Theoretical Economics 15 (2020), 923–954 1555-7561/20200923

Costly verification in collective decisions

Albin Erlanson
Department of Economics, University of Essex

Andreas Kleiner
Department of Economics, W. P. Carey School of Business, Arizona State University

We study how a principal should optimally choose between implementing a new policy and maintaining the status quo when information relevant for the decision is privately held by agents. Agents are strategic in revealing their information; the principal cannot use monetary transfers to elicit this information, but can verify an agent's claim at a cost. We characterize the mechanism that maximizes the expected utility of the principal. This mechanism can be implemented as a cardinal voting rule, in which agents can either cast a baseline vote, indicating only whether they are in favor of the new policy, or make specific claims about their type. The principal gives more weight to specific claims and verifies a claim whenever it is decisive.

Keywords. Collective decision, costly verification.

JEL classification. D71, D82.

1. Introduction

The usual mechanism design paradigm assumes that agents have private information and the only way to learn this information is by giving agents incentives to reveal it truthfully. This is a suitable model for many situations, most importantly when agents have private information about their preferences. But there are a number of environments where agents' private information is based on hard facts. This could enable an outside party to learn the private information of the agents, at a potentially significant cost.

For example, consider a chief executive officer (CEO) in a company who faces an investment decision. Board members have relevant information but could have misaligned incentives because the investment has different effects on different divisions. The CEO can take the information provided by a board member at face value or hire consultants to check various claims made by a board member. Another example is large mergers in the European Union, which must be approved by the European Commission.

Albin Erlanson: albin.erlanson@essex.ac.uk
Andreas Kleiner: andreas.kleiner@asu.edu

This paper was previously circulated under the title "Optimal social choice with costly verification." We are grateful to two anonymous referees, Daniel Krähmer, and Benny Moldovanu for detailed comments and discussions. We would also like to thank Hector Chade, Eddie Dekel, Francesc Dilmé, Navin Kartik, Jens Gudmundsson, Bart Lipman, and seminar participants at Bonn University and at various conferences and universities for insightful comments. Albin Erlanson gratefully acknowledges financial support from the European Research Council and the Jan Wallander and Tom Hedelius Foundation.

 https://doi.org/10.3982/TE3101

If a proposed merger has a potentially large impact and its evaluation is not clear, a detailed investigation is initiated. The Commission collects information from the merging companies, third parties, and competitors. According to the Commission, this investigation "typically involves more extensive information gathering, including companies' internal documents, extensive economic data, more detailed questionnaires to market participants, and/or site visits." The analyses carried out by the Commission on potential efficiency gains require that "claimed efficiencies must be verifiable" (European Union 2013). Last, consider the example, taken from Sweden, of the decision of whether a newly approved pharmaceutical drug should be subsidized. A producer of a drug can apply for a subsidy by providing arguments for the clinical and cost-effectiveness of the drug. Other stakeholders are also given an opportunity to participate in the deliberations by contributing information relevant to the decision. Importantly, the applicant and other stakeholders should provide documentation supporting their claims (Pharmaceutical Benefits Board 2019).

So as to study such situations, we formulate a model with costly verification in which a principal decides between introducing a new policy and maintaining the status quo. The principal's optimal choice depends on agents' private information, summarized by each agent i's type $t_i \in \mathbb{R}$. Agents can be in favor of or against the new policy, and they are strategic in revealing their information since it influences the decision made by the principal. We exclude monetary transfers, but before taking the decision, the principal can verify any agent and learn his information at a cost c_i. We determine the mechanism that maximizes the expected payoff of the principal; it optimally solves the trade-off between the benefits from using detailed information as input to the decision rule and the implied costs of verifying agents' claims to make the mechanism incentive compatible.

In the optimal mechanism, agents can vote in favor of or against the new policy; moreover, they have the option to report their exact type. If agent i reports his type, the principal adjusts the reported type by the verification cost c_i to obtain agent i's *net type*, which is $t_i - c_i$ if i votes in favor and $t_i + c_i$ if he votes against (see Figure 1 for an illustration). If an agent does not report his type the principal assumes this agent has a default net type, namely ω_i^+ if he voted in favor of the new policy and ω_i^- if he voted against. This induces *bunching*, since an agent who is in favor reports only his type if it is high enough and otherwise casts only a vote (and conversely if he is against). The optimal decision rule for the principal is then to implement the new policy whenever the sum of net types is positive. A report is *decisive* whenever it changes the decision compared to this agent not sending a report; in the optimal mechanism each decisive report is verified.

Our analysis provides at least two important insights for the design of mechanisms in applications similar to our model. We illustrate them by connecting our analysis to the European Commission's decision on whether to approve a merger. In a merger review, the Commission "analyses claimed efficiencies which the companies could achieve when merged together. If the positive effects of such efficiencies for consumers would outweigh the mergers' negative effects, the merger can be cleared" (European Union 2013). Our analysis suggests, first, that the Commission should not always use claimed efficiencies (which must be verifiable), but might benefit by assuming that a merger has

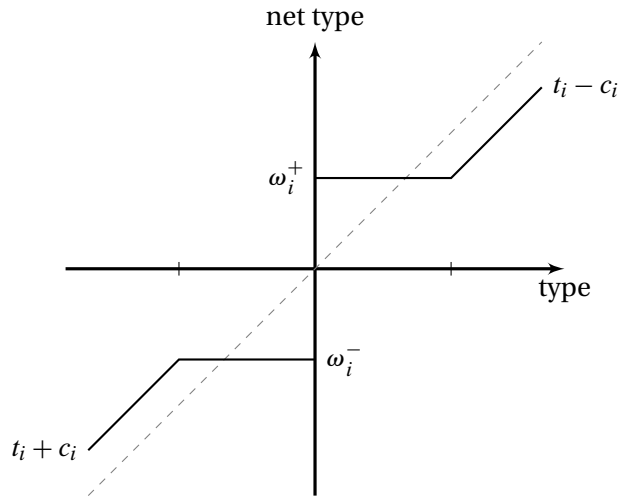

FIGURE 1. Illustration of how types are transformed to net types. The principal implements the new policy whenever the sum of net types is positive.

a predetermined estimated efficiency gain, even if no verifiable documentation is provided. Moreover, it might be beneficial to discount efficiency claims that are difficult and expensive to verify. Second, before starting the process of verifying claimed efficiencies and other reports, the Commission should first determine which reports are decisive and subsequently verify only them. While there are other verification rules that could be used, this is a particularly simple rule that is easy to implement in practice and it provides robust incentives for truth-telling.

To explain the intuition behind the optimal mechanism, we now describe in more detail our main results. We show first that the principal can, without loss of generality, use an incentive compatible direct mechanism, which can be implemented as follows. In the first step, agents communicate their information. For each profile of reports, a mechanism then provides answers to three questions: First, which reports should be verified (verification rule)? Second, what is the decision regarding the new policy (decision rule)? Finally, what is the penalty when someone is revealed to be lying? Because we can focus on incentive compatible mechanisms, penalties will be imposed only off the equilibrium path. The principal can, therefore, always choose the severest possible penalty, as this weakens incentive constraints but does not affect the decision made on the equilibrium path. In general, the principal can implement any decision rule by always verifying all agents. However, the principal has to make a trade-off between using detailed information for "good" decisions and incurring the costs of verification.

Key to solving the principal's problem is that incentive constraints are tractable. Each agent wants to send the report that maximizes the probability that his preferred decision is implemented. We show that if there is a profitable deviation for some type, any type that has a lower equilibrium probability of getting his preferred outcome also finds this deviation profitable. This suggests that incentive constraints are hardest to

satisfy for the types that have the lowest equilibrium probability of getting their preferred decision; we call these types the *worst-off types*.[1] It follows that a mechanism is incentive compatible if and only if it is incentive compatible for the worst-off types.

We can now explain how and why the optimal mechanism differs from the first best outcome. First, in the optimal mechanism the principal incurs costs of verification. Verifications are clearly necessary if information is private and, since the incentive constraints for worst-off types are exactly binding, the optimal mechanism uses costly verifications as rarely as possible. Second, the decision is distorted compared to the first-best because there is bunching at the bottom. This is optimal for the principal because, as observed above, incentive constraints are hardest to satisfy for worst-off types. Suppose instead there was no bunching at the bottom and a single type had the lowest probability of getting the preferred decision. Then any higher report has to be verified sometimes to make the worst-off type indifferent between reporting truthfully and deviating. Now if we increase the probability that the worst-off type gets his preferred outcome, this only changes the decision for this type, which has essentially no effect on the principal's expected utility from the decision. But this makes it less attractive for the worst-off type to claim to be of a different type and the principal can, therefore, verify all other types with a strictly lower probability. Thus, this change allows the principal to save on verification costs for almost all reports, but it changes only the decision for one type. This implies that the cost-saving effect dominates. We conclude that the original mechanism, with a single worst-off type, could not have been optimal and that the optimal mechanism must feature bunching at the bottom. Finally, the principal's first-best decision would be to implement the new policy whenever the sum of types is positive, but in the optimal mechanism, the principal uses net types instead to determine the decision, which introduces a further distortion. Whenever an agent's report t_i is verified, the principal pays the verification cost c_i. If the principal implements the new policy because agent i reported a high type, i's effect on the principal's payoff is only his net value $t_i - c_i$ and not his actual type t_i, because the principal has to pay the verification cost c_i. It is, therefore, optimal for the principal to distort the decision rule by using net types instead of true types.

The remainder of the paper is organized as follows. After reviewing relevant literature, we present in Section 2 our main model and describe the principal's objective. In Section 3, we discuss the optimal mechanism. We consider various extensions in Section 4, including an analysis of the optimal mechanism with imperfect verification. All proofs that are not found in the main body of the paper are relegated to the Appendix.

Related literature

There is a substantial literature on collective choice problems with two alternatives when monetary transfers are not possible. A particular strand of this literature, dating back to the seminal work of Rae (1969), assumes that agents have cardinal utilities and

[1] Since we allow for general utility functions, these are not necessarily the types with the lowest expected utility.

compares decision rules with respect to ex ante expected utilities. Because money cannot be used to elicit cardinal preferences, Pareto-optimal decision rules are simple and can be implemented as voting rules, where agents indicate only whether they are in favor of or against the policy (Schmitz and Tröger 2012, Azreli and Kim 2014). Introducing a technology to learn the agents' information allows a much richer class of decision rules to be implemented. Our main interest lies in understanding how this additional possibility allows for other implementable mechanisms and changes the optimal decision rule.

Townsend (1979) introduces costly verification in a principal–agent model with a risk-averse agent. Our model differs from his and the literature that builds on it (see, e.g., Gale and Hellwig 1985, Border and Sobel 1987), since monetary transfers are not feasible in our model. Allowing for monetary transfers yields different incentive constraints and economic trade-offs than in a model without money.

Recently, there has been growing interest in models with state verification that do not allow for transfers. Ben-Porath et al. (2014, henceforth BDL) consider a principal that wishes to allocate an indivisible good among a group of agents, and each agent's type can be learned at a given cost. The principal's trade-off is between allocating the object efficiently and incurring the cost of verification. BDL characterize the mechanism that maximizes the expected utility of the principal: it is a favored-agent mechanism, where a predetermined favored agent receives the object unless another agent claims a value above a threshold, in which case the agent with the highest (net) type gets the object. We study a similar model of costly verification and without transfers, but we are interested in optimal mechanisms in collective choice problems. In these problems more complex voting mechanisms are feasible, even in the absence of verification possibilities. More recently, Mylovanov and Zapechelnyuk (2017) study the allocation of an indivisible good when the principal always learns the private information of the agents but only after having made the allocation decision and having only limited penalties at his disposal. Halac and Yared (2019) introduce costly verification in a delegation setting and describe the conditions under which interval delegation with an "escape clause" is optimal.

Glazer and Rubinstein (2004) and Glazer and Rubinstein (2006) consider a situation in which an agent has private information about several characteristics and tries to persuade a principal to take a given action, and the principal can only check one of the agent's characteristics. Recently, Ben-Porath et al. (2019) study a class of mechanism design problems with evidence. They show that the optimal mechanism does not use randomization, commitment is not an issue, and robust incentive compatibility does not entail any cost. Additionally, they show that costly verification models can be embedded as evidence games as an alternative way to find optimal mechanisms, but the results on commitment and robustness do not apply to costly verification models.[2]

[2]For additional papers on mechanism design with evidence, see also Green and Laffont (1986), Bull and Watson (2007), Deneckere and Severinov (2008), and Ben-Porath and Lipman (2012).

2. Model and preliminaries

There is a principal and a set of agents $\mathcal{I} = \{1, 2, \dots, I\}$. The principal decides between implementing a new policy and maintaining the status quo.[3] Each agent holds private information, summarized by his type $t_i \in \mathbb{R}$. The payoff to the principal is $\sum_i t_i$ if the new policy is implemented, and it is normalized to 0 if the status quo remains. Monetary transfers are not possible. The private information held by the agents is verifiable. The principal can check agent i at a cost of c_i, in which case he learns the true type of agent i. Being verified imposes no costs on the agent. Agent i with type t_i obtains a utility of $u_i(t_i)$ if the policy is implemented and 0 otherwise. For example, if $u_i(t_i) = t_i$ for each agent, the principal maximizes utilitarian welfare; in general, the principal could have divergent preferences, for example, because he cares only about how the new policy affects him.[4] Types are drawn independently from the type space $T_i \subset \mathbb{R}$ according to the distribution function F_i with finite moments and density f_i. Let $t \equiv (t_i)_{i\in\mathcal{I}}$ and $T \equiv \prod_i T_i$.

The principal can design a mechanism and agents play a Bayesian Nash equilibrium in the game induced by the mechanism. A mechanism could potentially be an indirect and complicated dynamic mechanism that includes multiple rounds of communication and checking. However, we show in Appendix A.1 that it is without loss of generality to focus on direct mechanisms with truth-telling as a Bayesian Nash equilibrium. To allow for stochastic mechanisms we introduce a correlation device as a tool to correlate the decision rule with the verification rules. Assume that s is a random variable that is drawn independently of the types from a uniform distribution on $[0, 1]$ and observed only by the principal. A direct *mechanism* (d, a, ℓ) consists of a *decision rule* $d : T \times [0, 1] \to \{0, 1\}$, a profile of *verification rules* $a \equiv (a_i)_{i\in\mathcal{I}}$, where $a_i : T \times [0, 1] \to \{0, 1\}$, and a profile of *penalty rules* $\ell \equiv (\ell_i)_{i\in\mathcal{I}}$, where $\ell_i : T \times T_i \times [0, 1] \to \{0, 1\}$. In a direct mechanism (d, a, ℓ), each agent sends a message $m_i \in T_i$ to the principal. Given these messages, the principal verifies agent i if $a_i(m, s) = 1$. If no one is found to have lied, the principal implements the new policy if $d(m, s) = 1$.[5] If the verification reveals that agent i has lied, the new policy is implemented if and only if $\ell_i(m, t_i, s) = 1$, where t_i is agent i's true type. If more than one agent lied, it is arbitrary what decision to take. For each agent i, let $T_i^+ := \{t_i \in T_i | u_i(t_i) > 0\}$ denote the set of types that are in favor of the new policy, and let $T_i^- := \{t_i \in T_i | u_i(t_i) < 0\}$ denote the set of types that are against the policy. We assume that $t_i^- < t_i^+$ for all $t_i^- \in T_i^-$ and $t_i^+ \in T_i^+$. This assumption ensures a weak alignment between the agents' and the principal's preferences: if an agent is in favor of the new policy, this increases the principal's expected utility from implementing the policy. This implies that no agent has an incentive to misrepresent his ordinal type, for example, by claiming that he is in favor of the new policy while he actually is against the new policy. To simplify notation, we also assume that no agent is indifferent, so $T_i = T_i^+ \cup T_i^-$.

[3] We discuss in Section 4 how our analysis changes if the principal can decide between more than two actions.

[4] Another interpretation of the objective function, suggested by a referee, is that the principal is interested in the mean of an unknown parameter.

[5] With slight abuse of notation, we drop the realization of the randomization device as an argument whenever the correlation is irrelevant. In these cases, $\mathbb{E}_s[d(m, s)]$ is simply denoted as $d(m)$ and $\mathbb{E}_s[a_i(m, s)]$ is denoted as $a_i(m)$.

EXAMPLE 1. To illustrate a situation in which the general utility function $u_i(t_i)$ is useful, consider a principal deciding whether to provide a public good. Agents are privately informed about their value for the public good, which is always positive. The principal bears the cost k of providing the public good and maximizes the sum of the agents' values minus the potential cost of providing the public good.

This example can be mapped into our model by defining the type of an agent that values the public good at v_i to be $t_i = v_i - k/I$ and by setting $u_i(t_i) > 0$ for all t_i. Clearly, all agents are in favor, even if their types are negative, and the principal's payoff of providing the public good is $\sum t_i = \sum v_i - k$. ◊

Truth-telling is a Bayesian Nash equilibrium for the mechanism (d, a, ℓ) if and only if the mechanism (d, a, ℓ) is Bayesian incentive compatible, which is formally defined as follows.

DEFINITION 1. A mechanism (d, a, ℓ) is *Bayesian incentive compatible* (BIC) if, for all $i \in \mathcal{I}$ and all $t_i, t_i' \in T_i$,

$$\begin{aligned} &u_i(t_i) \cdot \mathbb{E}_{t_{-i},s}\big[d(t_i, t_{-i}, s)\big] \\ &\quad \geq u_i(t_i) \cdot \mathbb{E}_{t_{-i},s}\big[d\big(t_i', t_{-i}, s\big)\big[1 - a_i\big(t_i', t_{-i}, s\big)\big] + a_i\big(t_i', t_{-i}, s\big)\ell_i\big(t_i', t, s\big)\big]. \end{aligned}$$

The left-hand side of the equation in Definition 1 is the interim expected utility if agent i truthfully reports his type t_i and all others also report truthfully. The right-hand side is the interim expected utility if agent i instead lies and reports to be of type t_i'.

The aim of the principal is to find an incentive compatible mechanism that maximizes his expected utility. The expected utility of the principal for a given mechanism (d, a, ℓ) is

$$\mathbb{E}_t\left[\sum_i \big(d(t)t_i - a_i(t)c_i\big)\right],$$

where expectations are taken over the prior distribution of types.

Because the principal uses an incentive compatible mechanism, lies occur only off the equilibrium path and, therefore, do not directly enter the objective function. The principal can, therefore, always choose the severest possible penalty for a lying agent. This does not affect the outcome on the equilibrium path, but it weakens the incentive constraints. For example, if an agent is found to have lied and his true type supports the new policy, the penalty is to maintain the status quo. Henceforth, without loss of optimality, we assume that the principal uses this penalty scheme and we drop the reference to a profile of penalty functions when we describe a mechanism.

At this point, we have all the prerequisites and definitions required to formally state the aim of the principal:

$$\max_{d,a} \mathbb{E}_t\left[\sum_i \big(d(t)t_i - a_i(t)c_i\big)\right] \tag{P}$$

such that (d, a) is Bayesian incentive compatible.

The following lemma provides a characterization of Bayesian incentive compatible mechanisms.

LEMMA 1. *A mechanism* (d, a) *is Bayesian incentive compatible if and only if, for all* $i \in \mathcal{I}$ *and all* $t_i \in T_i$,

$$\inf_{t_i' \in T_i^+} \mathbb{E}_{t_{-i},s}\big[d\big(t_i', t_{-i}, s\big)\big] \geq \mathbb{E}_{t_{-i},s}\big[d(t_i, t_{-i}, s)\big[1 - a_i(t_i, t_{-i}, s)\big]\big],$$

$$\sup_{t_i' \in T_i^-} \mathbb{E}_{t_{-i},s}\big[d\big(t_i', t_{-i}, s\big)\big] \leq \mathbb{E}_{t_{-i},s}\big[d(t_i, t_{-i}, s)\big[1 - a_i(t_i, t_{-i}, s)\big] + a_i(t_i, t_{-i}, s)\big].$$

We call a type a *worst-off type* if the infimum (respectively, the supremum) in Lemma 1 is attained for this type. The intuition for Lemma 1 is as follows. First, because of the binary nature of the principal's decision, an agent maximizes his utility by sending a report that maximizes the probability of getting the preferred decision. Now if type t_i can increase this probability by deviating to a report t_i', any other type can use the same deviation t_i' to get the same probability (since types are distributed independently). By construction, worst-off types have the lowest probability of getting their preferred decision when being truthful. Thus, whenever some type has a profitable deviation, so do the worst-off types.

PROOF OF LEMMA 1. Let $i \in \mathcal{I}$. We consider two cases: one when agent i is in favor of the policy ($t_i' \in T_i^+$) and the other when agent i is against the policy ($t_i' \in T_i^-$).

Since $u_i(t_i) > 0$ for $t_i \in T_i^+$ and we can, without loss of generality, set $\ell_i(t', t_i, s) = 0$ for all t' and $t_i \in T_i^+$, we get that agent i with type $t_i' \in T_i^+$ has no incentive to deviate if and only if, for all $t_i \in T_i$,

$$\mathbb{E}_{t_{-i},s}\big[d\big(t_i', t_{-i}, s\big)\big] \geq \mathbb{E}_{t_{-i},s}\big[d(t_i, t_{-i}, s)\big[1 - a_i(t_i, t_{-i}, s)\big]\big]. \tag{1}$$

Since (1) is required to hold for all $t_i' \in T_i^+$, it must, in particular, hold for the infimum over T_i^+, which is equivalent to Definition 1 of BIC.

Similarly, since $u_i(t_i) < 0$ for $t_i \in T_i^-$ and we can, without loss of generality, set $\ell_i(t', t_i, s) = 1$ for all t' and $t_i \in T_i^-$, a type $t_i' \in T_i^-$, has no incentive to deviate if and only if, for all $t_i \in T_i$,

$$\mathbb{E}_{t_{-i},s}\big[d\big(t_i', t_{-i}, s\big)\big] \leq \mathbb{E}_{t_{-i},s}\big[d(t_i, t_{-i}, s)\big[1 - a_i(t_i, t_{-i}, s)\big] + a_i(t_i, t_{-i}, s)\big]. \tag{2}$$

Since (2) is required to hold for all $t_i' \in T_i^-$, it must, in particular, hold for the supremum over T_i^-, which is equivalent to Definition 1 of BIC. □

3. VOTING WITH EVIDENCE

In this section, we show that a voting-with-evidence mechanism is optimal, find optimal weights in a setting with two agents, and discuss comparative statics.

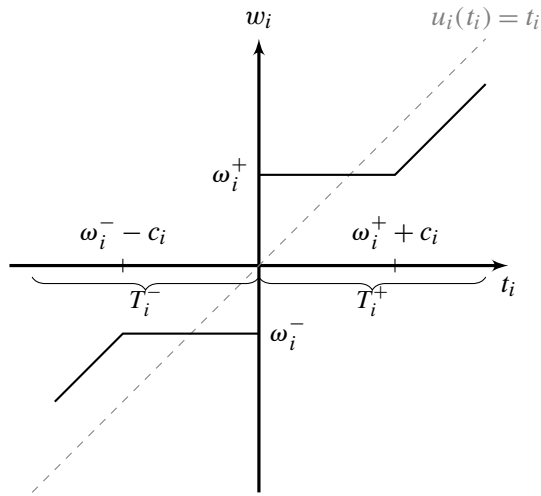

FIGURE 2. Example illustrating how weights are determined with utility $u_i(t_i) = t_i$.

3.1 *Optimal mechanism*

To formally define a voting-with-evidence mechanism, we define, given a collection of weights $\{\omega_i^+, \omega_i^-\}_{i\in\mathcal{I}}$ satisfying $\omega_i^- \leq \omega_i^+$, the *weight function* $w_i : T_i \to \mathbb{R}$ by

$$w_i(t_i) = \begin{cases} t_i + c_i & \text{if } t_i \in T_i^- \text{ and } t_i < \omega_i^- - c_i, \\ \omega_i^- & \text{if } t_i \in T_i^- \text{ and } t_i \geq \omega_i^- - c_i, \\ \omega_i^+ & \text{if } t_i \in T_i^+ \text{ and } t_i \leq \omega_i^+ + c_i, \\ t_i - c_i & \text{if } t_i \in T_i^+ \text{ and } t_i > \omega_i^+ + c_i. \end{cases}$$

For an illustration, see Figure 2. Given the weight functions w_i, we say that a mechanism is a *voting-with-evidence mechanism* if

$$d(t) = \begin{cases} 1 & \text{if } \sum w_i(t_i) > 0, \\ 0 & \text{if } \sum w_i(t_i) < 0, \end{cases}$$

and an agent i is verified if and only if he is decisive. An agent i is *decisive* at a profile of reports t if his preferred outcome is implemented and if the decision were to change if his report was replaced by his relevant cutoff ($\omega_i^+ + c_i$ if he is in favor and $\omega_i^- - c_i$ if he prefers status quo).

A voting-with-evidence mechanism can be interpreted as a cardinal voting rule, where agents have the option to make specific claims to gain additional influence. To see this, consider the following indirect mechanism. Each agent casts a vote either in favor of or against the new policy. In addition, agents can make claims about their information. If agent i does not make such a claim, his vote is weighted by the baseline weights ω_i^+ if he votes in favor of the new policy and by $-\omega_i^-$ if he votes against. If agent i supports the new policy and makes a claim t_i, his weight is increased to $t_i - c_i$. Similarly, if he opposes the new policy, his weight is increased to $-t_i + c_i$. The new policy

is implemented whenever the sum of weighted votes in favor is larger than the sum of the weighted votes against the new policy. An agent's claim is checked whenever he is decisive. This indirect mechanism indeed implements the same outcome as a voting-with-evidence mechanism. Any agents with weak or no information supporting their desired alternative prefer merely to cast a vote, whereas agents with sufficiently strong information make claims to gain additional influence over the outcome of the principal's decision. Note that the cutoffs already determine the default voting rule that is used if all agents cast votes.

A voting-with-evidence mechanism is particularly simple to describe when all agents have type-independent preferences, i.e., for each i, $u_i(t_i) > 0$ or $u_i(t_i) < 0$ for all t_i. For instance, consider the case of deciding on the provision of a public good, where the cost of provision of the public good is borne by the principal (this case is spelled out in detail in Example 1). Therefore, when agent i is always in favor of implementing the project, agent i is assigned a default type of $\omega_i^+ + c_i$, and the principal presumes i has the default type unless i reports differently. The principal reduces the reported (or presumed) type by the verification cost to obtain i's net type, and implements the policy whenever the sum of net types is positive. If an agent changes the outcome because he reports a type different from the default type, he will be verified.

Remark 1 (Ex post incentive compatibility of voting-with-evidence mechanisms). We now show that a voting-with-evidence mechanism is incentive compatible. We do so by showing that for every type, realization truth-telling is a best response. Let $t \in T$ be a profile of types, consider an agent i with type t_i, and assume that agent i is in favor of the new policy, i.e., $t_i \in T_i^+$. If $d(t_i, t_{-i}) = 1$, then agent i gets his preferred alternative and there is no beneficial deviation. Suppose instead that $d(t_i, t_{-i}) = 0$; then agent i can only change the decision by reporting some $t_i' > t_i$ and $t_i' > \omega_i^+ + c_i$. However, if $d(t_i', t_{-i}) = 1$, then agent i is decisive and will be verified. Agent i's true type t_i is revealed and the penalty is the retention of the status quo. Thus, agent i cannot gain by deviating to t_i'. A symmetric argument holds if agent i is against the new policy, i.e., $t_i \in T_i^-$. These arguments imply that truth-telling is an optimal response to truth-telling for every type realization and, therefore, independent of the beliefs the agents hold. We conclude that a voting-with-evidence mechanism is *ex post incentive compatible*.

We are now ready to state our main result.

Theorem 1. *A voting-with-evidence mechanism maximizes the expected utility of the principal.*

Appendix A.2 contains the proof of Theorem 1. We first prove it for finite type spaces and then extend the proof to infinite type spaces through an approximation argument. Before finding optimal weights for a voting-with-evidence mechanism in a two-agent example, we explain intuitively why these mechanisms are optimal.

A voting-with-evidence mechanism differs in three respects from the first-best mechanism. We argue that these inefficiencies have to be present in an optimal mechanism and that any additional inefficiencies make the principal worse off. First, the principal verifies all decisive agents and incurs the corresponding costs, which he would

not need to do if the information were public. Clearly, sometimes verifying agents is necessary to satisfy the incentive constraints for the given decision rule. Moreover, in a voting-with-evidence mechanism, the verification rules are chosen such that the incentive constraints are, in fact, binding: if the principal were to reduce the audit probability for some report, types in the bunching region would have a strict incentive to send this report. Thus, the principal cannot implement the given decision rule with lower verification costs.

The second inefficiency is introduced by replacing types with net types. Specifically, any report $t_i \in T_i^+$ and above $\omega_i^+ + c_i$ is replaced by the net type $t_i - c_i$. Similarly, types $t_i \in T_i^-$ and below $\omega_i^- - c_i$ are replaced by the net type $t_i + c_i$. Suppose we replace types of agent i by net types. Then, for a given profile of types, by replacing agent i's type with his net type, the decision either remains the same or it changes. First, if the decision remains the same, it does not matter whether the type or net type is used. Alternatively, if the decision changes, then agent i must be decisive with type t_i, but not with the net type. Therefore, the principal has to verify the agent if he uses the type t_i to decide on the policy so as to induce truthful reporting and incurs the cost of verification. Hence, the actual contribution of agent i to the principal's utility is his net type, $t_i - c_i$, and not t_i. Thus, the principal is made better off by using i's net type $t_i - c_i$ when determining his decision on the policy, anticipating that he will have to verify the agent whenever he is decisive.

The third inefficiency arises from the fact that all types below the cutoff $\omega_i^+ + c_i$ of an agent in favor of the policy are bunched together and receive the same weight: the baseline weight ω_i^+. Similarly, all types above the cutoff $\omega_i^- - c_i$ and against the policy are bunched together into the baseline weight ω_i^-. Suppose instead that in the optimal mechanism there was a type $t_i' \in T_i^+$ that uniquely had the lowest probability of getting his preferred decision: $\mathbb{E}[d(t_i', t_{-i})] < \mathbb{E}[d(t_i, t_{-i})]$ for all t_i. Increasing the probability with which this type gets his most preferred alternative does not affect the principal's expected utility directly (because this type is realized with probability 0). However, our characterization of incentive compatibility implies that changing this probability affects the audit probability for all other types $t_i \in T_i^+$:

$$\mathbb{E}_{t_{-i}}\big[a(t_i, t_{-i})\big] \geq \mathbb{E}_{t_{-i}}\big[d(t_i, t_{-i})\big] - \mathbb{E}_{t_{-i}}\big[d\big(t_i', t_{-i}\big)\big].$$

Therefore, changing the allocation on a null set allows the principal to save verification costs with strictly positive probability. This contradicts the idea that the original mechanism could be optimal and implies that any optimal mechanism has bunching "at the bottom."[6]

[6]More specifically, assume there is an agent i who is always in favor of the new policy and his type space is $T_i = [0, 1]$, and suppose $\mathbb{E}_{t_{-i}}[d(0, t_{-i})] < \mathbb{E}_{t_{-i}}[d(t_i, t_{-i})]$, so 0 is the only worst-off type. In particular, every report except 0 is sometimes be verified. Consider changing the decision rule so that, for any type $t_i \in [0, \varepsilon]$ and any t_{-i}, the probability of implementing the new policy is $\tilde{d}(t_i, t_{-i}) = \mathbb{E}[d(z, t_{-i})|z \leq \varepsilon]$, and the expected decision is unchanged for all other types of i and all other agents. It then follows from Lemma 1 that for any type above ε, the verification probability can be reduced by $\delta = \mathbb{E}_{t_{-i}}[\tilde{d}(0, t_{-i}) - d(0, t_{-i})] > 0$ and no type of agent i below ε is ever be verified. For ε sufficiently small, the saving in verification costs is on the order of $\delta(1 - \varepsilon)$ and, therefore, it outweighs the inefficiency induced to the decision rule, which is on the order of $\delta\varepsilon$. Hence, it could not have been optimal to have a unique worst-off type.

Remark 2. We comment briefly on the role of the assumption $t_i^- < t_i^+$ for all $t_i^- \in T_i^-$ and $t_i^+ \in T_i^+$. Without this assumption, we get a similar result to Theorem 1 except for the conclusion $\omega_i^+ \geq \omega_i^-$. We then have to check whether agents have an incentive to misreport their ordinal preference in this mechanism. As long as $\omega_i^+ \geq \omega_i^-$, all incentive constraints are satisfied even if the assumption $t_i^- < t_i^+$ is violated. Only if preferences are strongly misaligned, so that an agent being in favor makes the principal less eager to implement the new policy, we have to augment the mechanism by either (i) verifying agents even if they report in the bunching region or (ii) adjusting the weights so that $\omega_i^+ \geq \omega_i^-$ holds.

Remark 3. As noted before, any voting-with-evidence mechanism satisfies the strong notion of ex post incentive compatibility (see Remark 1). This implies that truthful reporting is an equilibrium irrespective of the prior beliefs or the information structure. This robustness of the voting-with-evidence mechanism is a desirable property that seems particularly useful regarding practical implementations.

Because the optimal mechanism is ex post incentive compatible and we allowed for any Bayesian incentive compatible mechanism, we conclude that the principal cannot save verification costs by implementing the mechanism only in Bayesian equilibrium. In the working paper version (Erlanson and Kleiner 2020), we explain this observation by showing that for any Bayesian incentive compatible mechanism in our model, there exists an equivalent ex post incentive compatible mechanism with the same expected verification costs. We see in Section 4.2 that this conclusion depends partly on the details of our model and we explore extensions in which this equivalence breaks down.

3.2 *Optimal weights and comparative statics for two agents*

We begin by characterizing the optimal weights in an utilitarian setting with two agents and then discuss comparative statics.[7]

Proposition 1. *Suppose* $I = 2$, $T_i^+ = \{t_i \in T_i | t_i \geq 0\}$, *and* $T_i^- = \{t_i \in T_i | t_i < 0\}$. *Let* ω_i^+ *and* ω_i^- *be implicitly defined by*

$$\mathbb{E}[t_i | t_i \geq 0] = \mathbb{E}\big[\max\{\omega_i^+, t_i - c_i\} | t_i \geq 0\big],$$
$$\mathbb{E}[t_i | t_i < 0] = \mathbb{E}\big[\min\{\omega_i^-, t_i + c_i\} | t_i < 0\big].$$

Then voting-with-evidence using weights ω_i^+ *and* ω_i^- *is optimal.*

To gain some intuition for the result in Proposition 1, suppose $\omega_1^+ > -\omega_2^-$ and consider slightly changing ω_1^+. This has an effect only if $t_2 + c_2 = -\omega_1^+$, so we condition throughout on this event. If ω_1^+ is slightly increased, then for any $t_1 > 0$, the project is implemented and no one is verified. Alternatively, if ω_1^+ is slightly decreased, there are

[7]With more than two agents, the weight of agent i not only affects the likelihood that i is decisive, but also has nontrivial effects on the probability that other agents are decisive. It is, therefore, more difficult to find closed-form solutions for the optimal weights ω_i^+ and ω_i^-.

two cases: if $t_1 - c_1 + t_2 + c_2 \geq 0$, the project is implemented and agent 1 is verified; otherwise the project is not implemented and agent 2 is verified. We obtain that ω_1^+ satisfies the first-order condition if

$$\int_0^\infty t_1 + t_2 \, dF_1(t_1) = \int_0^\infty (t_1 + t_2 - c_1)\mathbf{1}_{t_1 - c_1 + t_2 + c_2 \geq 0}(t_1) - c_2 \mathbf{1}_{t_1 - c_1 + t_2 + c_2 < 0}(t_2) \, dF_1(t_1).$$

Using $t_2 + c_2 = -\omega_1^+$, this can be rewritten as

$$\int_0^\infty t_1 \, dF_1(t_1) = \int_0^\infty (t_1 - c_1)\mathbf{1}_{t_1 - c_1 \geq \omega_1^+}(t_1) + \omega_1^+ \mathbf{1}_{t_1 - c_1 < \omega_1^+}(t_2) \, dF_1(t_1),$$

which yields the first condition in Proposition 1. An analogous argument heuristically explains the second condition.

Given the characterization of the optimal weights in Proposition 1, we can study how a change in the cost parameter c_i affects the optimal weights. Suppose that the cost of verifying agent i increases. Then the optimal weight ω_i^+ increases to allow for $\mathbb{E}[t_i | t_i \geq 0]$ to equal $\mathbb{E}[\max\{\omega_i^+, t_i - c_i\} | t_i \geq 0]$. Analogously, the increase in c_i implies that ω_i^- decreases. We conclude that as the cost of verifying an agent increases, the bunching region increases and the agent is verified less often. Another possible comparative static result concerns a second-order stochastic dominance change. Suppose the expected value of agent i's type t_i, conditional on him being in favor, increases. Then Proposition 1 implies that his optimal weight ω_i^+ increases as well.

4. Imperfect verification and robustness

In this section, we discuss the robustness of our results from various angles. In the first part, we relax the assumption of perfect verification. In the second part, we discuss briefly type-dependent costs of verification, interdependent preferences, a continuous decision on the level of the public good, and limited commitment.

4.1 *Imperfect verification*

Thus far, we have assumed that the verification technology works perfectly, that is, whenever the principal audits an agent, he learns the true type with probability 1. We now explore the extent to which the above results are robust to imperfect verification. We study a reduced form model and assume that in the event of an audit of agent i, the verification technology reveals the true type of agent i only with probability p, and with probability $1 - p$, the technology fails, in which case the output of the technology equals the report by the agent. Consequently, if the verification output differs from the reported type, the principal knows that the agent lied. However, if the output of the verification technology coincides with the reported type, the principal knows only that the agent was truthful or that the verification technology failed, but not which of these two cases applies. Moreover, we assume that multiple verifications of the same agent reveal no additional information.

To find the optimal mechanism, we first characterize Bayesian incentive compatibility in this new setting with imperfect verification. Similar to the case of perfect verification, the key incentive constraints are those for the worst-off types. The additional uncertainty of whether the verification technology managed to detect a lie implies that the worst-off type must get a higher expected probability of getting the preferred alternative.

LEMMA 2. *A mechanism (d, a) is Bayesian incentive compatible if and only if, for all $i \in \mathcal{I}$ and all $t_i \in T_i$,*

$$\inf_{t_i' \in T_i^+} \mathbb{E}_{t_{-i},s}\big[d\big(t_i', t_{-i}, s\big)\big] \geq \mathbb{E}_{t_{-i},s}\big[d(t_i, t_{-i}, s)\big[1 - p \cdot a_i(t_i, t_{-i}, s)\big]\big],$$

$$\sup_{t_i' \in T_i^-} \mathbb{E}_{t_{-i},s}\big[d\big(t_i', t_{-i}, s\big)\big] \leq \mathbb{E}_{t_{-i},s}\big[d(t_i, t_{-i}, s)\big[1 - p \cdot a_i(t_i, t_{-i}, s)\big] + p \cdot a_i(t_i, t_{-i}, s)\big].$$

The proof is analogous to the proof of Lemma 1.

The imperfectness of the verification technology implies that it is harder to satisfy the incentive constraints. Moreover, there is an upper bound on how much influence an agent can have in expectation. Since $a_i(t, s) \leq 1$ by feasibility and using Lemma 2, we get that any Bayesian incentive compatible mechanism satisfies

$$\forall t_i \in T_i^+: \quad \mathbb{E}_{t_{-i},s}\big[d(t_i, t_{-i}, s)\big] \leq \frac{1}{1-p} \inf_{t_i' \in T_i^+} \mathbb{E}_{t_{-i},s}\big[d\big(t_i', t_{-i}, s\big)\big], \tag{3}$$

$$\forall t_i \in T_i^-: \quad \mathbb{E}_{t_{-i},s}\big[d(t_i, t_{-i}, s)\big] \geq \frac{1}{1-p}\Big[\sup_{t_i' \in T_i^-} \mathbb{E}_{t_{-i},s}\big[d\big(t_i', t_{-i}, s\big)\big] - p\Big]. \tag{4}$$

This adds an additional constraint to the relaxed problem that essentially restricts the maximal influence an agent can have on the decision rule in any incentive compatible mechanism. The higher is the probability of failure $1 - p$ of the verification technology, the tighter is the bound and the less is the influence an agent can have.

THEOREM 2. *With imperfect verification as described above, an optimal mechanism sets $d(t) = 1$ if and only if $\sum_i w_i(t_i) > 0$, where*

$$w_i(t_i) = \begin{cases} \nu_i^- & \text{if } t_i \in T_i^- \text{ and } t_i \leq \nu_i^- - \dfrac{c_i}{p}, \\ t_i + \dfrac{c_i}{p} & \text{if } t_i \in T_i^- \text{ and } \nu_i^- < t_i + \dfrac{c_i}{p} < \omega_i^-, \\ \omega_i^- & \text{if } t_i \in T_i^- \text{ and } t_i \geq \omega_i^- - \dfrac{c_i}{p}, \\ \omega_i^+ & \text{if } t_i \in T_i^+ \text{ and } t_i \leq \omega_i^+ + \dfrac{c_i}{p}, \\ t_i - \dfrac{c_i}{p} & \text{if } t_i \in T_i^+ \text{ and } \nu_i^+ > t_i - \dfrac{c_i}{p} > \omega_i^+, \\ \nu_i^+ & \text{if } t_i \in T_i^+ \text{ and } t_i \geq \nu_i^+ + \dfrac{c_i}{p} \end{cases}$$

for some constants $\{\omega_i^+, \omega_i^-, \nu_i^+, \nu_i^-\}$ satisfying $\omega_i^- \leq \omega_i^+$.

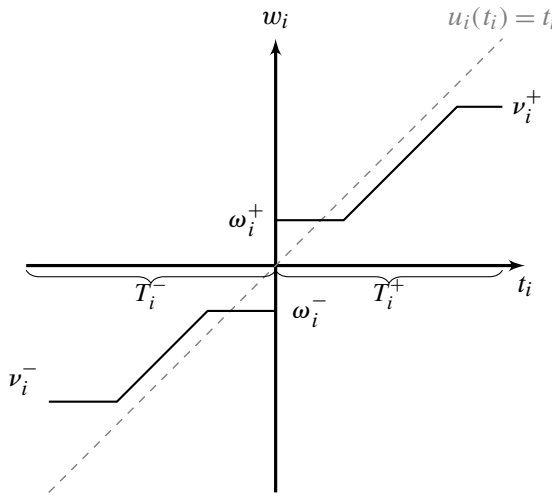

FIGURE 3. Example illustrating weights for imperfect verification with utility $u_i(t_i) = t_i$.

Compared to the optimal mechanism in the benchmark model with perfect verification, the optimal mechanism with imperfect verification can feature additional bunching regions at the extremes (see Figure 3). The reason is that any incentive compatible mechanism must restrict the maximal weight of an agent compared to the worst-off types. If a worst-off type could misreport his type and thereby increase the probability of getting his preferred outcome by too much, this misreport would be a profitable deviation even if this agent was always verified, simply because the verification technology sometimes fails to detect the lie. Therefore, any incentive compatible mechanism must cap the maximal weight an agent could get, inducing bunching for the extreme types.[8]

In contrast to the optimal mechanism with perfect verification, it is not enough to verify only decisive agents. Clearly, one should verify only an agent who gets his preferred outcome, but to induce truth-telling as a Bayes Nash equilibrium, one sometimes needs to verify agents who are not decisive. Suppose only decisive agents were verified. Clearly, agents with worst-off types would have an incentive to overstate their types because this could never hurt them (they are only verified if they are decisive, in which case the penalty is the outcome they would have obtained under truth-telling), and benefits them whenever the verification technology fails to detect their lie. Thus, agents sometimes need to be verified even when they are not decisive; by doing this sufficiently often, we can ensure that the mechanism is BIC. There are several ex post auditing rules that can make the optimal mechanism BIC, and we have not specified exactly which agents are going to be verified for a given realization of reports. We establish the existence of a feasible auditing rule in Lemma 8. This reasoning also implies that the optimal mechanism is not ex post incentive compatible if the verification technology is imperfect.

[8]This is reminiscent of the optimal mechanism in Mylovanov and Zapechelnyuk (2017), who study the optimal allocation of a prize when the winner is subject to a limited penalty if he makes a false claim. In their model, the limit on the penalty similarly requires that agents with the highest possible type are merely short-listed and do not win the prize with certainty.

REMARK 4. The additional bunching regions are the main qualitative difference of the optimal decision rule compared to the model with perfect verification. However, in many settings this difference does not even arise: if an optimal decision rule d (as described in Theorem 2) satisfies, for each i,

$$(1-p)\sup_{t_i\in T_i^+}\mathbb{E}_{t_{-i}}d(t_i,t_{-i})<\inf_{t_i\in T_i^+}\mathbb{E}_{t_{-i}}d(t_i,t_{-i}),$$

$$(1-p)\inf_{t_i\in T_i^-}\mathbb{E}_{t_{-i}}d(t_i,t_{-i})>\sup_{t_i\in T_i^-}\mathbb{E}_{t_{-i}}d(t_i,t_{-i}),$$

then $\nu_i^+=\infty$ and $\nu_i^-=-\infty$ (see the proof of Lemma 7). Therefore, the weight function looks qualitatively similar to that in the case of perfect verification. For example, if $p>\frac{1}{2}$, then the above conditions are always satisfied for a symmetric mechanism (d is symmetric around 0) in a symmetric environment ($f_i(t_i)=f_i(-t_i)$ and $T_i^+=-T_i^-$), because $\inf_{t_i\in T_i^+}\mathbb{E}_{t_{-i}}d(t_i,t_{-i})\geq\frac{1}{2}$ and $\sup_{t_i\in T_i^-}\mathbb{E}_{t_{-i}}d(t_i,t_{-i})\leq\frac{1}{2}$.

4.2 *Robustness*

In the remainder of the text we are going to keep the assumption of perfect verification and change some of our other assumptions to inquire which features of our analysis are robust.

Type-dependent cost function In our benchmark model we assume that the cost of verifying an agent depends only on the agent's identity and not on his true type. Alternatively, one could argue that it is more expensive to audit an agent who claims to have a large type and provides extensive documentation substantiating his claim. Similarly, one could argue that it is easier, and therefore cheaper, to verify an agent who has a low type. Here, we explain how our conclusions are altered if we allow for the audit cost to depend on the true type. Let $c_i(t_i)$ denote the audit cost for verifying agent i if his true type is t_i.

We observe first that the revelation principle still applies, and we can restrict attention to Bayesian incentive compatible direct mechanisms. Also, this change affects only the principal's utility and we can, therefore, use the characterization of Bayesian incentive compatibility as before. On the equilibrium path, the principal will verify only agents who are truthful, so the cost of verification is $c_i(t_i)$. To simplify the discussion, we assume that the net type $t_i-c_i(t_i)$ is increasing in t_i.[9] Using the same arguments as in our benchmark model, we can conclude that the optimal mechanism uses a weighting rule as in a voting-with-evidence mechanism, except that the weight of a report outside of the bunching region is now $t_i-c_i(t_i)$ instead of t_i-c_i (respectively, $t_i+c_i(t_i)$). The part of the weighting function outside the bunching region is, therefore, no longer a straight line with a slope of 1, but a potentially nonlinear increasing function instead. Other than that, the optimal mechanism is like a voting-with-evidence mechanism: the project is implemented if the sum of the weighted reports is positive, and an agent is verified if and only if he is decisive.

[9] If the net type $t_i-c_i(t_i)$ is not increasing, similar arguments can be applied after the types have been reordered.

Choosing the level of the public good In our benchmark model, we assume the principal takes a binary action by deciding whether to implement a public project. We relax this assumption here and analyze a principal who decides on the quantity $d \in [0,1]$ of a public good and assume that the principal pays a cost of $C(d)$ for providing the public good at level d. Assume that the cost function $C : [0,1] \to \mathbb{R}$ is continuously differentiable, increasing, and convex, and satisfies $C'(0) = 0$ and $\lim_{d\to 1} C(d) = +\infty$. All agents have preferences of the form $u(t_i, d) = t_i \cdot d$ and $t_i \in [0,1]$, i.e., agents always prefer more of the public good to less. The objective function of the principal is

$$\mathbb{E}_t\left[\sum_i \big[d(t)t_i - a_i(t)c_i\big] - C\big(d(t)\big)\right] \tag{5}$$

since he incurs the additional costs $C(d)$ of providing the public good.

We begin the discussion with the simplest setting of having only one agent. It follows again from Lemma 1 that any incentive compatible mechanism satisfies $\inf_{t'} d(t') \geq d(t)(1-a(t))$ and since audits are costly, it is optimal to choose the verification rule such that this holds as an equality. Plugging this into the objective function, we get that the principal maximizes $\mathbb{E}_t[d(t)t - (1 - \frac{\inf_{t'} d(t')}{d(t)})c - C(d(t))]$.[10]

The optimal decision rule d must, therefore, satisfy, for almost every t such that $d(t) > \inf_{t'} d(t')$, the first-order condition

$$t - c\frac{\inf_{t'} d\big(t'\big)}{d^2(t)} - C'\big(d(t)\big) = 0.$$

Therefore, as before there is a bunching region, and outside the bunching region we have downward distortions, i.e., too little public good is provided and this distortion is increasing in the verification cost. Note that in contrast to the previous analysis, where the quantity was either 0 or 1, optimal audits are now stochastic (as they satisfy $a(t) = 1 - \frac{\inf_{t'} d(t')}{d(t)}$). Intuition suggests that these conclusions for one agent carry over to the case with multiple agents if we impose ex post incentive compatibility instead of Bayesian incentive compatibility.

Let us now look briefly at the case with several agents and Bayesian incentive constraints. The characterization of Bayesian incentive compatibility in Lemma 1 continues to hold in this setting. Although incentive constraints remain tractable, solving the principal's problem turns out to be less tractable. The principal's optimization problem is to

[10] Observe that for an optimal decision rule, $\inf_{t'} d(t') > 0$. Suppose instead $\inf_{t'} d(t') = 0$. Given $\varepsilon > 0$, let $\delta(\varepsilon) = \text{Prob}(\{t | d(t) \leq \varepsilon\})$. If there is $\varepsilon > 0$ such that $\delta(\varepsilon) < 1$, we can change the decision rule such that $\inf_{t'} d(t') = \varepsilon$ by changing only the decision for types in $\{t | d(t) < \varepsilon\}$. This change increases the cost of public good provision by at most $\delta(\varepsilon)\varepsilon C'(\varepsilon)$, but decreases the cost of verifications by at least $(1 - \delta(\varepsilon))\varepsilon c$. Therefore, for ε small enough, this increases the principal's expected payoff. Alternatively, if $\delta(\varepsilon) = 1$ for all $\varepsilon > 0$, then $d(t) = 0$ for almost every t. Changing the decision rule to $d(t) = \varepsilon$ increases the cost of public good provision by at most $\varepsilon C'(\varepsilon)$ and increases the expected welfare of the principal by $\varepsilon \mathbb{E}[t]$. Since C is continuously differentiable and $C'(0) = 0$, we can, therefore, choose ε small enough such that this change increases the principal's expected welfare. We conclude that in any optimal mechanism $\inf_{t'} d(t') > 0$.

maximize (5) subject to

$$\mathbb{E}_{t_{-i}}\big[d(t_i,t_{-i})\big] \geq \inf_{t_i'} \mathbb{E}_{t_{-i}}\big[d\big(t_i',t_{-i}\big)\big] \geq \mathbb{E}_{t_{-i}}\big[d(t_i,t_{-i})\big(1-a_i(t_i,t_{-i})\big)\big]. \tag{6}$$

Consider first how to construct the optimal audit rule for a given decision rule d. Again, the optimal audit rule satisfies the second inequality in (6) as an equality. To achieve this in the most cost efficient way, we set, for each i and t_i, $a_i(t_i,t_{-i})=1$ for those t_{-i} such that $d(t_i,t_{-i})$ is largest until the second inequality in (6) binds. Thus, the optimal verification rule is deterministic. This is in contrast to the analysis above for the case of one agent. It also implies that there is no simple way to compute the verification costs necessary to implement a given decision rule and we cannot formulate the problem in a simple way with the decision rule being the only choice variable. Because of this, a complete analysis of the optimal decision rule in this case is beyond the scope of our paper.

Interdependent preferences Independent private values allow for a simple characterization of incentive compatibility: a mechanism is Bayesian incentive compatible if and only if it is Bayesian incentive compatible for the worst-off types. This observation does not carry over to models with interdependent preferences. While a complete analysis of this case is beyond the scope of this paper, we discuss below the incentives to misreport in a voting-with-evidence mechanism and possible improvements of this mechanism when preferences are interdependent. To fix ideas, for each $i\in\mathcal{I}$, suppose $T_i=[-1,1]$ and agent i's utility is given by $u_i(t)=t_i+\alpha\sum_{j\neq i}t_j$ if the policy is implemented and the type profile is t, where α satisfies $0<\alpha<1$. For our discussion below, consider a fixed voting-with-evidence mechanism.

If the level of interdependence, α, is high enough, then incentive constraints in the voting-with-evidence mechanism are not binding. Recall that in our benchmark model, reports are verified exactly to make worst-off types indifferent between lying and being truthful. If preferences are sufficiently interdependent, an agent with a small positive type might not want to deviate and send a large report even without verifications: his utility is mainly determined by other agents' types, and claiming a high type might lead to implementation of the new policy even though all others have negative types. This implies that one can reduce the verification probability of large reports without creating any incentives to misreport. However, one cannot lower the verification probability all the way to 0, as otherwise intermediate types have an incentive to send high reports. Which incentive constraints are binding in the optimal mechanism therefore depends on the details of preferences and type distributions. This implies that it is difficult to find the optimal mechanism. But the arguments so far suggest that one way to improve upon a voting-with-evidence mechanism might be to reduce the verification probabilities for high reports, at least if α, the degree of preference interdependence, is sufficiently large.

For moderate degrees of interdependence, α, all types above a threshold will prefer the new policy no matter what the types of all others are since they are only moderately affected by others' types. For these types, incentives are as in our benchmark model since these types will send a report to maximize the probability that the new policy is implemented. Furthermore, for small enough α, this is even true for some types in the bunching region of the voting-with-evidence mechanism. Since the worst-off

types were used to determine the verification probabilities, we cannot reduce the verification probabilities in the voting-with-evidence mechanism at all for small degrees of interdependence.

This suggests that there are only limited ways to improve upon voting-with-evidence mechanisms if α is small. One particularly simple way to improve upon a voting-with-evidence mechanism is to allow agents to abstain in this mechanism. Consider a setting where F_i is symmetric around 0 for each i and adjust the given voting-with-evidence mechanism by allowing for abstention and giving abstentions a weight of 0. Now an agent with a positive type close enough to 0 strictly prefers to abstain instead of casting a vote in favor, which would give weight ω_i^+. This allows for more information being transmitted to the principal without adding verification costs and this mechanism can, therefore, increase the principal's expected utility compared to a voting-with-evidence mechanism.

Limited commitment Following the standard approach in mechanism design, we assume the principal commits to a mechanism. There are several ways in which our optimal mechanism uses commitment of the principal, and our results would change if the principal could not commit. Most importantly, the principal commits to costly verifications although in equilibrium he never finds an agent lying. Second, as explained above, the decision rule is not the first-best for the principal since he distorts the decision by bunching agents and by using net types. This is similar to the use of commitment in standard mechanism design, where principals often commit to ex post inefficient outcomes. Third, in our model the principal commits to penalize an agent who is found to be lying. Note, however, that it is not necessary to use this third component to commit to unreasonable penalties. Suppose in a voting-with-evidence mechanism agent i deviates and reports t_i' although his true type is t_i. This is only relevant if agent i's report changes the outcome to the more preferred one for him, which implies that agent i's report is decisive. In this case, his report is audited and the penalty for agent i is to do the opposite of what agent i prefers. This coincides with the decision if agent i was truthful and reported t_i in the first place because agent i is decisive. In this sense it is not necessary to use commitment to carry out unreasonable penalties.

The fact that commitment matters is typical for models of costly verification, and contrasts with some models of evidence that show that commitment is not necessary (see, e.g., Ben-Porath et al. 2019). One reason for the difference is that, with costly verification, the principal anticipates the verification costs induced by a given decision rule and deviates from the first-best rule to reduce these costs. This effect is not present in models with evidence that have no verification costs.

Appendix

A.1 *Revelation principle*

In this section of the Appendix we show that it is without loss of generality to restrict attention to the class of direct mechanisms as we define them in Section 2. Similar versions of the revelation principle are obtained in Townsend (1988) and Ben-Porath et al. (2014). We proceed in two steps. The first step is a revelation principle argument where

we establish that any indirect mechanism can be implemented via a direct mechanism. In the second step we show that direct mechanisms can be expressed as a tuple (d, a, ℓ), where d specifies the decision, a_i specifies if agent i is verified, and ℓ_i specifies what happens if agent i is revealed to be lying.

Step 1: It is without loss of generality to restrict attention to direct mechanisms in which truth-telling is a Bayes Nash equilibrium. Let $(M_1, \dots, M_I, \tilde{x}, \tilde{y})$ be an indirect mechanism and let $M = \times_{i\in\mathcal{I}} M_i$, where each M_i denotes the message space for agent i, $\tilde{x} : M \times T \times [0,1] \to \{0,1\}$ is the decision function specifying whether the policy is implemented, and $\tilde{y} : M \times T \times \mathcal{I} \times [0,1] \to \{0,1\}$ is the verification function specifying whether an agent is verified.[11] Fix a Bayes Nash equilibrium σ of the game induced by the indirect mechanism.[12]

In the corresponding direct mechanism, let T_i be the message space for agent i. Define $x : T \times T \times [0,1] \to \{0,1\}$ as $x(t', t, s) = \tilde{x}(\sigma(t'), t, s)$ and $y : T \times T \times \mathcal{I} \times [0,1] \to \{0,1\}$ as $y(t', t, i, s) = \tilde{y}(\sigma(t'), t, i, s)$. Since σ is a Bayes Nash equilibrium in the original game, truth-telling is a Bayes Nash equilibrium in the game induced by the direct mechanism. This implies that in both equilibria, the same decision is taken and the same agents are verified.

Note that in any feasible direct mechanism, the decision whether to verify an agent cannot depend on his true type; hence, $y(t_i', t_{-i}, t_i', t_{-i}, i, s) = y(t_i', t_{-i}, t, i, s)$. Also, if agent i was not verified, the implementation decision cannot depend on his true type, $x(t, t, s) = x(t, t_i', t_{-i}, s)$.

Step 2: Any direct mechanism can be written as a tuple (d, a, ℓ)*, where* $d : T \times [0,1] \to \{0,1\}$*,* $a_i : T \times [0,1] \to \{0,1\}$*, and* $\ell_i : T \times T_i \times [0,1] \to \{0,1\}$. Let

$$d(t, s) = x(t, t, s),$$

$$a_i(t, s) = y(t, t, i, s),$$

$$\ell_i\big(t_i', t_{-i}, t_i, s\big) = x\big(t_i', t_{-i}, t_i, t_{-i}, s\big).$$

On the equilibrium path, (d, a, ℓ) implements the same outcome as (x, y) by definition. Suppose instead agent i of type t_i reports t_i' and all other agents report t_{-i} truthfully. Denoting $t' = (t_i', t_{-i})$, the decision taken in the mechanism (d, a, ℓ) if the type profile is t and the report profile is t' is

$$\begin{aligned}&\big[1 - a_i\big(t', s\big)\big]d\big(t', s\big) + a_i\big(t', s\big)\ell_i\big(t_i', t_i, t_{-i}, s\big)\\ &\quad = \big[1 - y\big(t', t', i, s\big)\big]x\big(t', t', s\big) + y\big(t', t', i, s\big)x\big(t', t, s\big)\end{aligned}$$

[11] To describe possibly stochastic mechanisms, we introduce a random variable s that is uniformly distributed on $[0,1]$ and observed only by the principal. This random variable is one way to correlate the verification and the decision on the policy.

[12] In the game induced by the indirect mechanism, whenever the principal verifies agent i, nature draws a type $\tilde{t}_i \in T_i$ as the outcome of the verification. Perfect verification implies that $\tilde{t}_i$ equals the true type of agent i with probability 1. The strategies $m_i \in M_i$ specify an action for each information set where agent i takes an action, even if this information set is never reached with strictly positive probability. In particular, they specify actions for information sets in which the outcome of the verification does not agree with the true type.

$$= \begin{cases} x(t', t, s) & \text{if } y(t', t', i, s) = 1, \\ x(t', t', s) & \text{if } y(t', t', i, s) = 0. \end{cases}$$

If $y(t', t', i, s) = 1$, the decision is $x(t', t, s)$ under both formulations. Instead, if $y(t', t', i, s) = 0$, then $y(t', t, i, s) = 0$ (since the decision to verify agent i cannot depend on his true type), and, hence, the decision on the policy must coincide with the case when agent i is verified and reports t_i', $x(t', t', s) = x(t', t, s)$ We conclude that the decision is the same in both formulations of the mechanism if one agent deviates. Since truth-telling is an equilibrium in the mechanism (x, y), it is, therefore, an equilibrium in the mechanism (d, a, ℓ), which consequently implements the same decision and verification rules.

A.2 *Proof of Theorem 1*

In this section of the Appendix we show that a voting-with-evidence mechanism maximizes the expected utility of the principal. The first step in the proof of Theorem 1 is to construct a relaxed problem for the principal where the optimization is only over decision rules, compared to jointly maximizing decision and verification rules in the original problem. The solution to the relaxed problem always yields weakly higher value than the solution to the original optimization problem (Lemma 3). In the second step, we show that the solution to the relaxed problem is a voting-with-evidence mechanism: first we establish this for finite type spaces (Lemma 4) and then extend the result to infinite type spaces (Lemma 5). To finish the proof we construct verification rules such that the solution to the relaxed problem is feasible for the original problem and achieves the same objective value. This proves Theorem 1.

We show that the problem below is a relaxed version of the principal's maximization problem as defined in (P):

$$\max_{0 \le d \le 1} \mathbb{E}_t \Bigg[\sum_i d(t)\big[t_i - \tilde{c}_i(t_i)\big] + c_i\Big(\mathbb{1}_{T_i^+}(t_i) \inf_{t_i' \in T_i^+} \mathbb{E}_{t_{-i}}\big[d(t_i', t_{-i})\big] - \mathbb{1}_{T_i^-}(t_i) \sup_{t_i' \in T_i^-} \mathbb{E}_{t_{-i}}\big[d(t_i', t_{-i})\big] \Big) \Bigg], \tag{R}$$

where $\mathbb{1}_{T_i^+}(t_i)$ denotes the indicator function for T_i^+, $\mathbb{1}_{T_i^-}(t_i)$ denotes the indicator function for T_i^-, and $\tilde{c}_i(t_i) = c_i$ if $t_i \in T_i^+$ and $\tilde{c}_i(t_i) = -c_i$ if $t_i \in T_i^-$.

For each mechanism (d, a), let $V_P(d, a)$ denote value of the objective in problem (P), and for each decision rule d, let $V_R(d)$ denote the objective value in problem (R).

LEMMA 3. *For any Bayesian incentive compatible mechanism* (d, a), $V_P(d, a) \le V_R(d)$.

PROOF. We have

$$V_P(d, a) = \mathbb{E}_t \Bigg[\sum_i d(t)\big[t_i - \tilde{c}_i(t_i)\big] + c_i \mathbb{1}_{T_i^+}(t_i)\big[d(t) - a_i(t)\big] - c_i \mathbb{1}_{T_i^-}(t_i)\big[d(t) + a_i(t)\big] \Bigg]$$

$$\leq \mathbb{E}_t\Bigg[\sum_i d(t)\big[t_i - \tilde{c}_i(t_i)\big] + c_i \mathbb{1}_{T_i^+}(t_i)\big[d(t)\big(1-a_i(t)\big)\big] - c_i \mathbb{1}_{T_i^-}(t_i)\big[d(t)\big(1-a_i(t)\big)+a_i(t)\big]\Bigg] \tag{7}$$

$$\leq \mathbb{E}_t\Bigg[\sum_i d(t)\big[t_i - \tilde{c}_i(t_i)\big] + c_i \mathbb{1}_{T_i^+}(t_i) \inf_{t_i' \in T_i^+} \mathbb{E}_{t_{-i}}\big[d\big(t_i', t_{-i}\big)\big] - c_i \mathbb{1}_{T_i^-}(t_i) \sup_{t_i' \in T_i^-} \mathbb{E}_{t_{-i}}\big[d\big(t_i', t_{-i}\big)\big]\Bigg] \tag{8}$$

$$= V_R(d).$$

The first inequality holds because $-a_i(t) \leq -d(t)a_i(t)$ and $d(t)a_i(t) \geq 0$. The second inequality follows from the fact that (d, a) is BIC. □

The significance of the relaxed problem lies in the fact that for any optimal solution d to problem (R), we can construct verification rules a such that (d, a) is feasible and $V_P(d, a) = V_R(d)$. This implies that d is part of an optimal solution to problem (P).

We now describe an optimal solution to the relaxed problem for finite type spaces.

Lemma 4. *Suppose that the type space T is finite. Problem (R) is solved by a voting-with-evidence mechanism.*

Proof. Let d^* denote an optimal solution to (R), let $\varphi_i^+ \equiv \inf_{t_i' \in T_i^+} \mathbb{E}_{t_{-i}}[d^*(t_i', t_{-i})]$ and $\varphi_i^- \equiv \sup_{t_i' \in T_i^-} \mathbb{E}_{t_{-i}}[d^*(t_i', t_{-i})]$, and observe that $\varphi_i^- \leq \varphi_i^+$.

Consider the auxiliary maximization problem

$$\max_{0 \leq d \leq 1} \mathbb{E}_t\Bigg[\sum_i d(t)\big[t_i - \tilde{c}_i(t_i)\big]\Bigg] \quad \text{such that for all } i \in \mathcal{I}, \tag{Aux}$$
$$\mathbb{E}_{t_{-i}}[d(t)] \geq \varphi_i^+ \quad \text{for all } t_i \in T_i^+,$$
$$\mathbb{E}_{t_{-i}}[d(t)] \leq \varphi_i^- \quad \text{for all } t_i \in T_i^-,$$

Clearly, d^* also solves the auxiliary problem. The Karush–Kuhn–Tucker theorem (Arrow et al. 1961, Luenberger 1969) implies that there exist Lagrange multipliers $\lambda_i^*(t_i)$ such that $\lambda_i^*(t_i) \geq 0$ for $t_i \in T_i^+$ and $\lambda_i^*(t_i) \leq 0$ for $t_i \in T_i^-$, and such that d^* maximizes

$$\mathcal{L}\big(d, \lambda^*\big) = \mathbb{E}_t\Bigg[\sum_i d(t)\big(t_i - \tilde{c}_i(t_i)\big)\Bigg] + \sum_i \sum_{t_i \in T_i} \big(\lambda_i^*(t_i)\big(\mathbb{E}_{t_{-i}}\big[d(t_i, t_{-i})\big] - \varphi_i(t_i)\big)\big)$$
$$= \sum_{t \in T} d(t) \sum_i \Bigg(t_i - \tilde{c}_i(t_i) + \frac{\lambda_i^*(t_i)}{f_i(t_i)}\Bigg) f(t) + \text{constant},$$

where

$$\varphi_i(t_i) := \begin{cases} \varphi_i^+ & \text{if } t_i \in T_i^+, \\ \varphi_i^- & \text{if } t_i \in T_i^-. \end{cases}$$

Setting $h_i^*(t_i) := t_i - c_i(t_i) + \frac{\lambda_i^*(t_i)}{f_i(t_i)}$ and ignoring the constant in the Lagrangian, we observe that d^* maximizes the function

$$g\big(d, h^*\big) = \sum_{t\in T}\sum_i d(t)f(t)h_i^*(t_i).$$

Let

$$\alpha_i^+ = \inf_{t_i\in T_i^+}\big\{t_i|\mathbb{E}_{t_{-i}}\big[d^*(t_i,t_{-i})\big] > \varphi_i^+\big\} - c_i,$$

$$\alpha_i^- = \sup_{t_i\in T_i^-}\big\{t_i|\mathbb{E}_{t_{-i}}\big[d^*(t_i,t_{-i})\big] < \varphi_i^-\big\} + c_i,$$

and define

$$\bar{h}_i(t_i) := \begin{cases} \dfrac{1}{\mu_i\big(A_i^+\big)}\displaystyle\sum_{t_i\in A_i^+} f_i(t_i)h_i^*(t_i) & \text{if } t_i \in T_i^+ \text{ and } t_i \le \alpha_i^+ + c_i, \\ \dfrac{1}{\mu_i\big(A_i^-\big)}\displaystyle\sum_{t_i\in A_i^-} f_i(t_i)h_i^*(t_i) & \text{if } t_i \in T_i^- \text{ and } t_i \ge \alpha_i^- - c_i, \\ t_i - \tilde{c}_i(t_i) & \text{otherwise,} \end{cases}$$

where $A_i^+ = \{t_i \in T_i^+ | t_i < \alpha_i^+ + c_i\}$, $A_i^- = \{t_i \in T_i^- | t_i > \alpha_i^- - c_i\}$, and $\mu_i(A)$ denotes the measure induced by F_i. Let $A_i^c = T_i \setminus (A_i^+ \cup A_i^-)$ and $A_i = A_i^+ \cup A_i^-$.

Claim 1. *The solution d^* also maximizes $g(d,\bar{h}) = \sum_{t\in T}\sum_i d(t)f(t)\bar{h}_i(t_i)$.*

Step 1: $\lambda^*(t_i) = 0$ for $t_i \in A_i^c$. Complementary slackness implies $\lambda_i^*(\alpha_i^+ + c_i) = 0$. Moreover, for every $t_i \in T_i^+$ such that $t_i > \alpha_i^+ + c_i$, we get $t_i - c_i + \frac{\lambda_i^*(t_i)}{f_i(t_i)} \ge \alpha_i^+$ and, hence, for every optimal solution to the Lagrangian d, that $\mathbb{E}_{t_{-i}}[d(t_i,t_{-i})] \ge \mathbb{E}_{t_{-i}}[d(\alpha_i^+ + c_i, t_{-i})] > \varphi_i^+$. This implies that for $t_i \in T_i^+ \cap A_i^c$, $\lambda_i^*(t_i) = 0$ by complementary slackness. Analogous arguments for $t_i \in T_i^- \cap A_i^c$ apply. Thus, $\lambda^*(t_i) = 0$ for $t_i \in A_i^c$.

Step 2: $g(d^*,h^*) = g(d^*,\bar{h})$. First observe that $h_i^*(t_i) = \bar{h}_i(t_i)$ for $t_i \in A_i^c$, $\varphi_i^+ = \mathbb{E}_{t_{-i}}[d^*(t_i,t_{-i})]$ for $t_i \in A_i^+$, and $\varphi_i^- = \mathbb{E}_{t_{-i}}[d^*(t_i,t_{-i})]$ for $t_i \in A_i^-$. This implies

$$\begin{aligned} g\big(d^*,h^*\big) &= \sum_i\Big[\sum_{t_i\in A_i} h_i^*(t_i)f_i(t_i)\mathbb{E}_{t_{-i}}\big[d^*(t)\big] + \sum_{t_i\in A_i^c} h_i^*(t_i)f_i(t_i)\mathbb{E}_{t_{-i}}\big[d^*(t)\big]\Big] \\ &= \sum_i\Big[\sum_{t_i\in A_i^+} h_i^*(t_i)f_i(t_i)\varphi_i^+ + \sum_{t_i\in A_i^-} h_i^*(t_i)f_i(t_i)\varphi_i^- + \sum_{t_i\in A_i^c} \bar{h}_i(t_i)f_i(t_i)\mathbb{E}_{t_{-i}}\big[d^*(t)\big]\Big] \\ &= \sum_i\Big[\sum_{t_i\in A_i^+} \bar{h}_i(t_i)f_i(t_i)\varphi_i^+ + \sum_{t_i\in A_i^-} \bar{h}_i(t_i)f_i(t_i)\varphi_i^- + \sum_{t_i\in A_i^c} \bar{h}_i(t_i)f_i(t_i)\mathbb{E}_{t_{-i}}\big[d^*(t)\big]\Big] \end{aligned}$$

$$= \sum_i \Bigg[\sum_{t_i \in A_i} \bar{h}_i(t_i) f_i(t_i) \mathbb{E}_{t_{-i}}\big[d^*(t)\big] + \sum_{t_i \in A_i^c} \bar{h}_i(t_i) f_i(t_i) \mathbb{E}_{t_{-i}}\big[d^*(t)\big] \Bigg]$$

$$= g\big(d^*, \bar{h}\big).$$

Step 3: $g(d^*, \bar{h}) = g(d^*, h^*) = \max_{0 \le d \le 1} g(d, h^*) \ge \max_{0 \le d \le 1} g(d, \bar{h})$. The first equality follows from Step 2 and the second holds because d^* maximizes $g(d, h^*)$ by construction.

Let $h_i : T_i \to \mathbb{R}$ be any real-valued function and for each such function h_i, define $H_i(t_i) := h_i(t_i) f_i(t_i)$ and denote $H_i \equiv (H_i(t_i))_{t_i \in T_i}$. Fix an agent $i \in \mathcal{I}$ and define a function $\Psi : \mathbb{R}^{|T_i|} \to \mathbb{R}$ as $\Psi(H_i) := \max_{0 \le d \le 1} \sum_{t \in T} d(t)[f_{-i}(t_{-i}) H_i(t_i) + \sum_{j \in \mathcal{I}_{-i}} f(t) h_j^*(t_j)]$. The function Ψ is convex, since it is a maximum over linear functions. It is also symmetric, since permuting the vector H_i does not change the value of Ψ. Thus, Ψ is Schur-convex. By construction, H_i^* (defined as $H_i^*(t_i) = h_i^*(t_i) f_i(t_i)$) majorizes $\bar{H}_i$ (defined as $\bar{H}_i(t_i) = \bar{h}_i(t_i) f_i(t_i)$). Therefore, we obtain that

$$\Psi\big(H_i^*\big) \ge \Psi(\bar{H}_i).$$

We have now shown that if we replace h_i^* for agent i with its average $\bar{h}_i$, we have that d^* remains the maximizer of $\max_{0 \le d \le 1} g(d, h^*_{\mathcal{I}_{-i}} h_i)$. By repeating this argument agent by agent we can conclude that

$$\max_{0 \le d \le 1} g\big(d, h^*\big) = \max_{0 \le d \le 1} \sum_{t \in T} \sum_{i \in \mathcal{I}} d(t) f_{-i}(t_{-i}) H_i^*(t_i)$$

$$\ge \max_{0 \le d \le 1} \sum_{t \in T} \sum_{i \in \mathcal{I}} d(t) f_{-i}(t_{-i}) \bar{H}_i(t_i) = \max_{0 \le d \le 1} g(d, \bar{h}).$$

This proves the Claim 1.

Hence, every solution to the Lagrangian can be described as

$$d(t) = \begin{cases} 1 & \text{if } \sum w_i(t_i) > 0, \\ 0 & \text{if } \sum w_i(t_i) < 0, \end{cases}$$

where

$$w_i(t_i) = \begin{cases} \omega_i^+ & \text{if } t_i \in T_i^+ \text{ and } t_i \le \alpha_i^+ + c_i, \\ \omega_i^- & \text{if } t_i \in T_i^- \text{ and } t_i \ge \alpha_i^- - c_i, \\ t_i - c_i(t_i) & \text{otherwise} \end{cases} \tag{9}$$

for constants $\{\omega_i^+, \omega_i^-\}_{i \in I}$. Since d^* maximizes the Lagrangian by assumption, we conclude that it takes this form.

Note that $\omega_i^+ \ge \sup_{t_i \in A_i^+} \{t_i - c_i\}$ since $\lambda_i^*(t_i) \ge 0$ for $t_i \in A_i^+$. Also, $\omega_i^+ \le \alpha_i^+$, since otherwise we would get, for $t_i \in A_i^+$, $\mathbb{E}_{t_{-i}}[d^*(t_i, t_{-i})] \ge \mathbb{E}_{t_{-i}}[d^*(\alpha_i^+ - c_i, t_{-i})] > \varphi_i^+$, contradicting the definition of A_i^+. Analogous arguments imply $\inf_{t_i \in A_i^-} \{t_i + c_i\} \le \omega^- \le \alpha_i^-$. This implies that we can replace α_i^+ (α_i^-) with ω_i^+ (ω_i^-) in the definition of the weight function w_i in (9) above without changing the outcome of the mechanism in any way. □

As the next step in the proof, we show that voting-with-evidence mechanisms are also optimal for an infinite type space.

LEMMA 5. *Suppose that T is an infinite type space. Problem (R) is solved by a voting-with-evidence mechanism.*

PROOF. Let F_i^+ and F_i^- denote the conditional distributions induced by F_i on T_i^+ and T_i^-, respectively. We first construct a discrete approximation of the type space. For $i \in \mathcal{I}$, $n \geq 1$, $l_i = 1, \ldots, 2^{n+1}$, let

$$S_i(n, l_i) := \begin{cases} \left\{ t_i \in T_i^+ \,\middle|\, \frac{l_i - 1}{2^n} \leq F_i^+(t_i) < \frac{l_i}{2^n} \right\} & \text{for } l_i \leq 2^n, \\ \left\{ t_i \in T_i^- \,\middle|\, \frac{l_i - 2^n - 1}{2^n} \leq F_i^-(t_i) < \frac{l_i - 2^n}{2^n} \right\} & \text{for } l_i > 2^n, \end{cases}$$

which form partitions of T_i^+ and T_i^-, and denote by $\mathcal{F}_i^n$ the set consisting of all possible unions of the $S_i(n, l_i)$. Let $l = (l_1, \ldots, l_n)$ and $S(n, l) = \prod_{i \in \mathcal{I}} S_i(n, l_i)$, which defines a partition of T, and denote by $\mathcal{F}^n$ the induced σ-algebra.

Let (R^n) denote the relaxed problem with the additional restriction that d is measurable with respect to $\mathcal{F}^n$. Then the constraint set has a nonempty interior and an optimal solution to (R^n) exists. Define $\tilde{t}_i(t_i) := \frac{1}{\mu_i(S_i(n,l_i))} \int_{S_i(n,l_i)} s \, dF_i$ for $t_i \in S_i(n, l_i)$, where μ_i denotes the measure induced by F_i. The arguments for finite type spaces imply that the following rule is an optimal solution to (R^n) for some $\omega_i^{+,n}$, $\omega_i^{-,n}$:

$$r_i^n(t_i) = \begin{cases} \omega_i^{+,n} - c_i & \text{if } t_i \in T_i^+ \text{ and } \tilde{t}_i(t_i) \leq \omega_i^{+,n}, \\ \omega_i^{-,n} + c_i & \text{if } t_i \in T_i^- \text{ and } \tilde{t}_i(t_i) \geq \omega_i^{-,n}, \\ \tilde{t}_i(t_i) - c_i(t_i) & \text{otherwise,} \end{cases}$$

$$d^n(t) = \begin{cases} 1 & \text{if } \sum r_i^n(t_i) > 0, \\ 0 & \text{if } \sum r_i^n(t_i) < 0. \end{cases}$$

Let $\omega_i^+ := \lim_{n \to \infty} \omega_i^{+n}$ and $\omega_i^- := \lim_{n \to \infty} \omega_i^{-,n}$ (by potentially choosing a convergent subsequence). Define

$$r_i(t_i) = \begin{cases} \omega_i^+ - c_i & \text{if } t_i \in T_i^+ \text{ and } \tilde{t}_i(t_i) \leq \omega_i^{+,n}, \\ \omega_i^- + c_i & \text{if } t_i \in T_i^- \text{ and } \tilde{t}_i(t_i) \geq \omega_i^{-,n}, \\ t_i - \tilde{c}_i(t_i) & \text{otherwise,} \end{cases}$$

$$d(t) = \begin{cases} 1 & \text{if } \sum r_i(t_i) > 0, \\ 0 & \text{if } \sum r_i(t_i) < 0. \end{cases}$$

Then, for all i and t_i, $\mathbb{E}_{t_{-i}}[d^n(t_i, t_{-i})] = \text{Prob}[\sum_{j \neq i} r_j^n(t_j) \geq -r_i^n(t_i)]$ converges pointwise almost everywhere to $\mathbb{E}_{t_{-i}}[d(t_i, t_{-i})]$. This implies that the marginals converge in L^1-norm and, hence, the objective value of d^n converges to the objective value of d. This

implies that d is an optimal solution to (R), since if there was a solution that achieves a strictly higher objective value, there would exist $\mathcal{F}^n$-measurable solutions that achieve a strictly higher objective value for all n large enough. Therefore, a voting-with-evidence mechanism solves problem (R). □

Now we have all the parts required to establish our main result, Theorem 1, that voting-with-evidence mechanisms are optimal.

Proof of Theorem 1. Denote by d^* the solution to problem (R). We first construct a verification rule a^* such that (d^*, a^*) is Bayesian incentive compatible and then argue that $V_P(d^*, a^*) = V_R(d^*)$. Given that $V_P(d, a) \leq V_R(d)$ holds for any incentive compatible mechanism, this implies that (d^*, a^*) solves (P).

Let a^* be such that agent i is verified whenever he is decisive. Then $a_i^*(t) = a_i^*(t)d^*(t)$ for all $t_i \in T_i^+$ (if $d^*(t) = 0$, then type $t_i \in T_i^+$ is not decisive) and $d^*(t) = d^*(t)[1 - a_i^*(t)]$ for all $t_i \in T_i^-$ (if $a_i^*(t) = 1$, then $d^*(t) = 0$). Hence, inequality (7) holds as an equality for (d^*, a^*).

Note that in mechanism (d^*, a^*), all incentive constraints are binding and, therefore, inequality (8) holds as an equality as well. We therefore conclude $V_P(d^*, a^*) = V_R(d^*)$. □

Proof of Proposition 1. Without loss of generality, suppose $\omega_1^+ \leq -\omega_2^-$ and consider changing ω_2^- (the other cases are analogous). This matters only if agent 2 has a negative type and agent 1 has a positive type. We consider two cases: (a) a change to $\omega_2^{-'}$ such that $\omega_1^+ \leq -\omega_2^{-'}$; (b) a change such that $\omega_1^+ > -\omega_2^{-'}$.

Case (a). Using weight $\omega_2^{-'}$ such that $-\omega_1^+ \geq \omega_2^{-'} > \omega_2^-$ instead of ω_2^- matters only if agent 1's type satisfies $\omega_2^- \leq -t_1 + c_1 \leq \omega_2^{-'}$. Conditional on such a type, the expected utility of the principal from using weight ω_2^- is 0. Alternatively, using weight $\omega_2^{-'}$ gives conditional expected utility of

$$\int_{-\infty}^{0} (t_1 + t_2 - c_1)\mathbf{1}_{t_1 - c_1 + t_2 + c_2 \geq 0} - c_2 \mathbf{1}_{t_1 - c_1 + t_2 + c_2 < 0}\, dF_2.$$

The definition of ω_2^- implies

$$\begin{aligned}\int_{-\infty}^{0} t_2\, dF_2 &= \int_{-\infty}^{0} \min\{\omega_2^-, t_2 + c_2\}\, dF_2 \\ &\leq \int_{-\infty}^{0} \min\{-t_1 + c_1, t_2 + c_2\}\, dF_2 \\ &= \int_{-\infty}^{0} (-t_1 + c_1)\mathbf{1}_{t_1 - c_1 + t_2 + c_2 \geq 0} + (t_2 + c_2)\mathbf{1}_{t_1 - c_1 + t_2 + c_2 < 0}\, dF_2.\end{aligned}$$

Subtracting $\int_{-\infty}^{0} t_2\, dF_2$ from both sides and multiplying by -1, this implies

$$0 \geq \int_{-\infty}^{0} (t_1 + t_2 - c_1)\mathbf{1}_{t_1 - c_1 + t_2 + c_2 \geq 0} - c_2 \mathbf{1}_{t_1 - c_1 + t_2 + c_2 < 0}\, dF_2$$

and, hence, the principal is better off using weight ω_2^-. Similar arguments also show that the principal is worse off using a cutoff $\omega_2^{-'} < \omega_2^-$.

Case (b). We can think of this case in two steps. First, a change such that $\hat{\omega}_2^- = -\omega_1^+$. As shown in Case (a), this reduces the principal's welfare. Second, a further change to $\omega_2^{-'}$, which changes only the decision if both agents cast a vote. The effect of this second change is nonpositive if and only if

$$\begin{aligned} 0 \geq \int_0^{\omega_1^+ + c_1} \int_{-\omega_1^- - c_2}^{0} & t_1 + t_2 \, dF_2 \, dF_1 \\ & + c_1\big[1 - F_1\big(\omega_1^+ + c_1\big)\big]\big[F_2(0) - F_2\big(-\omega_1^- - c_2\big)\big] \\ & - c_2\big[F_1\big(\omega_1^+ + c_1\big) - F_1(0)\big]F_2\big(-\omega_1^- - c_2\big). \end{aligned}$$

This is equivalent to

$$\begin{aligned} 0 \geq \big\{\mathbb{E}\big[t_1 | 0 \leq t_1 \leq \omega_1^+ + c_1\big] + \mathbb{E}\big[t_2 | 0 \geq t_2 \geq -\omega_1^+ - c_2\big] \\ \times\big\}\big[F_1\big(\omega_1^+ + c_1\big) - F_1(0)\big]\big[F_2(0) - F_2\big(-\omega_1^+ - c_2\big)\big] \\ + c_1\big[1 - F_1\big(\omega_1^+ + c_1\big)\big]\big[F_2(0) - F_2\big(-\omega_1^- - c_2\big)\big] \\ - c_2\big[F_1\big(\omega_1^+ + c_1\big) - F_1(0)\big]F_2\big(-\omega_1^- - c_2\big) \end{aligned}$$

or to

$$\begin{aligned} 0 \geq \mathbb{E}\big[t_1 | 0 \leq t_1 \leq \omega_1^+ + c_1\big] + \mathbb{E}\big[t_2 | 0 \geq t_2 \geq -\omega_1^+ - c_2\big] \\ + c_1 \frac{1 - F_1(0)}{F_1\big(\omega_1^+ + c_1\big) - F_1(0)} - c_1 - c_2 \frac{F_2(0)}{F_2(0) - F_2\big(-\omega_1^+ - c_2\big)} + c_2. \end{aligned} \tag{10}$$

However, the definition of ω_1^+ implies

$$\begin{aligned} \int_0^\infty t_1 \, dF_1 &= \int_{\omega_1^+ + c_1}^\infty t_1 - c_1 dF_1 + \big[F\big(\omega_1^+ + c_1\big) - F(0)\big]\omega_1^+ \\ \Leftrightarrow \quad \omega_1^+ &= \mathbb{E}\big[t_1 | 0 \leq t_1 \leq \omega_1^+ + c_1\big] - c_1 + c_1 \frac{1 - F_1(0)}{F_1\big(\omega_1^+ + c_1\big) - F_1(0)}. \end{aligned} \tag{11}$$

Similarly, the definition of ω_2^- and the fact that $\omega_2^- \leq -\omega_1^+$ imply

$$\begin{aligned} \mathbb{E}[t_2 | t_2 < 0] &= \mathbb{E}\big[\min\big\{\omega_2^- - c_2, t_2\big\} + c_2 | t_2 < 0\big] \\ &\leq \mathbb{E}\big[\min\big\{-\omega_1^+ - c_2, t_2\big\} + c_2 | t_2 < 0\big]. \end{aligned}$$

Rearranging this inequality yields

$$\mathbb{E}\big[t_2 | -\omega_1^+ - c_2 \leq t_2 < 0\big] - c_2 \frac{F_2(0)}{F_2(0) - F_2\big(\omega_1^+ - c_2\big)} \leq -\omega_1^+ - c_2. \tag{12}$$

Plugging (11) and (12) into (10), we see that (10) holds. We conclude that the principal is better off using weight ω_2^-. □

A.3 *Proof of Theorem 2*

Consider the relaxed problem

$$\max_{0\le d\le 1} \mathbb{E}_t\Bigg[\sum_i d(t)\Bigg[t_i - \frac{\tilde{c}_i(t_i)}{p} + \frac{c_i}{p}\Big(\mathbb{1}_{T_i^+}(t_i)\inf_{t_i'\in T_i^+}\mathbb{E}_{t_{-i}}\big[d\big(t_i', t_{-i}\big)\big] - \mathbb{1}_{T_i^-}(t_i)\sup_{t_i'\in T_i^-}\mathbb{E}_{t_{-i}}\big[d\big(t_i', t_{-i}\big)\big]\Big)\Bigg]\Bigg] \tag{$\tilde{\text{R}}$}$$

subject to (3) and (4).

For any mechanism (d,a), let $V_{\tilde{P}}(d,a)$ denote the expected utility of the principal given mechanism (d,a) and let $V_{\tilde{R}}(d)$ denote the value achieved by the decision rule d in the relaxed problem.

LEMMA 6. *For any mechanism (d,a) that is Bayesian incentive compatible in the imperfect verification setting, $V_{\tilde{P}}(d,a) \le V_{\tilde{R}}(d)$.*

PROOF. Note that Lemma 2 implies that

$$\forall t_i \in T_i^+: \quad \mathbb{E}_{t_{-i},s}\big[a_i(t_i,t_{-i},s)d(t_i,t_{-i},s)\big] \ge \frac{1}{p}\Big[\mathbb{E}_{t_{-i},s}\big[d(t_i,t_{-i})\big] - \inf_{t_i\in T_i^+}\mathbb{E}_{t_{-i},s}\big[d\big(t_i',t_{-i},s\big)\big]\Big],$$

$$\forall t_i \in T_i^-: \quad \mathbb{E}_{t_{-i},s}\big[a_i(t_i,t_{-i},s)\big[1-d(t_i,t_{-i},s)\big]\big] \ge \frac{1}{p}\Big[\sup_{t_i\in T_i^+}\mathbb{E}_{t_{-i},s}\big[d\big(t_i',t_{-i},s\big)\big] - \mathbb{E}_{t_{-i},s}\big[d(t_i,t_{-i})\big]\Big].$$

Hence,

$$\begin{aligned} V_{\tilde{P}}(d,a) &= \mathbb{E}_t\Big[\sum_i d(t)t_i - a_i(t)c_i\Big] \\ &\le \mathbb{E}_t\Big[\sum_i d(t)t_i - \mathbb{1}_{T_i^+}(t_i)d(t)a_i(t)c_i - \mathbb{1}_{T_i^-}(t_i)\big[1-d(t)\big]a_i(t)c_i\Big] \qquad (13)\\ &\le \sum_i \mathbb{E}_{t_i}\Big[\mathbb{E}_{t_{-i}}\big[d(t)\big]t_i - \mathbb{1}_{T_i^+}(t_i)\frac{1}{p}\Big[\mathbb{E}_{t_{-i},s}\big[d(t_i,t_{-i})\big] - \inf_{t_i\in T_i^+}\mathbb{E}_{t_{-i},s}\big[d\big(t_i',t_{-i},s\big)\big]\Big]c_i \\ &\quad - \mathbb{1}_{T_i^-}(t_i)\frac{1}{p}\Big[\sup_{t_i\in T_i^+}\mathbb{E}_{t_{-i},s}\big[d\big(t_i',t_{-i},s\big)\big] - \mathbb{E}_{t_{-i},s}\big[d(t_i,t_{-i})\big]\Big]c_i\Big] \qquad (14)\\ &= V_{\tilde{R}}(d). \end{aligned}$$

□

LEMMA 7. *Suppose T is finite. The decision rule stated in Theorem 2 solves problem ($\tilde{\text{R}}$).*

PROOF. Let d^* denote an optimal solution to the relaxed problem ($\tilde{\text{R}}$) above, and define $\varphi_i^+ \equiv \inf_{t_i'\in T_i^+}\mathbb{E}_{t_{-i}}[d^*(t_i',t_{-i})]$ and $\varphi_i^- \equiv \sup_{t_i'\in T_i^-}\mathbb{E}_{t_{-i}}[d^*(t_i',t_{-i})]$. Then d^* also solves the

problem

$$\max_{0\le d\le 1} \mathbb{E}_t\left[\sum_i d(t)\left[t_i - \frac{\tilde{c}_i(t_i)}{p}\right]\right]$$

$$\text{such that for all } i \in \mathcal{I}, \tag{Aux}$$

$$\varphi_i^+ \le \mathbb{E}_{t_{-i}} d(t) \le \frac{\varphi_i^+}{1-p} \quad \text{for all } t_i \in T_i^+,$$

$$\frac{\varphi_i^- - p}{1-p} \le \mathbb{E}_{t_{-i}} d(t) \le \varphi_i^- \quad \text{for all } t_i \in T_i^-.$$

The Karush–Kuhn–Tucker theorem implies that there exist Lagrange multipliers $\lambda_i(t_i)$ and $\mu_i(t_i)$ such that d^* maximizes the Lagrangian:

$$\begin{aligned}\mathcal{L}(d,\lambda,\mu) = {} & \mathbb{E}_t\left[\sum_i d(t)\left(t_i - \frac{\tilde{c}_i(t_i)}{p}\right)\right] \\ & + \sum_i \sum_{t_i \in T_i^+} \left(\lambda_i(t_i)\left(\mathbb{E}_{t_{-i}}\left[d(t_i,t_{-i})\right] - \varphi_i^+\right) + \mu_i(t_i)\left(\frac{\varphi_i^+}{1-p} - \mathbb{E}_{t_{-i}}\left[d(t_i,t_{-i})\right]\right)\right) \\ & + \sum_i \sum_{t_i \in T_i^-} \left(\lambda_i(t_i)\left(\mathbb{E}_{t_{-i}}\left[d(t_i,t_{-i})\right] - \varphi_i^-\right) + \mu_i(t_i)\left(\frac{\varphi_i^- - p}{1-p} - \mathbb{E}_{t_{-i}}\left[d(t_i,t_{-i})\right]\right)\right).\end{aligned}$$

Define $h_i(t_i) := t_i - \frac{\tilde{c}_i(t_i)}{p} + \frac{\lambda_i(t_i)+\mu_i(t_i)}{f_i(t_i)}$ and let

$$\alpha_i^+ = \inf_{t_i \in T_i^+}\left\{t_i \middle| \mathbb{E}_{t_{-i}}\left[d^*(t_i,t_{-i})\right] > \varphi_i^+\right\}, \qquad \alpha_i^- = \sup_{t_i \in T_i^-}\left\{t_i \middle| \mathbb{E}_{t_{-i}}\left[d^*(t_i,t_{-i})\right] < \varphi_i^-\right\},$$

$$\beta_i^+ = \sup_{t_i \in T_i^+}\left\{t_i \middle| \mathbb{E}_{t_{-i}}\left[d^*(t_i,t_{-i})\right] < \frac{\varphi_i^+}{1-p}\right\}, \qquad \beta_i^- = \inf_{t_i \in T_i^-}\left\{t_i \middle| \mathbb{E}_{t_{-i}}\left[d^*(t_i,t_{-i})\right] > \frac{\varphi_i^- - p}{1-p}\right\}.$$

Define $A_i^+ = \{t_i \in T_i^+ | t_i < \alpha_i^+\}$, $A_i^- = \{t_i \in T_i^- | t_i > \alpha_i^-\}$, $B_i^+ = \{t_i \in T_i^+ | t_i > \beta_i^+\}$, $B_i^- = \{t_i \in T_i^- | t_i < \beta_i^-\}$, and

$$\bar{h}_i(t_i) := \begin{cases} \dfrac{1}{\mu_i(A_i^+)} \displaystyle\sum_{t_i \in A_i^+} f_i(t_i) h_i(t_i) & \text{if } t_i \in A_i^+, \\ \dfrac{1}{\mu_i(B_i^+)} \displaystyle\sum_{t_i \in B_i^+} f_i(t_i) h_i(t_i) & \text{if } t_i \in B_i^+, \\ \dfrac{1}{\mu_i(A_i^-)} \displaystyle\sum_{t_i \in A_i^-} f_i(t_i) h_i(t_i) & \text{if } t_i \in A_i^-, \\ \dfrac{1}{\mu_i(B_i^-)} \displaystyle\sum_{t_i \in B_i^-} f_i(t_i) h_i(t_i) & \text{if } t_i \in B_i^-, \\ t_i - \tilde{c}_i(t_i) & \text{otherwise.} \end{cases}$$

The same arguments as in the proof of Lemma 4 imply that d^* maximizes

$$\sum_i \sum_t f(t)d(t)\bar{h}_i(t_i).$$

□

LEMMA 8. *Suppose T is infinite. The decision rule stated in Theorem 2 solves problem ($\tilde{R}$).*

The proof is analogous to the proof of Lemma 5 and, hence, is omitted.

PROOF OF THEOREM 2. Denote by d^* the solution to problem ($\tilde{R}$). For each i, define $q_i : T_i \to [0,1]$ as the solution to

$$\mathbb{E}_{t_{-i}}\big[d^*(t_i,t_{-i})\big[1-p\cdot q_i(t_i)\big]\big] = \inf_{t_i'\in T_i^+} \mathbb{E}_{t_{-i}}\big[d^*\big(t_i',t_{-i}\big)\big] \quad \text{for } t_i \in T_i^+,$$

$$\mathbb{E}_{t_{-i}}\Big[d^*(t_i,t_{-i})\big[1-p\cdot q_i(t_i)\big] = \sup_{t_i'\in T_i^-} \mathbb{E}_{t_{-i}}\big[d^*\big(t_i',t_{-i}\big)\big] - p\cdot q_i(t_i)\Big] \quad \text{for } t_i \in T_i^-.$$

We now show that a solution q_i exists. For $t_i \in T_i^+$, setting $q_i(t_i)=0$ yields

$$\mathbb{E}_{t_{-i}}\big[d^*(t_i,t_{-i})\big[1-pq_i(t_i)\big]\big] = \mathbb{E}_{t_{-i}}\big[d^*(t_i,t_{-i})\big] \geq \inf_{t_i'\in T_i^+} \mathbb{E}_{t_{-i}}\big[d^*\big(t_i',t_{-i}\big)\big]$$

and setting $q_i(t_i)=1$ yields

$$\mathbb{E}_{t_{-i}}\big[d^*(t_i,t_{-i})\big[1-pq_i(t_i)\big]\big] = \mathbb{E}_{t_{-i}}\big[d^*(t_i,t_{-i})[1-p]\big] \leq \inf_{t_i'\in T_i^+} \mathbb{E}_{t_{-i}}\big[d^*\big(t_i',t_{-i}\big)\big],$$

where the inequality follows from (3). The intermediate-value theorem hence implies the existence of a solution q_i. Similar arguments apply for $t_i \in T_i^-$.

Define

$$a_i^*(t) := \begin{cases} q_i(t_i) & \text{if } t_i \in T_i^+ \text{ and } d^*(t)=1, \\ q_i(t_i) & \text{if } t_i \in T_i^- \text{ and } d^*(t)=0, \\ 0 & \text{else.} \end{cases}$$

For each i and for all $t_i \in T_i^+$,

$$\inf_{t_i'\in T_i^+} \mathbb{E}_{t_{-i},s}\big[d^*\big(t_i',t_{-i},s\big)\big] = \mathbb{E}_{t_{-i},s}\big[d^*(t_i,t_{-i},s)\big[1-p\cdot a_i^*(t_i,t_{-i},s)\big]\big],$$

and for all $t_i \in T_i^-$,

$$\sup_{t_i'\in T_i^-} \mathbb{E}_{t_{-i},s}\big[d^*\big(t_i',t_{-i},s\big)\big] = \mathbb{E}_{t_{-i},s}\big[d^*(t_i,t_{-i},s)\big[1-p\cdot a_i^*(t_i,t_{-i},s)\big] + p\cdot a_i^*(t_i,t_{-i},s)\big].$$

Hence, (d^*,a^*) is Bayesian incentive compatible by Lemma 2 and inequality (14) holds as an equality. By construction, $t_i \in T_i^+$ implies $d(t)a_i^*(t) = a_i^*(t)$ and $t_i \in T_i^-$ implies $[1-d(t)]a_i^*(t) = a_i^*(t)$. Therefore, inequality (13) also holds as an equality and we conclude that $V_{\tilde{P}}(d^*,a^*) = V_{\tilde{R}}(d^*)$. Hence, (d^*,a^*) is optimal. □

References

Arrow, Kenneth J., Leonid Hurwicz, and Hirofumi Uzawa (1961), "Constraint qualifications in maximization problems." *Naval Research Logistics*, 8, 175–191. [944]

Azreli, Yaron and Semin Kim (2014), "Pareto efficiency and weighted majority rules." *International Economic Review*, 55, 1067–1088. [927]

Ben-Porath, Elchanan, Eddie Dekel, and Barton L. Lipman (2014), "Optimal allocation with costly verification." *American Economic Review*, 104, 3779–3813. [927, 941]

Ben-Porath, Elchanan, Eddie Dekel, and Barton L. Lipman (2019), "Mechanisms with evidence: Commitment and robustness." *Econometrica*, 87, 529–566. [927, 941]

Ben-Porath, Elchanan and Barton L. Lipman (2012), "Implementation with partial probability." *Journal of Economic Theory*, 147, 1689–1724. [927]

Border, Kim C. and Joel Sobel (1987), "Samurai accountant: A theory of auditing and plunder." *Review of Economic Studies*, 54, 525–540. [927]

Bull, Jesse and Joel Watson (2007), "Hard evidence and mechanism design." *Games and Economic Behavior*, 58, 75–93. [927]

Deneckere, Raymond and Sergei Severinov (2008), "Mechanism design with partial state verifiability." *Games and Economic Behavior*, 64, 487–513. [927]

Erlanson, Albin and Andreas Kleiner (2020), "Costly verification in collective decisions." Unpublished paper. Available at arXiv:1910.13979. [934]

European Union (2013), "Factsheet on merger control procedures.", Retrieved on March 30, 2019 at http://ec.europa.eu/competition/mergers/procedures_en.html. [924]

Gale, Douglas and Martin Hellwig (1985), "Incentive-compatible debt contracts: The one-period problem." *Review of Economic Studies*, 52, 647–663. [927]

Glazer, Jacob and Ariel Rubinstein (2004), "On optimal rules of persuasion." *Econometrica*, 72, 1715–1736. [927]

Glazer, Jacob and Ariel Rubinstein (2006), "A study in the pragmatics of persuasion: A game theoretical approach." *Theoretical Economics*, 1, 395–410. [927]

Green, Jerry R. and Jean-Jacques Laffont (1986), "Partially verifiable information and mechanism design." *Review of Economic Studies*, 53, 447–456. [927]

Halac, Marina and Pierre Yared (2019), "Commitment vs. flexibility with costly verification." Unpublished paper, Columbia University and NBER. [927]

Luenberger, David G. (1969), *Optimization by Vector Space Methods*. Wiley, New York. [944]

Mylovanov, Tymofiy and Andriy Zapechelnyuk (2017), "Optimal allocation with ex post verification and limited penalties." *American Economic Review*, 107, 2666–2694. [927, 937]

Pharmaceutical Benefits Board (2019), "Information about subsidizing pharmaceutical drugs in Sweden.", Retrieved on April 17, 2019 at https://www.tlv.se/in-english/medicines/apply-for-reimbursement/the-decision-process.html. [924]

Rae, Douglas W. (1969), "Decision-rules and individual values in constitutional choice." *American Political Science Review*, 63, 40–56. [926]

Schmitz, Patrick W. and Thomas Tröger (2012), "The (sub-)optimality of the majority rule." *Games and Economic Behavior*, 74, 651–665. [927]

Townsend, Robert M. (1979), "Optimal contracts and competitive markets with costly state verification." *Journal of Economic Theory*, 21, 265–293. [927]

Townsend, Robert M. (1988), "Information constrained insurance: The revelation principle extended." *Journal of Monetary Economics*, 21, 411–450. [941]

Co-editor Simon Board handled this manuscript.

Manuscript received 13 December, 2017; final version accepted 3 November, 2019; available online 19 November, 2019.