that $U(x, 0) \leq V(x, 0)$; this follows from the fact that the averages of U and V on

the set of types sending the same message, in this case $\Omega_2(0)$, must be the same. Similarly, there is some type $(0, y) \in \Omega_1(0)$ such that $U(0, y) \leq V(0, y)$.

Consider type (x, y) . Suppose $(x, y) \in \Omega_1(x)$, this implies $y > 0$ (because $(x, 0) \in \Omega_2(0)$), and then $U(x, y) > V(x, 0) \geq U(x, 0)$ (the first inequality follows since $V(x, z) > V(x, 0)$ for all (x, z) , because for $t < 1$, $V(x, \cdot)$ is strictly increasing). Then type $(x, 0)$ has a profitable deviation, which is to send message $(x, *)$, a contradiction. We get a similar contradiction if $(x, y) \in \Omega_2(y)$.

Consider the remaining case, $(x, y) \in \Omega_0$. Notice that we must have $x, y > 0$ in this case. Now, we have

$$\begin{aligned} U(x, 0) + U(0, y) &\leq V(x, 0) + V(0, y) \\ &= (1 - t)(x + y) \\ &< (1 - t)(x + y) + t \min\{x, y\} \\ &= V(x, y) = U(x, y); \end{aligned}$$

the latter follows from $(x, y) \in \Omega_0$. However, this inequality contradicts Lemma C1. This contradiction completes the proof. \square

Lemma C3. *Suppose $(a_1, a_2) \in \Omega_1(a_1)$, and either $a_1 > 0$ or $t < 1$ (or both). Then for any $y \in [0, a_2]$, $(a_1, y) \in \Omega_1(a_1)$. Similarly, if $(a_1, a_2) \in \Omega_2(a_2)$ and either $a_2 > 0$ or $t < 1$, then for any $x \in [0, a_1]$, $(x, a_2) \in \Omega_2(a_2)$.*

Proof. Take the first part of the statement (the proof of the second part is analogous). It suffices to prove that $(a_1, 0) \in \Omega_1(a_1)$; then the result will follow from connectedness. The statement is trivial if $a_2 = 0$, so suppose $a_2 > 0$. Now suppose, to obtain a contradiction, that $(a_1, 0) \notin \Omega_1(a_1)$. This implies that either $(a_1, 0) \in \Omega_0$ or $(a_1, 0) \in \Omega_2(0)$. In the former case, $U(a_1, 0) = V(a_1, 0) < U(a_1, a_2)$, because $V(a_1, 0) < V(a_1, z)$ for all $(a_1, z) \in \Omega_1(a_1)$, and $U(a_1, a_2)$ is the mean of such values. (This argument makes use of the assumption $a_1 > 0$ in case $t = 1$, but does not need this assumption if $t < 1$.) This implies that type $(a_1, 0)$ has a profitable deviation, a contradiction. So suppose $(a_1, 0) \in \Omega_2(0)$, and consider the following cases separately.

Case 1: $t = 1$, so $V(a_1, a_2) = \min\{a_1, a_2\}$. In this case, $U(a_1, 0) = 0$ (because this is the value of V at all types in $\Omega_2(0)$). As before, this shows that type $(a_1, 0)$ has a profitable deviation.

Case 2: $t < 1$ and $a_1 = 0$. In this case, we have $(0, a_2) = (a_1, a_2) \in \Omega_1(a_1) = \Omega_1(0)$ and $(a_1, 0) \in \Omega_2(0)$; taken together, these contradict Lemma C2.

Case 3: $t < 1$ and $a_1 > 0$. We have $(a_1, 0) \in \Omega_2(0)$. Moreover, writing $b \equiv \omega_2(0)$, the entire segment from $(0, 0)$ to $(b, 0)$, perhaps excluding point $(b, 0)$, is in $\Omega_2(0)$; this follows from Case 2, which we have already proven (applied as in the second part of the Lemma). Furthermore, we must have $b \geq 2a_1$, because otherwise $U(a_1, 0) < V(a_1, 0)$, and $(a_1, 0)$ would deviate to reporting $(a_1, *)$. Take some $\varepsilon \in (0, \min\{\frac{b-a_1}{2}, \frac{a_2}{2}\})$; we have $(b - \varepsilon, 0) \in \Omega_2(0)$ and thus $U(b - \varepsilon, 0) = \frac{1-t}{2}b$ (we cannot guarantee that $(b, 0) \in \Omega_2(0)$, hence we use $(b - \varepsilon, 0)$ instead).

Now consider type $(b - \varepsilon, a_2)$. We have $U(b - \varepsilon, 0) = \frac{1-t}{2}b < (1-t)(b - \varepsilon) = V(b - \varepsilon, 0)$, so if $(b - \varepsilon, a_2) \in \Omega_1(b - \varepsilon)$, then $(b - \varepsilon, 0)$ would deviate to reporting $(b - \varepsilon, *)$. Suppose that $(b - \varepsilon, a_2) \in \Omega_2(a_2)$, then $(0, a_2) \notin \Omega_2(a_2)$ (otherwise connectedness would imply that $(a_1, a_2) \in \Omega_2(a_2)$, but this is not the case), and we already proved that $(0, a_2) \notin \Omega_1(0)$. Thus, $(0, a_2) \in \Omega_0$, but then $(0, a_2)$ would have a profitable deviation to reporting $(*, a_2)$, a contradiction. This establishes that $(b - \varepsilon, a_2) \in \Omega_0$.

Now, by Lemma C1, $U(b - \varepsilon, a_2) \leq U(b - \varepsilon, 0) + U(a_1, a_2)$. Moreover, we know that $U(a_1, a_2) \leq U(a_1, 0) = U(b - \varepsilon, 0)$ (the equality is because types $(a_1, 0)$ and $(b - \varepsilon, 0)$ both send message $(*, 0)$, and the inequality is because otherwise $(a_1, 0)$ would deviate to $(a_1, *)$). Combining gives $U(b - \varepsilon, a_2) \leq 2U(b - \varepsilon, 0) = (1-t)b$. On the other hand, because we have established $(b - \varepsilon, a_2) \in \Omega_0$, we get $U(b - \varepsilon, a_2) = V(b - \varepsilon, a_2) \geq (1-t)(b - \varepsilon + a_2)$. This is a contradiction. \square

Lemma C4. *The receiver's objective $\mathcal{W}(\mathcal{M})$ satisfies*

$$\mathcal{W}(\mathcal{M}) = \frac{1}{L^2} \left(\int_0^L \beta(a_1, \omega_1(a_1), t) da_1 + \int_0^L \beta(a_2, \omega_2(a_2), t) da_2 \right),$$

where

$$\beta(x, y, t) = \begin{cases} \frac{1}{4}y^2 & \text{if } y \leq x; \\ \frac{1}{4} \left(y - t \frac{(y-x)^2}{y} \right)^2 & \text{if } x < y \leq \left(\sqrt{\frac{1}{1-t}} + 1 \right) x; \\ \frac{1}{4(1-t)} \left(y - t \frac{y^2 - x^2}{y} \right)^2 & \text{if } y > \left(\sqrt{\frac{1}{1-t}} + 1 \right) x; \end{cases}$$

(where if $t = 1$, $\sqrt{\frac{1}{1-t}} = \infty$).

Proof. We have (the factors $\frac{1}{L^2}$ before the integrals reflect the density of the distribution):

$$\begin{aligned}\mathcal{W}(\mathcal{M}) &= \int_{\Omega_0} |U(a_1, a_2) - V(a_1, a_2)| d\Phi \\ &\quad + \frac{1}{L^2} \int_0^1 \int_0^{\omega_1(a_1)} |U(a_1, a_2) - V(a_1, a_2)| da_2 da_1 \\ &\quad + \frac{1}{L^2} \int_0^1 \int_0^{\omega_2(a_2)} |U(a_1, a_2) - V(a_1, a_2)| da_1 da_2.\end{aligned}$$

Notice that the first term is zero (because of truth-telling, $U(a_1, a_2) = V(a_1, a_2)$ for $(a_1, a_2) \in \Omega_0$). Consider the inner integral of the second term. For any fixed $a_1 > 0$ (i.e. for almost all a_1), the set of types integrated in this term are exactly the set of types $\Omega_1(a_1)$ that send message $(a_1, *)$ in equilibrium (perhaps with the exception of type $(a_1, \omega_1(a_1))$). To see why this is the case, notice that $\omega_1(a_1) = 0$ if and only if $\Omega_1(a_1)$ is empty or a singleton $\{(a_1, 0)\}$, so the statement is true. If $\omega_1(a_1) > 0$, then from the definition of $\omega_1(a_1)$ it follows that $\Omega_1(a_1)$ is nonempty, in which case $(a_1, 0) \in \Omega_1(a_1)$ by Lemma C3, and thus $\Omega_1(a_1)$ spans the types from $(a_1, 0)$ to $(a_1, \omega_1(a_1))$.

In the following, assume that $\omega_1(a_1) > 0$, for otherwise the contribution of this inner integral equals zero as it should. We have that $U(a_1, a_2)$ is the same for (almost) all types being integrated and that $\int_0^{\omega_1(a_1)} U(a_1, a_2) da_2 = \int_0^{\omega_1(a_1)} V(a_1, a_2) da_2$. From this, it is straightforward to get

$$U(a_1, a_2) = \begin{cases} a_1 + \frac{\omega_1(a_1)}{2} - ta_1 & \text{if } \omega_1(a_1) \leq a_1; \\ a_1 + \frac{\omega_1(a_1)}{2} - t \frac{(a_1)^2 + (\omega_1(a_1))^2}{2\omega_1(a_1)} & \text{if } \omega_1(a_1) > a_1. \end{cases}$$

Notice that due to $V(a_1, a_2)$ being increasing in a_2 and continuous, there is a value of $z(a_1) \in (0, \omega_1(a_1))$ such that $V(a_1, z(a_1)) = U(a_1, z(a_1))$ (and this value of z is unique unless $t = 1$). By solving the equation, we find

$$z(a_1) = \begin{cases} \frac{\omega_1(a_1)}{2} & \text{if } \omega_1(a_1) \leq a_1; \\ \frac{1}{2} \left(\omega_1(a_1) - t \frac{(\omega_1(a_1) - a_1)^2}{\omega_1(a_1)} \right) & \text{if } a_1 < \omega_1(a_1) \leq \left(\sqrt{\frac{1}{1-t}} + 1 \right) a_1; \\ \frac{1}{2} \left(\omega_1(a_1) - \frac{t}{1-t} \frac{(a_1)^2}{\omega_1(a_1)} \right) & \text{if } \omega_1(a_1) > \left(\sqrt{\frac{1}{1-t}} + 1 \right) a_1. \end{cases}$$

Now, direct computation of the integral shows that $\int_0^{\omega_1(a_1)} |U(a_1, a_2) - V(a_1, a_2)| da_2 =$

$\beta(a_1, \omega(a_1), t)$ in each of the cases; notice that for this calculation, it is convenient to use

$$\begin{aligned}
& \int_0^{\omega(a_1)} |U(a_1, a_2) - V(a_1, a_2)| da_2 \\
&= \int_0^{z(a_1)} (U(a_1, a_2) - V(a_1, a_2)) da_2 + \int_{z(a_1)}^{\omega(a_1)} (V(a_1, a_2) - U(a_1, a_2)) da_2 \\
&= 2 \int_0^{z(a_1)} (U(a_1, a_2) - V(a_1, a_2)) + \int_0^{\omega(a_1)} (V(a_1, a_2) - U(a_1, a_2)) da_2 \\
&= 2 \int_0^{z(a_1)} (U(a_1, a_2) - V(a_1, a_2)),
\end{aligned}$$

which follows from $\int_0^{\omega(a_1)} V(a_1, a_2) da_2 = \int_0^{\omega(a_1)} U(a_1, a_2) da_2$.

We have thus shown that the contribution of $\Omega_1(a_1)$ is indeed given by $\beta(a_1, \omega(a_1), t)$. To account for other sets $\Omega_1(\cdot)$, we need to integrate this over a_1 , and to account for sets $\Omega_2(\cdot)$, we need to add a similar term. This completes the proof. \square

Proof of Proposition 15. For completeness, we start by verifying the statement (made in the text) that for $V = Z_t$, we have $\eta = \frac{t}{2}$, and moreover with uniform Φ , we have $\xi = 2\eta = t$. For any a_1 and a_2 , $V(a_1, \frac{t}{2}a_1) + V(\frac{t}{2}a_2, a_2) = (1 - \frac{t}{2})a_1 + (1 - \frac{t}{2})a_2 = a_1 + a_2 - t \frac{a_1 + a_2}{2} \geq a_1 + a_2 - t \cdot \max\{a_1, a_2\} = V(a_1, a_2)$, whereas for any $\eta' < \frac{t}{2}$, we can take $a_1 = a_2 = x$ and then $V(x, \eta'x) + V(\eta'x, x) = 2(1 + \eta' - t)x < (2 - t)x = V(x, x)$. This confirms $\eta = \frac{t}{2}$. Moreover, $V(a_1, a_2)$ is linear in a_2 for $a_2 \leq a_1$, so for $\xi' \leq 1$ we have $\mathbb{E}_{0 \leq a_2 \leq \xi' a_1} V(a_1, a_2) = V(a_1, \frac{\xi'}{2}a_1)$ which leads to $\xi = 2\eta$. Proposition 13 now implies that the proposed mechanism \mathcal{M}_t is valid.

For mechanism \mathcal{M}_t , $\omega_1(a_1) = ta_1$ and $\omega_2(a_2) = ta_2$, so Lemma C4 implies that

$$\mathcal{W}(\mathcal{M}_t) = \frac{1}{L^2} \left(\int_0^L \frac{(ta_1)^2}{4} da_1 + \int_0^L \frac{(ta_2)^2}{4} da_2 \right) = \frac{t^2 L}{6}.$$

Now take any connected semi-direct valid mechanism \mathcal{M}' , and let us show that $\mathcal{W}(\mathcal{M}') \geq \mathcal{W}(\mathcal{M}_t) = \frac{t^2 L}{6}$. In the light of Lemma C4, it suffices to prove that for any $x \in (0, L)$, $\beta(x, \omega'_1(x), t) + \beta(x, \omega'_2(x), t) \geq \frac{(tx)^2}{2}$, where $\omega'_1(x)$ and $\omega'_2(x)$ are defined for the mechanism \mathcal{M}' .

It is easy to show that $\beta(x, y, t)$ is nondecreasing in y , and in particular if $y > x$, then $\beta(x, y, t) \geq \beta(x, x, t) = \frac{x^2}{4}$. Now suppose first that $\omega'_1(x) \geq x$. We then have $\beta(x, \omega'_1(x), t) \geq \frac{x^2}{4}$, and also that $\beta(x, \omega'_2(x), t) \geq 0$. Their sum is therefore at least $\frac{x^2}{4}$, which exceeds $\frac{(tx)^2}{2}$ for $t < \sqrt{\frac{1}{2}}$. If $\omega'_2(x) \geq x$, the needed inequality is obtained in a

similar way.

So consider the case $\omega'_1(x) < x$ and $\omega'_2(x) < x$. Let us show that it cannot be that $(0, x) \in \Omega'_1(0)$. Suppose, to obtain a contradiction, that this is the case, then by Lemma C2, $(x, 0) \notin \Omega'_2(0)$. This, together with $\omega'_1(x) < x$, implies that $U'(x, 0) \leq \frac{1}{2}(V(x, x) + V(x, 0)) = \frac{1}{2}(x + x - tx + x - tx) = (\frac{3}{2} - t)x$. Let $y = \omega'_1(0)$. Lemma C3 implies that $\Omega'_1(0)$ consists of the segment from $(0, 0)$ to $(0, y)$ (possibly excluding the endpoint $(0, y)$, if $y > x$). If $y = x$, then we have $U'(0, x) = \frac{1-t}{2}x$ and, clearly, $(x, x) \in \Omega'_0$, and then $U'(x, x) - U'(x, 0) - U'(0, x) \geq (2-t)x - (\frac{3}{2} - t)x - \frac{1-t}{2}x = \frac{tx}{2} > 0$, contradicting Lemma C1. If $y > x$, take $\varepsilon \in (0, \frac{y-x}{2})$; then $U(0, y - \varepsilon) = \frac{1-t}{2}y$. We similarly have $(x, y - \varepsilon) \in \Omega'_0$, and $U'(x, y - \varepsilon) - U'(x, 0) - U'(0, y - \varepsilon) = x + (1-t)(y - \varepsilon) - (\frac{3}{2} - t)x - \frac{1-t}{2}y \geq \frac{(1-t)(y-x-2\varepsilon)+tx}{2} > 0$, again contradicting Lemma C1. This shows that $(0, x) \notin \Omega'_1(0)$; we can similarly show that $(x, 0) \notin \Omega'_2(0)$.

These results imply that both in case $(x, 0) \in \Omega'_0$ and in case $(x, 0) \in \Omega'_1(x)$, the payoff of type $(x, 0)$ under \mathcal{M}' is given by $U'(x, 0) = (1-t)x + \frac{\omega'_1(x)}{2}$; similarly, we have $U'(0, x) = (1-t)x + \frac{\omega'_2(x)}{2}$. Since we must have $(x, x) \in \Omega'_0$, we have $U'(x, x) = (2-t)x$. Now by Lemma C1, $U'(x, x) \leq U'(x, 0) + U'(0, x)$, which simplifies to $\omega'_1(x) + \omega'_2(x) \geq 2tx$. We must thus have $(\omega'_1(x))^2 + (\omega'_2(x))^2 \geq 2(tx)^2$, with equality only being achieved if $\omega'_1(x) = \omega'_2(x) = tx$. But given that $\omega'_1(x) < x$ and $\omega'_2(x) < x$,

$$\beta(x, \omega'_1(x), t) + \beta(x, \omega'_2(x), t) = \frac{1}{4}(\omega'_1(x))^2 + \frac{1}{4}(\omega'_2(x))^2 \geq \frac{(tx)^2}{2},$$

and the inequality is strict unless $\omega'_1(x) = \omega'_2(x) = tx$.

We have thus shown that $\mathcal{W}(\mathcal{M}') \geq \mathcal{W}(\mathcal{M}_t) = \frac{t^2 L}{6}$, and the equality is strict unless \mathcal{M}' coincides with \mathcal{M}_t for almost all types, which proves that for $t < \sqrt{\frac{1}{2}}$, \mathcal{M}_t is the optimal mechanism within the class considered, and is essentially unique in that. \square

To prove Proposition 16, we first need to prove several further Lemmas (Lemmas C5–C16). In all of them, we assume that $V(a_1, a_2) = Z_t(a_1, a_2)$ for $t = 1$ without mentioning this explicitly.

Let us call a semi-direct connected valid mechanism an *admissible* mechanism for brevity. In Lemmas C5, C6, C7, C8, and C9, we document further properties that any admissible mechanism \mathcal{M} must satisfy.

Lemma C5. *For any $x \in [0, L]$, either $\omega_1(x) \geq x$ or $\omega_2(x) \geq x$.*

Proof. If $x = 0$, this is trivial. Take $x > 0$ and suppose, to obtain a contradiction, that $\omega_1(x) < x$ and $\omega_2(x) < x$. Then type $(x, x) \in \Omega_0$, and thus $U(x, x) = x$. Consider

type $(x, 0)$; if $(x, 0) \in \Omega_0$ or $(x, 0) \in \Omega_2(0)$, then $U(x, 0) = 0$, and if $(x, 0) \in \Omega_2(x)$, then $U(x, 0) = \frac{\omega_1(x)}{2} < \frac{x}{2}$, so in either case, $U(x, 0) < \frac{x}{2}$. Similarly, $U(0, x) < \frac{x}{2}$. This contradicts Lemma C1, which completes the proof. \square

Lemma C6. *Suppose that for some $x \in (0, L)$ and $i \in \{1, 2\}$, $\omega_i(x) > x$. Then for all $y \in (x, \omega_i(x))$, $\omega_i(y) > y$, and for all $y \in \left[\frac{x+\omega_i(x)}{2}, L\right]$, $\omega_i(y) \geq \omega_i(x)$.*

Proof. Assume throughout the proof that $i = 1$; the proof for $i = 2$ is identical. Take some $y > x$; if $\omega_1(y) \geq \omega_1(x)$, then both statements hold (because $\omega_1(x) > x$) and we are done. So consider the remaining case, where $\omega_1(y) < \omega_1(x)$.

Let us show that $\omega_1(y) = 0$ is impossible. Indeed, if $\omega_1(y) = 0$, then $U(y, 0) = 0$; $U(y, x) = \min\{x, y\} = x$, and $U(0, x) \leq \frac{x}{2}$, because $\omega_2(x) \leq x$ as $\omega_1(x) > x$. We thus have $U(y, x) > U(y, 0) + U(0, x)$, which contradicts Lemma C1. We thus proceed assuming that $\omega_1(y) \in (0, \omega_1(x))$.

Take $\varepsilon > 0$ such that $\varepsilon < \omega_1(x) - \omega_1(y)$, $\varepsilon < \frac{\omega_1(x)-x}{2}$, and also $\varepsilon < \omega_1(x) - y$ if the latter is positive. Then $\varepsilon < \omega_1(x) - \omega_1(y)$ implies $(y, \omega_1(x) - \varepsilon) \in \Omega_0$, and thus $U(y, \omega_1(x) - \varepsilon) = \min\{y, \omega_1(x) - \varepsilon\}$.

Notice that we have $U\left(y, \frac{\omega_1(y)}{2}\right) \leq \frac{\omega_1(y)}{2}$, because $\omega_1(y) > 0$ implies $\left(y, \frac{\omega_1(y)}{2}\right) \in \Omega_1(y)$ by Lemma C3, and we have $V(y, z) = \min\{y, z\} \leq z$ for all $z \in [0, \omega_1(y)]$, which implies that the average cannot exceed $\frac{\omega_1(y)}{2}$. In addition, we have $U\left(\frac{x}{2}, \omega_1(x) - \varepsilon\right) \leq \frac{x}{2}$ (this is true for obvious reasons if $\left(\frac{x}{2}, \omega_1(x) - \varepsilon\right) \in \Omega_0$ or $\left(\frac{x}{2}, \omega_1(x) - \varepsilon\right) \in \Omega_1\left(\frac{x}{2}\right)$, whereas if $\left(\frac{x}{2}, \omega_1(x) - \varepsilon\right) \in \Omega_2(\omega_1(x) - \varepsilon)$ it follows since $\omega_2(\omega_1(x) - \varepsilon) \leq x$, which is true because $(x, \omega_1(x) - \varepsilon) \in \Omega_1(x)$). We now use Lemma C1 to obtain:

$$\frac{x}{2} + \frac{\omega_1(y)}{2} \geq U\left(\frac{x}{2}, \omega_1(x) - \varepsilon\right) + U\left(y, \frac{\omega_1(y)}{2}\right) \geq U(y, \omega_1(x) - \varepsilon) = \min\{y, \omega_1(x) - \varepsilon\}.$$

Now consider the following cases. Suppose that $y < \omega_1(x)$, we have $\min\{y, \omega_1(x) - \varepsilon\} = y$ by the choice of ε , which implies $\omega_1(y) \geq 2y - x > y$, since $y > x$. Moreover, if $y \geq \frac{x+\omega_1(x)}{2}$, we get $\omega_1(y) \geq 2y - x \geq x + \omega_1(x) - x = \omega_1(x)$. Now suppose that $y \geq \omega_1(x)$, we have $\min\{y, \omega_1(x) - \varepsilon\} = \omega_1(x) - \varepsilon$, which implies $\omega_1(y) \geq 2\omega_1(x) - 2\varepsilon - x > \omega_1(x) + (\omega_1(x) - x - 2\varepsilon) > \omega_1(x)$ by the choice of ε . This completes the proof. \square

Lemma C7. *Suppose that for some $z \in (0, L)$ and $i \in \{1, 2\}$, $\omega_i(z) = z$. Then for any $x \in (0, z)$, $\omega_1(x) \leq z$ and $\omega_2(x) \leq z$, and for any $x \in (z, L)$, $\omega_1(x) \geq z$ and $\omega_2(x) \geq z$.*

Proof. Suppose that $i = 1$, so $\omega_1(z) = z$ (the case $i = 2$ is similar).

First, we show that $\omega_2(x) \leq z$ for $x < z$. Indeed, for such x we have $(x, z) \in \Omega_1(x)$, and the result follows by definition of $\omega_2(x)$ (and Lemma C3).

Second, we show that $\omega_1(x) \geq z$ for $x > z$. Suppose not, so for some $x > z$ we have $\omega_1(x) < z$. Take any $r \in \left(\frac{\omega_1(x)+2z}{3}, z\right)$ (in this way, $r > \omega_1(x)$) and consider the type (x, r) . We must have $(x, r) \in \Omega_0$, so $U(x, r) = r$. Notice that since $(z, r) \in \Omega_1(z)$ due to $\omega_1(z) = z > r$, we have $U(z, r) = \frac{z}{2}$, and also $U\left(x, \frac{r+\omega_1(x)}{4}\right) \leq \frac{r+\omega_1(x)}{4}$ (indeed, if $\left(x, \frac{r+\omega_1(x)}{4}\right) \in \Omega_0$ then it holds as equality, whereas if $\left(x, \frac{r+\omega_1(x)}{4}\right) \in \Omega_1(x)$, then $U\left(x, \frac{r+\omega_1(x)}{4}\right) = \frac{\omega_1(x)}{2} < \frac{r+\omega_1(x)}{4}$ as we took $r > \omega_1(x)$). We thus have

$$\frac{z}{2} + \frac{r + \omega_1(x)}{4} \geq U(z, r) + U\left(x, \frac{r + \omega_1(x)}{4}\right) \geq U(x, r) = r,$$

which implies $2z + \omega_1(x) \geq 3r$. But this contradicts the choice of r ; this contradiction proves that we must have $\omega_1(x) \geq z$.

Third, we show that $\omega_2(x) \geq z$ for $x > z$. Suppose not, so for some $x > z$ we have $\omega_2(x) < z$. Consider the type (z, x) ; we must have $(z, x) \in \Omega_0$, so $U(z, x) = z$. Notice that we have $U\left(z, \frac{z}{2}\right) = \frac{z}{2}$, and $U\left(\frac{z}{4}, x\right) < \frac{z}{2}$ (indeed, if $\left(\frac{z}{4}, x\right) \in \Omega_0$, then $U\left(\frac{z}{4}, x\right) = \frac{z}{4} < \frac{z}{2}$; if $\left(\frac{z}{4}, x\right) \in \Omega_1\left(\frac{z}{4}\right)$, then $U\left(\frac{z}{4}, x\right) \leq \frac{z}{4} < \frac{z}{2}$ because all types $\left(\frac{z}{4}, y\right) \in \Omega_1\left(\frac{z}{4}\right)$ have $V\left(\frac{z}{4}, y\right) \leq \frac{z}{4}$; and if $\left(\frac{z}{4}, x\right) \in \Omega_2(x)$, then $U\left(\frac{z}{4}, x\right) \leq \frac{\omega_2(x)}{2} < \frac{z}{2}$ as we asserted that $\omega_2(x) < z$). Thus, by Lemma C1, we have

$$\frac{z}{2} + \frac{z}{2} > U\left(z, \frac{z}{2}\right) + U\left(\frac{z}{4}, x\right) \geq U(z, x) = z.$$

This contradiction shows that $\omega_2(x) \geq z$.

Fourth, we show that $\omega_1(x) \leq z$ for $x < z$. Indeed, for such x we have $(x, r) \in \Omega_2(r)$ for any $r \in (z, L)$ as we just proved, and $\omega_1(x) \leq z$ follows from the definition of $\omega_1(x)$. This completes the proof. \square

Lemma C8. *Suppose that for some $x, y \in (0, L)$, $\omega_1(x) > x$ and $\omega_2(y) > y$. Then there exists $z \in (\min\{x, y\}, \max\{x, y\})$ such that $\omega_i(z) = z$ for some $i \in \{1, 2\}$.*

Proof. It is impossible that $x = y$, for otherwise (x, x) would have to be both in $\Omega_1(x)$ and $\Omega_2(x)$. Suppose $x < y$; the case $x > y$ is considered similarly.

Define z by $z = \sup\{r \in (x, y) : \omega_1(r) > r\}$; it is well-defined because the set is nonempty, because by Lemma C6, any $x' \in (x, \min\{\omega_1(x), y\})$ belongs to it. Let us

show that it satisfies the requirement. Consider three cases. First, if $\omega_1(z) = z$, then we are done.

Second, suppose $\omega_1(z) > z$. This implies, in particular, that $z \neq y$, because $\omega_1(z) > z$ and $\omega_2(z) > z$ are incompatible. Now, by Lemma C6, it must be that for $s \in (z, \min\{\omega_1(z), y\})$, $\omega_1(s) > s$. This, however, violates the definition of z as the supremum.

Third, suppose that $\omega_1(z) < z$; this implies $z > x$ since $\omega_1(x) > x$. By Lemma C5, $\omega_2(z) \geq z$. By the definition of supremum, there are arbitrarily small $\varepsilon \in (0, z - x)$ such that $\omega_1(z - \varepsilon) > z - \varepsilon$, in particular, one can pick $\varepsilon < z - \omega_1(z)$. Notice that $\omega_2(z) \geq z$ implies that $\omega_1(z - \varepsilon) \leq z$. Take $\delta \in (z - \omega_1(z - \varepsilon), \varepsilon)$ and consider the point $(z, z - \delta)$. Since $\omega_1(z - \varepsilon) > z - \delta$ by the choice of δ , we have $\omega_2(z - \delta) \leq z - \varepsilon$, and since also $\omega_1(z) < z - \varepsilon < z - \delta$, we have $(z, z - \delta) \in \Omega_0$, which implies $U(z, z - \delta) = z - \delta$. Now take $\theta \in (0, \frac{z - \varepsilon}{2})$; we have $U(\theta, z - \delta) \leq \frac{z - \varepsilon}{2}$ (if $(\theta, z - \delta) \in \Omega_2(z - \delta)$ it follows from $\omega_2(z - \delta) \leq z - \varepsilon$, and if $(\theta, z - \delta) \in \Omega_0$ or $(\theta, z - \delta) \in \Omega_1(\theta)$ this is obviously true). By similar logic, we have $U(z, \theta) < \frac{z - \varepsilon}{2}$ because $\omega_1(z) < z - \varepsilon$. Then by Lemma C1,

$$z - \varepsilon \geq U(\theta, z - \delta) + U(z, \theta) \geq U(z, z - \delta) = z - \delta,$$

which implies $\delta \geq \varepsilon$. However, this contradicts the choice of δ . This contradiction completes the proof. \square

Lemma C9. *The set of points $\{z : \omega_1(z) = z \neq \omega_2(z)\}$ is at most countable, and so is $\{z : \omega_1(z) \neq z = \omega_2(z)\}$.*

Proof. It suffices to prove the result for the first set. This set is a union of two sets, $Z_1 = \{z : \omega_1(z) = z < \omega_2(z)\}$ and $Z_2 = \{z : \omega_1(z) = z > \omega_2(z)\}$. Let us show that if $z \in Z_1$, then there is $\varepsilon > 0$ such that $(z, z + \varepsilon) \cap Z_1 = \emptyset$. This is trivial, since it suffices to take $\varepsilon = \omega_2(z) - \omega_1(z)$; then for $x \in (z, z + \varepsilon)$, $\omega_1(x) = x$ would contradict Lemma C6. This means that there is a bijection from Z_1 to a certain set of nonintersecting intervals, which is at most countable (since each interval contains a rational number).

Now let us prove that if $z \in Z_2$, then there is $\varepsilon > 0$ such that $(z - \varepsilon, z) \cap Z_2 = \emptyset$. Take $\varepsilon = z - \omega_2(z)$ and consider any $r \in (z - \varepsilon, z) = (\omega_2(z), z)$. Suppose, to obtain a contradiction, that $r \in Z_2$, then $\omega_1(r) = r > \omega_2(r)$. Consider the point (r, z) . Since $\omega_1(r) = r < z$ and $\omega_2(z) < r$ by the choice of r , we have $(r, z) \in \Omega_0$, and thus $U(r, z) = r$. Now take $\delta \in (0, \min\{\omega_1(r), \max\{\omega_2(z), \frac{r}{2}\}\})$ sufficiently small, we would have $U(\delta, z) < \frac{r}{2}$ (if $\omega_2(z) > 0$, this follows from $U(\delta, z) \leq \frac{\omega_2(z)}{2} < \frac{r}{2}$, and if $\omega_2(z) = 0$, then it

follows straightforwardly both if $(\delta, z) \in \Omega_0$ and if $(\delta, z) \in \Omega_1(\delta)$, and $U(r, \delta) = \frac{\omega_1(r)}{2} = \frac{r}{2}$. This implies

$$\frac{r}{2} + \frac{r}{2} > U(\delta, z) + U(r, \delta) \geq U(r, z) = r,$$

a contradiction, which shows that for the chosen ε , $(z - \varepsilon, z) \cap Z_2 = \emptyset$. But then, as before, we obtain a bijection from Z_2 to a certain set of nonintersecting intervals, which is at most countable. This completes the proof. \square

Lemma C10. *Suppose that \mathcal{M} is an admissible mechanism. Then there is an admissible mechanism \mathcal{M}' with $\mathcal{W}(\mathcal{M}') = \mathcal{W}(\mathcal{M})$ and such that the sets $\Omega'_1(0)$ and $\Omega'_2(0)$ are empty.*

Proof. Define \mathcal{M}' as follows. Take any (a_1, a_2) . If $(a_1, a_2) \in \Omega_1(0)$, then let $(a_1, a_2) \in \Omega'_0$, and $p'_1(a_1, a_2) = 1$, $p'_2(a_1, a_2) = 0$ (so dimension 1 is verified for sure). Similarly, if $(a_1, a_2) \in \Omega_2(0)$, then let $(a_1, a_2) \in \Omega'_0$, and $p'_1(a_1, a_2) = 0$, $p'_2(a_1, a_2) = 1$. Otherwise, let (a_1, a_2) send the same message in \mathcal{M}' as in \mathcal{M} , and let the verification probabilities for all messages not defined earlier be the same in \mathcal{M}' as in \mathcal{M} (with the exception of $(0, *)$ and $(*, 0)$, which are not allowed under the new mechanism). The posterior beliefs are defined in the natural way.

Notice that in \mathcal{M}' , each type gets the same equilibrium payoff as in \mathcal{M} , and a deviation by any type (x, y) to send the message that type (a_1, a_2) sends in equilibrium will result in the same payoff under \mathcal{M}' as under \mathcal{M} . This implies that all incentive compatibility constraints are satisfied, and the mechanism \mathcal{M}' is admissible and satisfies the requirement. This completes the proof. \square

Lemma C11. *Suppose that \mathcal{M} is an admissible mechanism. Then there is an admissible mechanism \mathcal{M}' with $\mathcal{W}(\mathcal{M}') = \mathcal{W}(\mathcal{M})$ such that the conclusions of Lemma C10 are satisfied, and also for $z \in (0, L)$, $\omega'_1(z) = z$ implies $\omega'_2(z) \geq z$ and $\omega'_2(z) = z$ implies $\omega'_1(z) \geq z$.*

Proof. By Lemma C10 we may assume that \mathcal{M} satisfies its requirements. Let us define the mechanism \mathcal{M}' as follows. Take any (a_1, a_2) . If $0 \leq a_2 < a_1 < L$ and $\omega_1(a_1) < a_1 = \omega_2(a_1)$, then type (a_1, a_2) will send message $(a_1, *)$ under \mathcal{M}' (and after such message, dimension 1 will be verified for sure); in this case, the set of types that send message $(a_1, *)$ would be $\{(a_1, a_2) : a_2 < a_1\}$, which is connected. If $0 \leq a_1 < a_2 < L$ and $\omega_2(a_2) < a_2 = \omega_1(a_2)$, then type (a_1, a_2) will send message $(*, a_2)$ under \mathcal{M}' (and after such message,

dimension 2 will be verified for sure); then, similarly, message $(*, a_2)$ is sent by a connected set of types. In all other cases, type (a_1, a_2) will send the same message under \mathcal{M}' as under \mathcal{M} , and verification probabilities following these messages will be the same. The posterior beliefs are defined in a natural way.

Notice that in the new mechanism \mathcal{M}' , each message is sent by a connected set of types. To see this, notice that if a_1 is such that $\omega_1(a_1) < a_1 = \omega_2(a_1)$, then message $(a_1, *)$ is sent by a connected set in \mathcal{M}' by construction. Otherwise, if a_1 is such that $\omega_1(a_1) < a_1 = \omega_2(a_1)$ does not hold, then the set of types that send $(a_1, *)$ in \mathcal{M}' is a subset of the set of those that send this message in \mathcal{M} (i.e., $\Omega'_1(a_1) \subset \Omega_1(a_1)$), which implies that $\omega'_1(a_1) \leq \omega_1(a_1)$. Let us prove that $a_2 < \omega_2(a_1)$ implies $(a_1, a_2) \in \Omega'_1(a_1)$; this would immediately imply that $\Omega'_1(a_1)$ is connected. By construction, if $\omega_1(a_1) < a_1 = \omega_2(a_1)$ does not hold, then $(a_1, a_2) \notin \Omega'_1(a_1)$ is possible only if $a_2 \in (a_1, L)$ and $\omega_2(a_2) < a_2 = \omega_1(a_2)$. By Lemma C6, $\omega_1(a_2) = a_2$ implies, in particular, that $\omega_2(a_1) \leq a_2$, because $a_1 < a_2$. But we took $a_2 < \omega_2(a_1)$, so we get a contradiction, showing that $\Omega'_1(a_1)$ is connected for each a_1 . We can similarly show that $\Omega'_2(a_2)$ is connected for each a_2 . This shows that \mathcal{M}' is connected, and it is semi-direct by construction.

It is easy to see, again by construction, that mechanism \mathcal{M}' satisfies the required property (that for $z \in (0, L)$, $\omega'_1(z) = z$ implies $\omega'_2(z) \geq z$ and $\omega'_2(z) = z$ implies $\omega'_1(z) \geq z$). Furthermore, by Lemma C9, the set of types (a_1, a_2) for which payoffs may have changed has measure zero, and as a result $\mathcal{W}(\mathcal{M}') = \mathcal{W}(\mathcal{M})$. It remains to prove that the new mechanism is incentive compatible.

We need to show that no type (x, y) could benefit from deviating and sending some other equilibrium message (we may assume that non-equilibrium messages are forbidden). Suppose that it benefits from sending a message that type (a_1, a_2) sent. It suffices to consider cases where either (a) $\max\{a_1, a_2\} \in (0, L)$, $a_1 \neq a_2$, and $\omega'_1(\max\{a_1, a_2\}) = \omega'_2(\max\{a_1, a_2\}) = \max\{a_1, a_2\}$, or (b) $\max\{x, y\} \in (0, L)$, $x \neq y$, and $\omega'_1(\max\{x, y\}) = \omega'_2(\max\{x, y\}) = \max\{x, y\}$ (or perhaps both); in other cases, the deviation cannot be profitable because it was not under mechanism \mathcal{M} .

Consider Case (a) first. Without loss of generality, suppose $a_1 > a_2$; the opposite case is considered similarly. In \mathcal{M}' , (a_1, a_2) sends message $(a_1, *)$, so for the deviation to be profitable, it must be that $x = a_1$ and $y \geq a_1$ (because $\omega'_1(a_1) = a_1$). For such a type, a deviation yields a payoff $U'(a_1, a_2) = \frac{a_1}{2}$. On the other hand, in case $(x, y) \in \Omega'_0$, we have $U'(x, y) = \min\{x, y\} = a_1$. In case $(x, y) \in \Omega'_2(y)$, notice that by construction, $\omega'_1(a_1) = \omega'_2(a_1) = a_1$ may only be true if we had either $\omega_1(a_1) = a_1$ or $\omega_2(a_1) = a_1$ (or both) in mechanism \mathcal{M} ; however, by Lemma C7, this means that $\omega_2(y) \geq a_1$, and by construction, this property is preserved in \mathcal{M}' , so $\omega'_2(y) \geq a_1$, which implies that

$U'(x, y) \geq \frac{a_1}{2}$ in this case. So, in either case, the deviation is not profitable, and the case $a_1 < a_2$ is considered similarly.

Now consider Case (b), and again without loss of generality, suppose $x > y$. We have that type (x, y) gets $U'(x, y) = \frac{x}{2}$. A deviation to sending the same message as (a_1, a_2) may only be profitable if either $a_1 = x$ or $a_2 = y$. Suppose $a_1 = x$; for this to be a deviation it must be that $(a_1, a_2) \notin \Omega_1(x)$, and in particular $a_2 \geq x$. If $(a_1, a_2) \in \Omega'_2(a_2)$, then the deviation is caught for sure, hence is not profitable. If $(a_1, a_2) \in \Omega'_0$, it must be that $(a_1, a_2) \in \Omega_0$ in mechanism \mathcal{M} , and the verification probabilities were the same. In addition, under mechanism \mathcal{M} , type $(x, 0)$ got either payoff $U(x, 0) = 0$ or $U(x, 0) = \frac{\omega_1(x)}{2} \leq \frac{x}{2}$; thus, if (x, y) has a profitable deviation to (a_1, a_2) in \mathcal{M}' , then for type $(x, 0)$ this deviation was profitable under \mathcal{M} , which is impossible since \mathcal{M} is admissible. Now suppose that $a_2 = y$. If $a_1 < x$, then $\omega'_1(x) = \omega'_2(x) = x$ implies that $U(a_1, a_2) \leq \frac{x}{2}$, so this is not a profitable deviation. If $a_1 > x$, then similarly to Case (a), $\omega'_1(x) = \omega'_2(x) = x$ is only possible if either $\omega_1(x) = x$ or $\omega_2(x) = x$ (or both) in mechanism \mathcal{M} , and Lemma C7 implies that $\omega_1(a_1) \geq x$, which by construction means that $\omega'_1(a_1) \geq x$. This means that $(a_1, a_2) \in \Omega'_1(a_1)$, and the deviation to (a_1, a_2) will be detected for sure, which means that it is not profitable in this case as well. This completes the proof. \square

Lemma C12. *Suppose that \mathcal{M} is an admissible mechanism. Then there is an admissible mechanism \mathcal{M}' with $\mathcal{W}(\mathcal{M}') = \mathcal{W}(\mathcal{M})$ such that the conclusions of Lemmas C10 and C11 are satisfied, and also $\omega'_1(z) = \omega'_2(z) = z$ implies $(z, z) \in \Omega'_0$ and also $\omega'_1(L) = \omega'_2(L) = L$.*

Proof. By Lemmas C10 and C11 we may assume that \mathcal{M} satisfies their conclusions. Let us define the mechanism \mathcal{M}' as follows. We require that types (x, L) for $x < L$ send message $(*, L)$ (with dimension 2 verified for sure) and types (L, y) for $y < L$ send message $(L, *)$ (with dimension 1 verified for sure). We furthermore require that if $\omega_1(z) = \omega_2(z) = z$ for $z > 0$, then (z, z) sends message (z, z) , following which each dimension is verified with probability $\frac{1}{2}$ (in this way, the property $\omega'_1(z) = \omega'_2(z) = z$ is preserved). In all other cases, type (a_1, a_2) will send the same message under \mathcal{M}' as under \mathcal{M} , and verification probabilities following these messages will be the same. The beliefs are defined in a natural way. Then mechanism \mathcal{M}' satisfies the requirements (notice that $(0, 0) \in \Omega_0$ because \mathcal{M} satisfies the conclusions of Lemma C10, and thus $(0, 0) \in \Omega'_0$); furthermore, the payoffs may only have changed for a set of types of measure zero, so $\mathcal{W}(\mathcal{M}') = \mathcal{W}(\mathcal{M})$.

Let us show that mechanism \mathcal{M}' is incentive compatible. First, observe that any deviation to $(*, L)$ or $(L, *)$ will be caught, except if the deviating type is (L, L) , for which such

deviation is clearly unprofitable. If type (x, L) with $x < L$ deviates to reporting (L, L) , he moves from getting $U'(x, L) = \frac{L}{2}$ for sure to getting $U'(L, L) = L$ with probability $\frac{1}{2}$, so this is not profitable. Deviating to reporting $(x, *)$ cannot be profitable as such types cannot get more than $\frac{L}{2}$, so the only possible deviation by this type is to report (x, y) for $y < L$ and $(x, y) \in \Omega'_0$. This cannot be profitable if $x = y$, because this would imply that $\omega'_1(x) = \omega'_2(x) = x$, so by construction each dimension is verified with probability $\frac{1}{2}$, and in expectation such deviation would yield $\frac{x}{2} < \frac{L}{2}$. If $x \neq y$, then we must have $(x, y) \in \Omega_0$ in \mathcal{M} by construction. Then if type (x, L) had a profitable deviation to (x, y) in \mathcal{M}' , then so would type $(x, 0)$ in \mathcal{M}' (because its equilibrium payoff is $\frac{\omega_1(x)}{2}$), and since its payoff under \mathcal{M} is the same, this would be a profitable deviation under \mathcal{M} as well. However, this is impossible.

Let us now consider the deviations that involve types (z, z) in cases $\omega'_1(z) = \omega'_2(z) = z$, but not types (x, L) or (L, y) . Suppose some type (x, y) would benefit from deviating to (z, z) ; then either $x = z$ or $y = z$. Suppose $x = z$; if $x < z$, we have $U'(x, z) = \frac{z}{2}$, and in case of deviation the expected payoff would be $\frac{z}{2}$ as well, because each dimension is verified with probability $\frac{1}{2}$. If instead $x > z$, then $\omega'_1(z) = \omega'_2(z) = z$ means that $\omega_1(z) = \omega_2(z) = z$ (because $z < L$ in this case), and this means that $\omega_1(x) \geq z$ by Lemma C7, so $U'(x, z) \geq \frac{z}{2}$, so this is not a profitable deviation. We would similarly prove that deviations from (x, y) with $y = z$ cannot be profitable.

The remaining case to consider is deviation from type (z, z) to some type that was not affected by the modifications above. By construction, the payoff of type (z, z) is $U'(z, z) = z$, whereas under mechanism \mathcal{M} this type's payoff could not be higher (it was either z or $\frac{z}{2}$). Thus, if some deviation from type (z, z) to another type that was not affected by the modification is profitable under \mathcal{M}' , it would be profitable under \mathcal{M} , which is impossible. This completes the proof. \square

To formulate the next result, we introduce the following definition. We call a mechanism \mathcal{M} *pseudo-admissible* if it is connected, semi-direct, satisfies the Bayesian updating property, $\Omega_1(0) = \Omega_2(0) = \emptyset$, $\omega_1(z) > 0$ implies $(z, 0) \in \Omega_1(z)$, $\omega_2(z) > 0$ implies $(0, z) \in \Omega_2(z)$, $\omega_1(z) \geq z$ for each $z \in [0, L]$, and for each $(a_1, a_2) \in \Omega_0$, $\omega_1(a_1)\omega_2(a_2) \geq (a_1)^2$. Notice that we do not require \mathcal{M} to be incentive compatible; this requirement is replaced by: (a) a requirement that may be fulfilled by a modification of any admissible mechanism by Lemma C10; (b) a property that holds for any admissible mechanism by Lemma C3; and (c) by two conditions on $\omega_1(\cdot)$ and $\omega_2(\cdot)$ that will be satisfied by the mechanism that we construct in the proof of the next Lemma.

Lemma C13. *Suppose that \mathcal{M} is an admissible mechanism. Then there is a pseudo-admissible mechanism \mathcal{M}' with $\mathcal{W}(\mathcal{M}') = \mathcal{W}(\mathcal{M})$.*

Proof. We may assume that mechanism \mathcal{M} satisfies the conclusions of Lemmas C10, C11, and C12. Throughout this proof, we will, for a vector $a = (a_1, a_2)$, define $\|a\| = \max\{a_1, a_2\}$.

Define sets Z_0, Z_1, Z_2 as follows. For type (a_1, a_2) , if $\omega_1(\|a\|) = \omega_2(\|a\|) = \|a\|$, then $(a_1, a_2) \in Z_0$; if $\omega_1(\|a\|) > \|a\| \geq \omega_2(\|a\|)$, then $(a_1, a_2) \in Z_1$, and if $\omega_2(\|a\|) > \|a\| \geq \omega_1(\|a\|)$, then $(a_1, a_2) \in Z_2$. These three sets are pairwise disjoint, furthermore, we have $Z_0 \cup Z_1 \cup Z_2 = \Omega$ (indeed, $\omega_1(\|a\|) > \|a\|$ and $\omega_2(\|a\|) > \|a\|$ are incompatible; $\omega_1(\|a\|) < \|a\|$ and $\omega_2(\|a\|) < \|a\|$ is ruled out by Lemma C5; whereas $\omega_1(\|a\|) = \|a\| > \omega_2(\|a\|)$ is ruled out because the conclusions of Lemma C11 and C12 are satisfied; the same is true for $\omega_2(\|a\|) = \|a\| > \omega_1(\|a\|)$). Importantly, there is symmetry, in that $(a_1, a_2) \in Z_i$ if and only if $(a_2, a_1) \in Z_i$, for each $i \in \{0, 1, 2\}$.

Let us prove that if two types send the same message under \mathcal{M} , then they are in the same Z_i . To see this, notice that if $(a_1, a_2) \in \Omega_0$, then it is the only type that sends this message, so the statement holds trivially. Suppose that $(a_1, a_2), (a_1, y) \in \Omega_1(a_1)$; without loss of generality suppose that $a_2 < y$. Now consider the following cases. If $a_2 < y \leq a_1$, then $\|a\| = \|(a_1, y)\| = a_1$, and the two types belong to the same set by construction. If $a_2 \leq a_1 < y$, then $(a_1, y) \in \Omega_1(a_1)$ implies $\omega_1(a_1) > a_1$. This means that $\|a\| = a_1 < \omega_1(a_1)$, so $(a_1, a_2) \in Z_1$. Now, $(a_1, y) \in \Omega_1(a_1)$ implies $\omega_2(y) \leq a_1 < y$, and then we must have $\omega_1(y) > y$ (otherwise, this would contradict the conclusions of Lemma C11 if $y < L$ or Lemma C12 if $y = L$), which implies $(a_1, y) \in Z_1$. In the remaining case $a_1 < a_2 < y$, $(a_1, y) \in \Omega_1(a_1)$ implies $\omega_1(a_1) \geq y > a_2$, and $\|a\| = a_2 < \omega_1(a_1)$, so $(a_1, a_2) \in Z_1$. Similarly to the previous case, $(a_1, y) \in \Omega_1(a_1)$ implies $\omega_2(y) \leq a_1 < y$, and thus $\omega_1(y) > y$, so $(a_1, y) \in Z_1$. So, if $(a_1, a_2), (a_1, y) \in \Omega_1(a_1)$, then the two types belong to the same Z_i , and we can consider the types $(a_1, a_2), (x, a_2) \in \Omega_2(a_2)$ in a similar way. We have thus established that if two types send the same message under \mathcal{M} , then they are in the same Z_i .

Define mechanism \mathcal{M}' as follows. If $(a_1, a_2) \in Z_0 \cup Z_1$, then type (a_1, a_2) sends the same message in \mathcal{M}' as in \mathcal{M} . If $(a_1, a_2) \in Z_2$, then type (a_1, a_2) sends the message “symmetric” to what type (a_2, a_1) did in equilibrium. Specifically, if $(a_2, a_1) \in \Omega_0$, then (a_1, a_2) sends message (a_1, a_2) . If $(a_2, a_1) \in \Omega_1(a_2)$, then (a_1, a_2) sends message $(*, a_2)$. If $(a_2, a_1) \in \Omega_2(a_1)$, then (a_1, a_2) sends message $(a_1, *)$. The beliefs are defined in the natural way. Notice that we do not have to define verification probabilities (though this may be done in a natural way), because we do not claim incentive compatibility of \mathcal{M}' . Given the properties above, we have that if two types (a_1, a_2) and (x, y) send the same

message in \mathcal{M} , then if they are in $Z_0 \cup Z_1$ they send the same message in \mathcal{M}' , and if they are in Z_2 , then (a_2, a_1) and (y, x) send the same message in \mathcal{M}' . We therefore have $\mathcal{W}(\mathcal{M}') = \mathcal{W}(\mathcal{M})$; furthermore, by construction, \mathcal{M}' satisfies all the properties of a pseudo-admissible mechanism, except perhaps for the last one. We thus need to prove that for $(a_1, a_2) \in \Omega'_0$, $\omega'_1(a_1)\omega'_2(a_2) \geq (a_1)^2$.

Let us first show that for any $(a_1, a_2) \in \Omega'_0$, $U'(a_1, a_2) \leq U'(a_1, 0) + U'(0, a_2)$. Suppose first that $(a_1, a_2) \in Z_0$; then by definition of Z_0 , we must have $a_1 = a_2 = \|a\|$, and $\omega_1(\|a\|) = \omega_2(\|a\|) = \|a\|$. By construction, we have $\omega'_1(\|a\|) = \omega'_2(\|a\|) = \|a\|$ as well. Thus, $U'(a_1, a_2) = \|a\|$ and $U'(a_1, 0) = U'(0, a_2) = \frac{\|a\|}{2}$, so the inequality is satisfied.

Consider the case $(a_1, a_2) \in Z_1$. This means $\omega_1(\|a\|) > \|a\| \geq \omega_2(\|a\|)$, and by construction $\omega'_1(\|a\|) > \|a\| \geq \omega'_2(\|a\|)$; now, $(a_1, a_2) \in \Omega'_0$ is only possible if $a_1 < a_2$. Now, $(a_1, a_2) \in \Omega'_0$ implies $\omega'_2(a_2) \leq a_1$, and since $(a_1, a_2) \in Z_1$, we have $\omega_2(a_2) \leq a_1$.

There are three possibilities. Suppose that $(a_1, 0) \in Z_0$. Then $\omega_1(a_1) = a_1$, and by Lemma C7, $\omega_2(a_2) \geq a_1$. Thus, we have $\omega'_2(a_2) = \omega_2(a_2) = a_1$, and consequently $U'(a_1, a_2) = a_1$, $U'(a_1, 0) = \frac{\omega'_1(a_1)}{2} = \frac{\omega_1(a_1)}{2} = \frac{a_1}{2}$, and $U'(0, a_2) = \frac{\omega'_2(a_2)}{2} = a_1$, and the inequality follows.

Now suppose that $(a_1, 0) \in Z_1$. In this case, the inequality $U(a_1, a_2) \leq U(a_1, 0) + U(0, a_2)$, which holds by Lemma C1, implies $U'(a_1, a_2) \leq U'(a_1, 0) + U'(0, a_2)$ because the corresponding payoffs are identical under \mathcal{M} and under \mathcal{M}' .

The last possibility is that $(a_1, 0) \in Z_2$; in other words, $\omega_2(a_1) > a_1$. We also have $\omega_2(a_2) \leq a_1 < a_2$, and thus by Lemma C8 there is some $z \in (a_1, a_2]$ such that $\omega_i(z) = z$ for some $i \in \{1, 2\}$. If $z < a_2$, then by Lemma C7, we must have $\omega_2(a_2) \geq z > a_1$, which contradicts $\omega_2(a_2) \leq a_1$, whereas if $z = a_2$, then $\omega_2(a_2) \geq z > a_1$ follows from the fact that \mathcal{M} satisfies the conclusion of Lemma C11, again a contradiction. Thus, we have shown that $(a_1, a_2) \in Z_1$ implies that $U'(a_1, a_2) \leq U'(a_1, 0) + U'(0, a_2)$ holds.

Now consider the case $(a_1, a_2) \in Z_2$; it is very similar to the previous case. We have $\omega_2(\|a\|) > \|a\| \geq \omega_1(\|a\|)$ and thus, by construction, $\omega'_1(\|a\|) > \|a\| \geq \omega'_2(\|a\|)$. As before, $(a_1, a_2) \in \Omega'_0$ implies $a_1 < a_2$, and we must have $\omega'_2(a_2) \leq a_1$; since $(a_1, a_2) \in Z_2$, we must have $\omega_1(a_2) \leq a_1$.

Again, consider three possibilities. Suppose that $(a_1, 0) \in Z_0$. Then $\omega_1(a_1) = a_1$, and by Lemma C7, $\omega_1(a_2) \geq a_1$. Thus, we have $\omega'_2(a_2) = \omega_1(a_2) = a_1$, and consequently $U'(a_1, a_2) = a_1$, $U'(a_1, 0) = \frac{\omega'_1(a_1)}{2} = \frac{\omega_1(a_1)}{2} = \frac{a_1}{2}$, and $U'(0, a_2) = \frac{\omega'_2(a_2)}{2} = a_1$, and the inequality follows.

Now suppose that $(a_1, 0) \in Z_1$, which means $\omega_1(a_1) > a_1$. Since $\|a\| = a_2$, we also have $\omega_2(a_2) > a_2$, which, by Lemma C8, implies that there is some $z \in (a_1, a_2]$ such that $\omega_i(z) = z$ for some $i \in \{1, 2\}$. If $z < a_2$, then by Lemma C7 we must have

$\omega_1(a_2) \geq z > a_1$, and if $z = a_2$, we have $\omega_1(a_2) \geq z > a_1$ because \mathcal{M} satisfies the conclusion of Lemma C11; in either case we get a contradiction with $\omega_1(a_2) \leq a_1$.

The last case is $(a_1, 0) \in Z_2$. By Lemma C1 applied to mechanism \mathcal{M} , we have $U(a_2, a_1) \leq U(0, a_1) + U(a_2, 0)$, and since we have in this case $U'(a_1, a_2) = U(a_2, a_1)$, $U'(a_1, 0) = U(0, a_1)$, $U'(0, a_2) = U(a_2, 0)$, we get the required inequality $U'(a_1, a_2) \leq U'(a_1, 0) + U'(0, a_2)$ in this case as well.

Now let us use this inequality to prove that for $(a_1, a_2) \in \Omega'_0$, $\omega'_1(a_1)\omega'_2(a_2) \geq (a_1)^2$. Notice that if $a_1 = 0$ then the inequality is satisfied automatically, so suppose $a_1 > 0$. Since $\omega'_1(a_1) \geq a_1$, we have $a_2 \geq a_1$, and thus $U'(a_1, a_2) = \min\{a_1, a_2\} = a_1$. We also have $U'(a_1, 0) = \frac{a_1}{\omega'_1(a_1)} \frac{a_1}{2} + \frac{\omega'_1(a_1) - a_1}{\omega'_1(a_1)} a_1 = a_1 - \frac{(a_1)^2}{2\omega'_1(a_1)}$, and $U'(0, a_2) = \frac{\omega'_2(a_2)}{2}$. Using the inequality we proved earlier, we have

$$a_1 = U'(a_1, a_2) \leq U'(a_1, 0) + U'(0, a_2) = a_1 - \frac{(a_1)^2}{2\omega'_1(a_1)} + \frac{\omega'_2(a_2)}{2},$$

so $\frac{\omega'_2(a_2)}{2} \geq \frac{(a_1)^2}{2\omega'_1(a_1)}$, which implies $\omega'_1(a_1)\omega'_2(a_2) \geq (a_1)^2$. This completes the proof that \mathcal{M}' is a pseudo-admissible mechanism. \square

Lemma C14. *Suppose that \mathcal{M} is a pseudo-admissible mechanism. Then*

$$\mathcal{W}(\mathcal{M}) = \frac{1}{L^2} \int_0^L \left(a_1 - \frac{(a_1)^2}{2\omega_1(a_1)} \right)^2 da_1 + \frac{1}{L^2} \int_0^L \left(\frac{\omega_2(a_2)}{2} \right)^2 da_2.$$

Proof. This follows immediately from Lemma C4 in case $t = 1$, where we simplified $\omega_1(a_1) - \frac{(\omega_1(a_1) - a_1)^2}{\omega_1(a_1)} = 2 \left(a_1 - \frac{(a_1)^2}{2\omega_1(a_1)} \right)$. \square

Lemma C15. *Suppose that \mathcal{M} is a pseudo-admissible mechanism. Then there is a pseudo-admissible mechanism \mathcal{M}' such that for each z , $\omega'_1(z) \leq \omega_1(z)$, $\omega'_2(z) \leq \omega_2(z)$ and $\omega'_1(z)$ and $\omega'_2(z)$ are monotonically increasing; in this mechanism, $\mathcal{W}(\mathcal{M}') \leq \mathcal{W}(\mathcal{M})$.*

Proof. Define mechanism \mathcal{M}' as follows. Let type (a_1, a_2) send message $(a_1, *)$ in \mathcal{M}' if and only if all $x \geq a_1$, type (x, a_2) sends message $(x, *)$ in \mathcal{M} . Let type (a_1, a_2) send message $(*, a_2)$ in \mathcal{M}' if and only if all $y \geq a_2$, type (a_1, y) sends message $(*, y)$ in \mathcal{M} . (Notice that no type (a_1, a_2) satisfies both properties, so these definitions are not contradictory.) Otherwise, let type (a_1, a_2) send message (a_1, a_2) . Since incentive compatibility is not required for pseudo-admissible mechanisms, the probabilities of verification are immaterial. The posterior beliefs are defined in a natural way.

In mechanism \mathcal{M}' , we have $\omega'_1(a_1) = \inf_{x \geq a_1} \omega_1(x)$; this implies that $\omega'_1(a_1) \leq \omega_1(a_1)$, $\omega'_1(a_1)$ is monotonically increasing, and also $\omega'_1(a_1) \geq a_1$ (since $\omega_1(x) \geq x \geq a_1$ for each $x \geq a_1$). Similarly, $\omega'_2(a_2) = \inf_{y \geq a_2} \omega_2(y)$; this implies that $\omega'_2(a_2) \leq \omega_2(a_2)$ and $\omega'_2(a_2)$ is monotonically increasing. Furthermore, Lemma C14 implies $\mathcal{W}(\mathcal{M}') \leq \mathcal{W}(\mathcal{M})$, because a decrease in $\omega_1(a_1)$ or $\omega_2(a_2)$ may only make $\mathcal{W}(\mathcal{M})$ smaller (notice that $a_1 - \frac{(a_1)^2}{2\omega_1(a_1)} \geq 0$, because $\omega_1(a_1) \geq a_1$). It is easy to see that the other properties of a pseudo-admissible mechanism are satisfied as well, except that $\omega'_1(a_1)\omega'_2(a_2) \geq (a_1)^2$ for any $(a_1, a_2) \in \Omega'_0$, which we prove explicitly.

Take any $(a_1, a_2) \in \Omega'_0$, and let us prove that $\omega'_1(a_1)\omega'_2(a_2) \geq (a_1)^2$. Take any $\varepsilon > 0$. Let us show that there is $x \geq a_1$ such that $(x, a_2) \notin \Omega_1(x)$ and $\omega_1(x) \leq \omega'_1(a_1) + \varepsilon$. Indeed, suppose not; then there is $\tilde{x} \geq a_1$ satisfying $\omega_1(\tilde{x}) \leq \omega'_1(a_1) + \varepsilon$ by the definition of the infimum, and furthermore $(\tilde{x}, a_2) \in \Omega_1(\tilde{x})$. This is only possible if $\omega_1(\tilde{x}) \geq a_2$, which implies $\omega'_1(a_1) \geq a_2 - \varepsilon$. Now take any $\hat{x} \geq a_1$ with $\omega_1(\hat{x}) > \omega'_1(a_1) + \varepsilon$, then $\omega_1(\hat{x}) > a_2$, and thus $(\hat{x}, a_2) \in \Omega_1(\hat{x})$. However, this means that for all $\hat{x} \geq a_1$, regardless of whether $\omega_1(\hat{x}) \leq \omega'_1(a_1) + \varepsilon$ or not, $(\hat{x}, a_2) \in \Omega_1(\hat{x})$, which by construction of mechanism \mathcal{M}' means that (a_1, a_2) should send message $(a_1, *)$ in \mathcal{M}' , which contradicts $(a_1, a_2) \in \Omega'_0$.

We can similarly show that there is $y \geq a_2$ such that $(a_1, y) \notin \Omega_2(y)$ and $\omega_2(y) \leq \omega'_2(a_2) + \varepsilon$. Now, take such x and y and consider (x, y) . Notice that $(x, y) \in \Omega_0$: indeed, if $(x, y) \in \Omega_1(x)$, then $(x, a_2) \in \Omega_1(x)$, contradicting the choice of x ; similarly, if $(x, y) \in \Omega_2(y)$, then $(a_1, y) \in \Omega_2(y)$, contradicting the choice of y . Now, since \mathcal{M} is pseudo-admissible and $(x, y) \in \Omega_0$, we have $\omega_1(x)\omega_2(y) \geq x^2$. Therefore, using $x \geq a_1$, we have

$$\begin{aligned} \omega'_1(a_1)\omega'_2(a_2) - (a_1)^2 &\geq (\omega_1(x) - \varepsilon)(\omega_2(y) - \varepsilon) - x^2 \\ &\geq \varepsilon^2 - \varepsilon\omega_1(x) - \varepsilon\omega_2(y). \end{aligned}$$

Now, since $\omega_1(x)$ and $\omega_2(y)$ are bounded and ε may be chosen arbitrarily low, we must have $\omega'_1(a_1)\omega'_2(a_2) \geq (a_1)^2$. This shows that mechanism \mathcal{M}' is pseudo-admissible, which completes the proof. \square

Lemma C16. *For any pseudo-admissible mechanism with monotone $\omega_1(z)$ and $\omega_2(z)$,*

$$\mathcal{W}(\mathcal{M}) \geq \frac{2}{15}L + \frac{1}{L^2} \iint_A \frac{x^4}{2y^3} dx dy,$$

where A is defined as $A = \{(a_1, a_2) \in \Omega : a_2 \geq \omega_1(a_1)\}$.

Proof. We use the formula from Lemma C14. Take the first term and rewrite it as:

$$\begin{aligned}
& \frac{1}{L^2} \int_0^L \left(x - \frac{x^2}{2\omega_1(x)} \right)^2 dx \\
&= \frac{1}{L^2} \int_0^L \left(x - \frac{x^2}{2M} \right)^2 dx \\
&\quad - \frac{1}{L^2} \left(\int_0^L \left(x - \frac{x^2}{2M} \right)^2 dx - \int_0^L \left(x - \frac{x^2}{2\omega_1(x)} \right)^2 dx \right) \\
&= \frac{2}{15}L - \frac{1}{L^2} \int_0^L \left(\left(x - \frac{x^2}{2M} \right)^2 - \left(x - \frac{x^2}{2\omega_1(x)} \right)^2 \right) dx \\
&= \frac{2}{15}L - \frac{1}{L^2} \int_0^L \left(\int_{\omega_1(x)}^L 2 \left(x - \frac{x^2}{2y} \right) \frac{x^2}{2y^2} dy \right) dx \\
&= \frac{2}{15}L - \frac{1}{L^2} \iint_A \left(\frac{x^3}{y^2} - \frac{x^4}{2y^3} \right) dx dy.
\end{aligned}$$

To proceed with the second term, consider the function $f(y)$ defined for $y \in [\omega_1(0), L]$ and given by

$$f(y) = \inf \{ x \in [0, L] : \omega_1(x) \geq y \};$$

it is well-defined because $\omega_1(x) \geq x$ for all x , and in particular $\omega_1(L) = L$, so the set is nonempty. (Notice that if $\omega_1(x)$ is strictly increasing and continuous, then $f(y)$ is its inverse: $f(y) = (\omega_1)^{-1}(y)$.)

Let us show that for all $y \in (\omega_1(0), L)$, $\omega_2(y) \geq \frac{(f(y))^2}{y}$. Suppose first that $f(y) > \omega_2(y)$. Then for $\varepsilon \in (0, f(y) - \omega_2(y))$, $(f(y) - \varepsilon, y) \in \Omega_0$. Thus, $(\omega_1(f(y) - \varepsilon)) \omega_2(y) \geq (f(y) - \varepsilon)^2$. Now, we have $\omega_1(f(y) - \varepsilon) < y$ by definition of $f(y)$, so $\omega_2(y) \geq \frac{(f(y) - \varepsilon)^2}{y}$, and since this is true for arbitrarily small ε , we have $\omega_2(y) \geq \frac{(f(y))^2}{y}$. Second, suppose that $f(y) = \omega_2(y)$. Notice that we must have $f(y) \leq y$; indeed, $f(y) > y$ would imply that for $x \in (y, f(y))$, we would have $\omega_1(x) < y < x$, which contradicts $\omega_1(x) \geq x$. Thus, $1 \geq \frac{f(y)}{y}$, and multiplying this inequality by equality $\omega_2(y) = f(y)$ we get $\omega_2(y) \geq \frac{(f(y))^2}{y}$. Lastly, suppose that $f(y) < \omega_2(y)$. Since $\omega_2(y) \leq y$ (because $\omega_1(x) \geq x$ for all x), we have $f(y) < y$, so $1 > \frac{f(y)}{y}$, and multiplying this inequality by inequality $\omega_2(y) > f(y)$ we get that the required inequality $\omega_2(y) \geq \frac{(f(y))^2}{y}$ holds.

Using this inequality, we rewrite the second term from Lemma C14 as

$$\begin{aligned}
\frac{1}{L^2} \int_0^L \left(\frac{\omega_2(y)}{2} \right)^2 dy &\geq \frac{1}{L^2} \int_{\omega_1(0)}^L \left(\frac{(f(y))^2}{2y} \right)^2 dy \\
&= \frac{1}{L^2} \int_0^L \frac{1}{4y^2} (f(y))^4 dy \\
&= \frac{1}{L^2} \int_0^L \frac{1}{4y^2} \left(\int_0^{f(y)} 4x^3 dx \right) dy \\
&= \frac{1}{L^2} \iint_A \frac{x^3}{y^2} dx dy.
\end{aligned}$$

We therefore have

$$\begin{aligned}
\mathcal{W}(\mathcal{M}) &\geq \frac{2}{15}L - \frac{1}{L^2} \iint_A \left(\frac{x^3}{y^2} - \frac{x^4}{2y^3} \right) dx dy + \frac{1}{L^2} \iint_A \frac{x^3}{y^2} dx dy \\
&= \frac{2}{15}L + \frac{1}{L^2} \iint_A \frac{x^4}{2y^3} dx dy.
\end{aligned}$$

This completes the proof. \square

Proof of Proposition 16. First, notice that for the mechanisms in the Proposition, $\mathcal{W}(\mathcal{M}) = \frac{2}{15}L$; this follows immediately from Lemma C14. Suppose that \mathcal{M}' is an admissible mechanism that does not coincide with either of the mechanisms described in the proposition almost everywhere; then $\omega'_1(a_1) < L$ for a positive measure of a_1 and $\omega'_2(a_2) < L$ for a positive measure of a_2 . By Lemma C13, there is a pseudo-admissible mechanism \mathcal{M}'' such that $\mathcal{W}(\mathcal{M}'') = \mathcal{W}(\mathcal{M}')$, however, by construction, $\omega''_1(a_1) < L$ for a positive measure of a_1 . Now, by Lemma C15, there is a pseudo-admissible mechanism \mathcal{M}''' such that $\omega''_1(a_1)$ and $\omega''_2(a_2)$ are monotone, $\omega'''_1(a_1) < L$ for a positive measure of a_1 , and $\mathcal{W}(\mathcal{M}''') \leq \mathcal{W}(\mathcal{M}'')$. Now, by Lemma C16, $\mathcal{W}(\mathcal{M}''') > \frac{2}{15}L$; therefore, we have $\mathcal{W}(\mathcal{M}') = \mathcal{W}(\mathcal{M}'') \geq \mathcal{W}(\mathcal{M}''') > \frac{2}{15}L$. Thus, $\frac{2}{15}L$ is indeed the minimum of $\mathcal{W}(\mathcal{M})$ for an admissible mechanism, and it is achieved by exactly two mechanisms, up to differences on a set of measure zero. This completes the proof. \square

Proof of Proposition 17. Suppose, for contradiction, that there is a mechanism that achieves full learning using at most k tests. Let a be a particular type such that $k < k(a)$. For each dimension $i = 1, \dots, n$, let q_i be the probability that the sender gets tested on dimension i if his type is a (and he follows the equilibrium strategy). Then $\sum_i q_i$ is the expected number of tests performed when the sender is type a ; hence, $\sum_i q_i \leq k <$

$k(a)$. By definition of $k(a)$, the vector (q_1, \dots, q_n) must violate some constraint of the linear program in the proposition statement: that is, there is some nonempty subset S of coordinates such that $V(a|_S) < \left(1 - \sum_{j \notin S} q_j\right) V(a)$.

Now if the sender's true type is $a|_S$, but he imitates type a , he will be caught only if some dimension $j \notin S$ is tested, which happens with probability at most $\sum_{j \notin S} q_j$. (Note that this is true even though the probability of testing dimension j "off-path" — i.e. once a lie on another dimension has been detected — need not equal q_j . The reason is because, for each $j \notin S$, the probability of j being the *first* dimension outside of S that gets tested must be the same for the deviating $a|_S$ as it would be for a truthful sender of type a , and so is at most q_j .) Therefore, with probability at least $1 - \sum_{j \notin S} q_j$, the lie is not detected. Hence, the sender receives payoff at least $\left(1 - \sum_{j \notin S} q_j\right) V(a) > V(a|_S)$. So the sender would rather deviate, and the mechanism is not valid. This contradiction completes the proof. \square