Costly Verification in Collective Decisions*

Albin Erlanson Andreas Kleiner

11th December 2017

First draft November 2015

Abstract

We study how a principal should optimally choose between implementing a new policy and maintaining the status quo when the information relevant for the decision is privately held by agents. Agents are strategic in revealing their information, but the principal can verify an agent's information at a given cost. We exclude monetary transfers. When is it worthwhile for the principal to incur the cost and learn an agent's information? We characterize the mechanism that maximizes the expected utility of the principal. The evidence is verified whenever it is decisive for the principal's decision. This mechanism can be implemented as a weighted majority voting rule, where agents are given additional weight if they provide evidence for their information. The evidence is verified whenever it is decisive for the principal's decision. Additionally, we find a general equivalence between Bayesian and ex-post incentive compatible mechanisms in this setting. Finally, we extend our analysis of optimal mechanisms to imperfect verification.

Keywords: Collective decision; Costly verification

JEL classification: D82, D71

*This paper was previously circulated under the title "Optimal Social Choice with Costly Verification". We are grateful to 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. Erlanson: Department of Economics at Stockholm School of Economics, albin.erlanson@hhs.se; Kleiner: Department of Economics, W.P. Carey School of Business at Arizona State University, andreas.kleiner@asu.edu.

1 Introduction

A principal decides whether to implement a new policy. Information that is relevant for the decision is privately held by strategic agents, but we assume that the principal can learn an agent's information at a cost. This is a suitable model for situations in which private information is not only about the preferences of individuals but instead depends on hard evidence, which can be reviewed by the principal. The possibility to learn private information gives the principal an additional tool when designing decision rules and the principal has to decide when he considers it worthwhile to learn an agent's information and how to design an optimal decision rule.

One example, taken from Sweden, that corresponds to our model is the decision of whether a newly approved pharmaceutical drug should be subsidized. This is determined by the Dental and Pharmaceutical Benefits Board (TLV). The producer of the 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 TLV's decision. Importantly, the applicant and other stakeholders should provide documentation supporting their claims made to the board. Clinical effectiveness is documented by reporting the results of clinical trials, while evidence of cost-effectiveness should be provided through analysis derived from a health economic model. The TLV can verify the information provided, but it is costly to do so. For example, the TLV occasionally has to build its own health-economic models or hire external experts to evaluate the evidence that was provided, which entails significant costs. When should the TLV invest effort and money to verify the evidence? What decision rule should the TLV use to decide on the subsidy?

The usual mechanism design paradigm cannot be applied to address these questions because it assumes that information is not verifiable. Instead, 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. 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 deciding, the principal can learn the information of each agent at a given cost. We show that the optimal mechanism can be implemented as a weighted majority voting rule, where agents are given additional weight if they provide evidence supporting their position on the new policy. The evidence is verified whenever it is decisive for the principal's decision. Moreover, we show that for any decision rule there exists an equivalent decision rule that can be robustly implemented without requiring additional verification.

Our analysis provides three insight for design of mechanisms in applications similar to our model. First, only types with strong evidence in support of their preferred alternative should be asked to provide evidence, and types with weak evidence should be bunched together. This reduces the incentives for types with weak evidence to mimic types with

stronger evidence and thereby saves on verification costs since types with stronger evidence can be verified less frequently. Second, evidence should not always be verified. Instead, the principal should determine which agents are decisive and verify only those agents. Third, the principal should take the verification cost into account when evaluating an agent's information.

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, and it 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 will 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 have a tractable structure. A mechanism is incentive compatible if and only if it is incentive compatible for the "worst-off" types. These are the types that have the lowest probability of getting their preferred alternative. If there is a profitable deviation for some type, this deviation will also be profitable for the worst-off types because they have the lowest probability of getting their preferred alternative on the equilibrium path. Because only incentive constraints for the worst-off types matter and additional verification is costly, the optimal verification rule makes the worst-off types exactly indifferent between reporting truthfully and lying. This is true independent of what the optimal decision rule is.

The optimal mechanism can be implemented as a voting rule with flexible weights. Each agent votes in favor of or against the new policy. The decision rule compares the sum of weighted votes in favor of with the sum of weighted votes against the new policy, and the alternative with the highest sum is chosen. Agents that do not provide evidence have baseline weights attached to their votes. If an agent claims to have evidence strongly supporting his preferred alternative, he gains additional weight in the voting rule that corresponds to the importance of his information. We say that an agent provides *decisive* evidence if the policy decision would change if the agent simply voted for his preferred alternative, instead of providing the evidence. In the optimal mechanism, all decisive evidence is verified. Consequently, in equilibrium, agents with weak evidence in favor of their preferred alternative will merely cast a vote, and only agents with strong evidence in favor of their preferred alternative will provide the evidence to the principal.

In the optimal mechanism, an agent is verified whenever he presents decisive evidence. This implies that he cannot gain by deviating, no matter what the others' types are: whenever a deviation affects the outcome, the principal will detect the deviation and

punish the deviator. The strategies we describe therefore form an ex-post equilibrium, which does not depend on the beliefs of the agents. This is a desirable feature of any mechanism because it implies that it can be robustly implemented and does not rely on detailed information about the beliefs of the agents. This robustness is not only a property of the optimal mechanism. Indeed, the principal can obtain robustness of any Bayesian incentive compatible mechanism without incurring additional verification costs; for any Bayesian incentive compatible mechanism there exists an equivalent mechanism that induces the same interim expected decision and verification rules and for which truth-telling is an ex-post equilibrium.

In the last part of the paper, we relax the assumption of perfect verification and consider a situation with imperfect verification. We assume that with some probability the verification technology does not work. The optimal mechanism can still be described as a voting rule with flexible weights. Similar to before, agents get a constant weight if their types are small in absolute terms. However, if the imperfectness of the verification technology is severe enough, there also has to be an upper bound on how much weight an agent can get to maintain incentive compatibility. Otherwise, an agent who is almost sure that he will not get his preferred outcome would use an extreme deviation hoping that his deviation will not be detected.

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 compares decision rules with respect to ex-ante expected utilities. Because money cannot be used to elicit cardinal preferences, Pareto-optimal decision rules are very 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).[1] 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 building 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, Dekel and Lipman (2014, henceforth BDL) consider a principal that wishes to allocate an indivisible good among a group of agents, and

[1] See also Gershkov, Moldovanu and Shi (2016) for a recent extension to settings with more than two alternatives.

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 pre-determined 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 (2017) 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, Dekel and Lipman (2017) 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.[2]

Our result on the equivalence between Bayesian and ex-post incentive compatible mechanisms relates our work to several papers that establish an equivalence between Bayesian and dominant-strategy incentive compatible mechanisms in settings with transfers (Manelli and Vincent 2010, Gershkov, Goeree, Kushnir, Moldovanu and Shi 2013). Since incentive constraints take a different form in our model, the economic mechanisms underlying our equivalence are also different. To prove the equivalence, we use mathematical tools due to Gutmann, Kemperman, Reeds and Shepp (1991) that were introduced into the mechanism design literature by Gershkov et al. (2013).

The remainder of the paper is organized as follows. In Section 2, we present the model and describe the principal's objective. In Section 3, we introduce voting-with-evidence mechanisms and discuss their optimality. Appendix A.2 contains the proof of the optimality of the voting-with-evidence mechanisms. We establish an equivalence of Bayesian and ex-post incentive compatible mechanisms in Section 4. In Section 5, we consider a case with imperfect verification and identify the optimal mechanism in this setting. Section 6 concludes the paper. All proofs not found in the main body of the paper are relegated to the Appendix.

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

2 Model and Preliminaries

There is a principal and a set of agents $\mathcal{I} = \{1, 2, \ldots, I\}$. The principal decides between implementing a new policy and maintaining the status quo. Each agent holds private information, summarized by his type t_i. The payoff to the principal is $\sum_i t_i$ if the new policy is implemented, and it is normalized to zero 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 zero 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 only cares about how the new policy affects himself. 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 $t \equiv (t_i)_{i \in N}$ and $T \equiv \prod_i T_i$.

The principal can design a mechanism, and truth-telling by the agents should be 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 only observed 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 $d(m, s) = 1$.[3] If the verification reveals that at least one agent has lied, the principal considers the lie by the agent with the lowest index, call it agent j^*, and implements the new policy if and only if $\ell_{j*}(m, t_{j*}, s) = 1$, where t_{j*} is agent j^*'s true type.

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 $t_i^+ \in T_i^+$.[4] To simplify notation, we also assume that $T_i = T_i^+ \cup T_i^-$.

Truth-telling is a Bayesian Nash equilibrium for the mechanism (d, a, ℓ) if and only

[3] With slight abuse of notation, we will 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 $a_i(m)$.

[4] This will imply 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 is actually against the new policy.

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 $t_i, t_i' \in T_i$,

$$u_i(t_i') \cdot \mathbb{E}_{t_{-i},s}[d(t_i', t_{-i}, s)] \geqslant u_i(t_i') \cdot \mathbb{E}_{t_{-i},s}[d(t_i, t_{-i}, s)[1 - a_i(t_i, t_{-i}, s)] + a_i(t_i, t_{-i}, s)\ell_i(t_i, t_{-i}, t_i', s)].$$

The left-hand side 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

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

where expectations are taken over the prior distribution of types.

Because the principal uses an incentive compatible mechanism, lies will occur only off the equilibrium path and will therefore not directly enter the objective function. The principal can therefore always choose the severest possible penalty for a lying agent. This will 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 will be to maintain the status quo. Henceforth, without loss of optimality, we assume that the principal uses this penalty scheme and, we will 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:

$$\max_{d,a} \mathbb{E}_t\Big[\sum_i (d(t)t_i - a_i(t)c_i)\Big] \tag{P}$$
$$\text{s.t. } (d, a) \text{ being Bayesian incentive compatible.}$$

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 $t_i \in T_i$,*

$$\inf_{t_i' \in T_i^+} \mathbb{E}_{t_{-i},s}[d(t_i', t_{-i}, s)] \geqslant \mathbb{E}_{t_{-i},s}[d(t_i, t_{-i}, s)[1 - a_i(t_i, t_{-i}, s)]] \qquad \textit{and}$$
$$\sup_{t_i' \in T_i^-} \mathbb{E}_{t_{-i},s}[d(t_i', t_{-i}, s)] \leqslant \mathbb{E}_{t_{-i},s}[d(t_i, t_{-i}, s)[1 - a_i(t_i, t_{-i}, s)] + a_i(t_i, t_{-i}, s)].$$

Proof. Let $i \in \mathcal{I}$. We will consider two cases, one when agent i is in favor of the policy ($t_i' \in T_i^+$), and the other case is 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 $t_i \in T_i$,

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

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 wlog 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 $t_i \in T_i$,

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

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

$$w_i(t_i) = \begin{cases} \omega_i^+ & \text{if } 0 \leqslant t_i \leqslant \omega_i^+ + c_i \\ \omega_i^- & \text{if } 0 > t_i \geqslant \omega_i^- - c_i \\ t_i - c_i & \text{if } t_i > \omega_i^+ + c_i \\ t_i + c_i & \text{if } t_i < \omega_i^- - c_i. \end{cases}$$

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 were 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 weighted majority voting rule, where agents have the additional 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

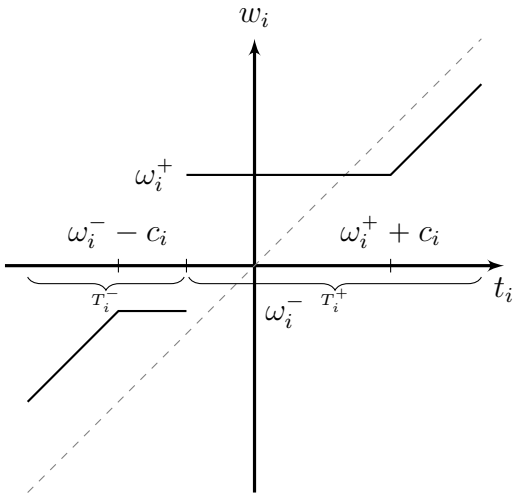


Figure 1: Example illustrating how weights are determined.

by the baseline weights ω_i^+ and $-\omega_i^-$ if he votes in favor of or against the new policy, respectively. 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 are larger than the sum of the weighted votes against the new policy. An agent's claim will be 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 will prefer to merely cast a vote, whereas agents with sufficiently strong information will 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.

Remark 1 (Ex-post incentive compatibility of voting-with-evidence mechanisms)**.** We will now show that a voting-with-evidence mechanism is incentive compatible. We will 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 will be 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 independently of the beliefs the agents hold. We conclude that a voting-with-evidence mechanism is *ex-post incentive compatible* (see Section 4 for more on ex-post incentive compatibility).

We are now ready to state our main result.

Theorem 1. *Voting-with-evidence 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 the proof is extended to infinite type spaces through an approximation argument. Before illustrating a voting-with-evidence mechanism in a two-agent example, we will give intuition for why these mechanisms are optimal.

A voting-with-evidence mechanism differs in three respects from the first-best mechanism. We will argue that these inefficiencies have to be present in an optimal mechanism and that any additional inefficiencies will 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, verifying decisive agents is necessary to satisfy the incentive constraints. Moreover, in a voting-with-evidence mechanism, the verification rules are chosen such that the incentive constraints are in fact binding. 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 $t_i + c_i$. The reason that this is part of an optimal mechanism has to do with decisiveness and when the policy decision changes. If it is the case that by replacing t_i with the net type $t_i - c_i$ the outcome changes, then agent i must be decisive if his altered report were t_i. However, then the principal has to verify him to induce truthful reporting and incurs the cost of verification. Therefore, the actual contribution of agent i to the principal's utility is his net type, $t_i - c_i$, 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.

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 were a unique worst-off type. Increasing the probability with which this type gets his most preferred alternative has no negative effect (because it is realized with probability 0), but this allows the principal to verify all other types (which are realized with probability 1) with a strictly lower probability. Therefore, bunching of types that become the worst-off types must be part of any optimal mechanism.

To summarize, there is an optimal mechanism that bunches types in favor of the new policy (and types against the policy) with weak information supporting their position, and that uses net types instead of true types when determining the decision; these are distinctive features of a voting-with-evidence mechanism.

We end this section by illustrating a voting-with-evidence mechanism in an example with two agents and showing how to determine the optimal baseline weights in this example. We assume that both agents prefer the new policy to the status quo, independent

of their types. The voting-with-evidence mechanism is illustrated in Figure 2a. For report profiles above the solid line, the sum of the altered reports is positive. Thus, in this region, the policy will be implemented. If instead report profiles are below the solid line, the status quo remains.

If the reported types induce the status quo, no agent makes a decisive claim. The same is true if both agents report a very high type, as a claim is not decisive when the claim reported by the other agent already induces the principal to implement the new policy. Both agents are decisive if both report intermediate types that induce the policy, but if any of them were to replace their reported type with the baseline report, the policy would not be implemented.

To determine the optimal baseline weights, we use a first-order approach.[5] Consider a slight increase in the baseline weight of agent 1. This matters only if this changes the decision given agent 2's type t_2'; that is, this is only relevant if $\omega_1^+ + t_2' - c_2 = 0$. Therefore, suppose that agent 2's type is t_2' and that agent 1's type is below $\omega_1^+ + c_1$. If baseline weight ω_1^+ is used, the policy will not be implemented.[6] However, if the cutoff is slightly increased, then the new policy will be implemented, agent 2 becomes decisive, and therefore, agent 2 has to be verified. Hence, the principal's expected utility changes by

$$f_2(t_2') \int_{-\infty}^{\omega_1^+ + c_1} t_1 + t_2' - c_2 \; dF_1.$$

Since the new policy will be implemented at type profile $(\omega_1^+ + c_1, t_2')$ under the higher baseline weight, agent 1 is not decisive at profiles (t_1, t_2') for $t_1 > \omega_1^+ + c_1$ (for these profiles he would be decisive if the smaller baseline weights were used). Consequently, the principal can save on verification costs, which increases his utility by

$$f_2(t_2') \int_{\omega_1^+ + c_1}^{\infty} c_1 \; dF_1.$$

At the optimal baseline weights, these two effects sum to zero. Exploiting that $t_2' - c_2 = -\omega_1^+$, this yields the following first-order condition for the optimal baseline weight for agent 1:

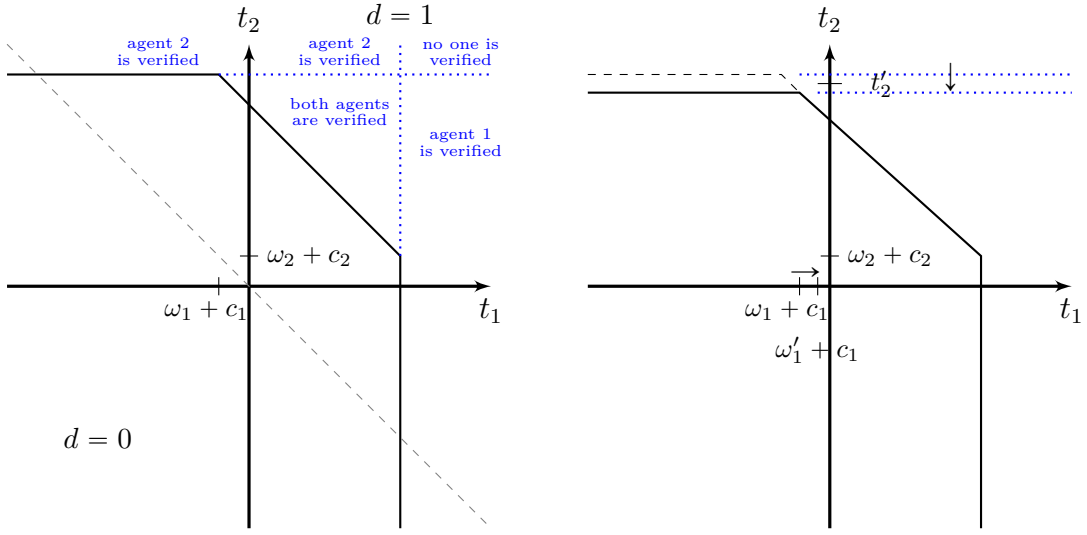
$$\int_{-\infty}^{\omega_1^+ + c_1} t_1 - (\omega_1^+ + c_1) dF_1 = -c_1.$$

A symmetric first-order condition can be derived for the optimal baseline weights for agent 2:

$$\int_{-\infty}^{\omega_2^+ + c_2} t_2 - (\omega_2^+ + c_2) dF_2 = -c_2.$$

[5] This approach can be extended to the general case with I agents and general preferences for the agents, but it becomes less tractable. The main reason for this is that the optimal cutoff for one agent is in general not independent of the other agents' optimal cutoffs. This makes the optimization problem more convoluted and the first-order conditions more complicated.

[6] Assuming that the status quo remains if the sum of the weights w_i is equal to 0.



(a) A voting-with-evidence mechanism.

(b) Optimal baseline weights.

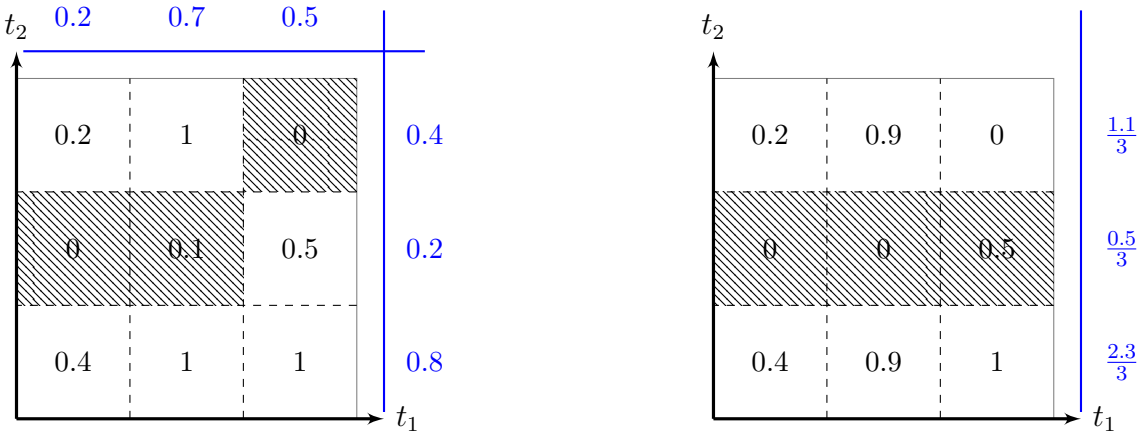
Figure 2: Illustration of a voting-with-evidence mechanism and optimal baseline weights in a two agent example.

This implies that an increase in verification costs increases the optimal baseline weights. Since it is costlier to verify an agent, the principal adjusts the decision rule to ensure that this agent is less frequently decisive. A first-order stochastic dominance shift in the distribution of types similarly increases the optimal baseline weights.

4 BIC-EPIC equivalence

A voting-with-evidence mechanism is not only Bayesian incentive compatible, but it also satisfies the stronger notion of ex-post incentive compatibility (see Remark 1). This robustness of the voting-with-evidence mechanism is a desirable property of any mechanism that one wish to use in real-life applications because optimal strategies are independent of beliefs and information structure. Reducing the number of assumptions about common knowledge and weakening the informational requirements places the theoretical analysis underpinning the design on firmer ground (Wilson (1987) and Bergemann and Morris (2005)).

Because the optimal mechanism is ex-post incentive compatible we conclude that the principal cannot gain by weakening the incentive constraints. A natural question to ask is why the principal cannot save on verification costs by implementing the optimal mechanism in Bayesian equilibrium instead of ex-post equilibrium. We show that the answer lies in a general equivalence between Bayesian and ex-post incentive compatible mechanisms. For *every* BIC mechanism, there exists an ex-post incentive compatible mechanism that induces the same interim expected decision and verification rules; since



(a) Decision rule d and its marginals in blue.

(b) Verification probabilities that are necessary for EPIC.

Figure 3: Failure of a naive BIC-EPIC equivalence.

the interim expected decision and verification rules determine the expected utility of the principal, this implies that an ex-post incentive compatible mechanism is optimal within the whole class of BIC mechanisms.

Recall that a mechanism (d, a) is BIC if and only if, for all i and t_i,

$$\inf_{t_i' \in T_i^+} \mathbb{E}_{t_{-i},s}[d(t_i', t_{-i}, s)] \geqslant \mathbb{E}_{t_{-i},s}[d(t_i, t_{-i}, s)[1 - a_i(t_i, t_{-i}, s)]] \quad \text{and} \tag{3}$$

$$\sup_{t_i' \in T_i^-} \mathbb{E}_{t_{-i},s}[d(t_i', t_{-i}, s)] \leqslant \mathbb{E}_{t_{-i},s}[d(t_i, t_{-i}, s)[1 - a_i(t_i, t_{-i}, s)] + a_i(t_i, t_{-i}, s)]. \tag{4}$$

Analogously, a mechanism (d, a) is ex-post incentive compatible (EPIC) if and only if, for all i, t_i and t_{-i},

$$\inf_{t_i' \in T_i^+} \mathbb{E}_{s}[d(t_i', t_{-i}, s)] \geqslant \mathbb{E}_{s}[d(t_i, t_{-i}, s)[1 - a_i(t_i, t_{-i}, s)]] \quad \text{and} \tag{5}$$

$$\sup_{t_i' \in T_i^-} \mathbb{E}_{s}[d(t_i', t_{-i}, s)] \leqslant \mathbb{E}_{s}[d(t_i, t_{-i}, s)[1 - a_i(t_i, t_{-i}, s)] + a_i(t_i, t_{-i}, s)]. \tag{6}$$

Not every BIC mechanism is EPIC. More important, not every decision rule that can be implemented in a Bayesian equilibrium can be implemented in an ex-post equilibrium with the same verification costs, as the following example illustrates.

Example 1. *Suppose that $\mathcal{I} = \{1, 2\}$ and that agent 2 is always in favor of the new policy. Each type profile is equally likely and the decision rule d is shown in Figure 3a. The shaded areas indicate type profiles that induce the lowest probabilities of accepting the new policy for agent 2. We focus on incentive constraints for agent 2.*

Lemma 1 shows that it is sufficient to ensure incentive compatibility for the "worst-off" types, which are the intermediate types in this example. Since the intermediate types are the worst-off, they never need to be verified. If high (low) types are verified with probability

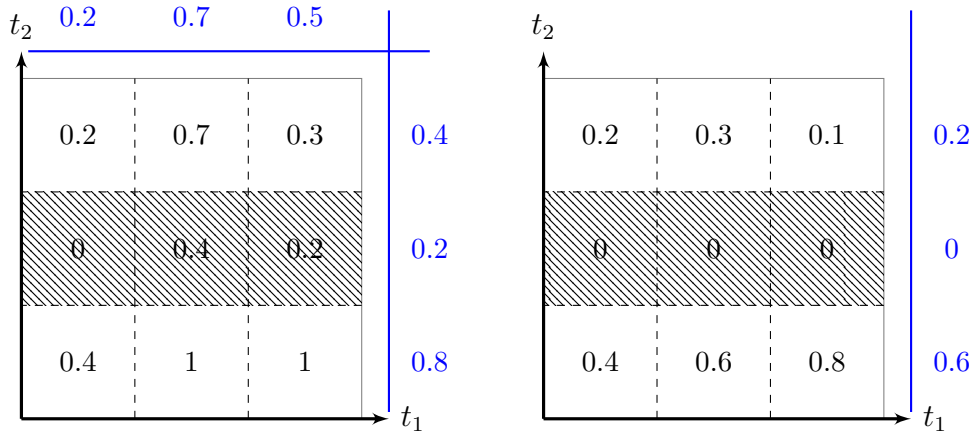
0.2 (0.6), then the Bayesian incentive constraints for the worst-off types are exactly binding. If we instead want to implement the decision rule d in an ex-post equilibrium, the cost of verification increases. For example, intermediate types must be verified with probability 0.5 if agent 1's type is high. In expectation, agent 2 must be verified with probability $\frac{0.5}{3}$ if he has an intermediate type, with probability $\frac{1.1}{3}$ if he has a high type, and with probability $\frac{2.3}{3}$ if he has a low type (the verification probabilities for each profile of reports are given in Figure 3b).

As Example 1 above illustrates, we cannot simply take a BIC mechanism, maintain the same decision rule, and expect that the mechanism will also be EPIC without increasing the verification costs. This is in line what should be expected since for a mechanism to be EPIC, incentive constraints must hold pointwise and not only in expectation. The reason for this is that in general the left-hand side of (3) is greater than the expected value of the left-hand side of (5); that is, $\inf_{t_i' \in T_i^+} \mathbb{E}_{t_{-i},s}[d(t_i', t_{-i}, s)]$ is generally larger than $\mathbb{E}_{t_{-i}}\big[\inf_{t_i' \in T_i^+} \mathbb{E}_s[d(t_i', t_{-i}, s)]\big]$. A decision rule can be implemented in ex-post equilibrium at the same costs as in Bayesian equilibrium if and only if the expectation operator commutes with the infimum/supremum operator, which is a strong requirement. However, it turns out that for every function, there exists another function that induces the same marginals and for which the expectation operator commutes with the infimum/supremum operator. We will use this result to establish an equivalence between BIC and EPIC mechanisms.

Theorem 2. *Let $A = \times_i A_i \subseteq \mathbb{R}^I$, let t_i be independently distributed with an absolutely continuous distribution function F_i, and let $g : A \to [0,1]$ be a measurable function. Then there exists a function $\hat{g} : A \to [0,1]$ with the same marginals, i. e., for all i, $\mathbb{E}_{t_{-i}}[g(\cdot, t_{-i})] = \mathbb{E}_{t_{-i}}[\hat{g}(\cdot, t_{-i})]$ almost everywhere, such that for all $B \subseteq A_i$,*

$$\inf_{t_i \in B} \mathbb{E}_{t_{-i}}[\hat{g}(t_i, t_{-i})] = \mathbb{E}_{t_{-i}}[\inf_{t_i \in B} \hat{g}(t_i, t_{-i})] \text{ and}$$
$$\sup_{t_i \in B} \mathbb{E}_{t_{-i}}[\hat{g}(t_i, t_{-i})] = \mathbb{E}_{t_{-i}}[\sup_{t_i \in B} \hat{g}(t_i, t_{-i})].$$

We will illustrate the idea behind the proof of Theorem 2 by assuming that A is finite. The argument in our proof uses Theorem 6 in Gutmann et al. (1991). This theorem shows that for any matrix with elements between 0 and 1 and with increasing row and column sums, there exists another matrix consisting of elements between 0 and 1 with the same row and column sums, and whose elements are increasing in each row and column. To use this result, we reorder A such that the marginals of g are weakly increasing. Then, Theorem 6 in Gutmann et al. (1991) implies that there exists a function $\hat{g}$ that induces the same marginals and is pointwise increasing. For this function, there is an argument t_i for each i that independent of t_{-i} minimizes $\hat{g}(\cdot, t_{-i})$. This implies that the expectation operator commutes with the infimum operator, i.e., $\mathbb{E}_{t_{-i}}[\inf_{t_i \in A} \hat{g}(t_i, t_{-i})] = \inf_{t_i \in A} \mathbb{E}_{t_{-i}}[\hat{g}(t_i, t_{-i})]$. This basic idea sketched above is extended via an approximation argument to a complete proof in Appendix A.3.



(a) Corresponding decision rule $\hat{d}$ and its marginals in blue.

(b) EPIC verification rule $\hat{a}_2$ for decision rule $\hat{d}$ and its marginals in blue.

Figure 4: Illustration of the BIC-EPIC equivalence.

Building on Theorem 2, we can establish an equivalence between BIC and EPIC mechanisms. To define this equivalence formally, we call $\mathbb{E}_{t_{-i}}[d(t_i, t_{-i})]$ the *interim decision rule* and $\mathbb{E}_{t_{-i}}[a_i(t_i, t_{-i})]$ the *interim verification rules* of a mechanism (d, a).

Definition 2. Two mechanisms (d, a) and $(\hat{d}, \hat{a})$ are *equivalent* if they induce the same interim decision and verification rules almost everywhere.

Now we can state the equivalence between BIC and EPIC mechanisms.

Theorem 3. *For any BIC mechanism (d, a), there exists an equivalent EPIC mechanism $(\hat{d}, \hat{a})$.*

There are two steps in the construction of an equivalent EPIC mechanism $(\hat{d}, \hat{a})$. In the first step, we use Theorem 2 to obtain a decision rule $\hat{d}$ with the same interim decisions as d and such that for $\hat{d}$ the expectation operator commutes with the infimum/supremum. This implies that the left-hand sides of (3) and (4), respectively are equal to the expected values of the left-hand sides of (5) and (6), respectively. In the second step, we construct a verification rule $\hat{a}$ such that all incentive constraints hold as equalities for $(\hat{d}, \hat{a})$. By potentially adding some verification, we obtain a verification rule $\hat{a}$ with the same interim verification rule as a. Thus, we have constructed an equivalent EPIC mechanism $(\hat{d}, \hat{a})$ from the BIC mechanism (d, a).

Example 1 (ctd)**.** *Figure 4b depicts the decision rule $\hat{d}$, which has the same marginals as d. Note that intermediate types of agent 2 always induce the lowest probability of accepting the proposal, independent of the type of agent 1. This implies that the expected value of the infimum equals the infimum of the expected value, that is,*

$$\inf_{t_2} \mathbb{E}_{t_1}[\hat{d}(t)] = \mathbb{E}_{t_1}[\inf_{t_2} \hat{d}(t)].$$

Figure 4b shows a verification rule $\hat{a}$ such that $(\hat{d}, \hat{a})$ is EPIC. The expected verification probabilities are the same as those necessary for implementation in Bayesian equilibrium.

The economic mechanisms behind our equivalence are different from those underlying the BIC-DIC equivalence in a standard social choice setting with transfers (with linear utilities and one-dimensional, private types (Gershkov et al. 2013)). In the standard setting, an allocation rule can be implemented with appropriate transfers in Bayesian equilibrium if and only if its marginals are increasing and in dominant strategies if and only if it is pointwise increasing. In contrast, monotonicity is neither necessary nor sufficient for implementability in our model.

Note that there is no equivalence between Bayesian and dominant-strategy incentive compatible mechanisms in our setting, as the following example illustrates. The lack of private goods to punish agents if there are multiple deviators implies that agents care whether the other agents are truthful.

Example 2. *Suppose that $\mathcal{I} = \{1, 2, 3\}$, verification costs are 0 for each agent, and $T_i^+ = \{t_i | t_i \geqslant 0\}$ and $T_i^- = \{t_i | t_i < 0\}$. Consider the voting-with-evidence mechanism with cutoffs $\omega_i^+ + c_i = 1$ and $\omega_i^- - c_i = -1$ for all i. Let $t = (-5, 2, 2)$. Given truthful reporting, the voting-with-evidence mechanism specifies that $d(t) = 0$. Suppose that agent 2 deviates from truth-telling and instead reports being of type $t_2' = 6$. Now he is decisive, and the principal verifies him. After observing the true types $(-5, 2, 2)$, the principal has to punish the lie by agent 2 and maintain the status quo to induce truthful reporting. However, this creates an incentive for agent 3 to misreport. He could report $t_3' = 6$, and then no agent is decisive; hence, no one is verified, and the voting-with-evidence mechanism specifies that $d(t_1, t_2', t_3') = 1$. The voting-with-evidence mechanism is therefore not dominant-strategy incentive compatible, no matter how we specify the mechanism off-equilibrium.*

The equivalence between Bayesian and ex-post incentive compatible mechanisms can be established in other models without money but with verification. We believe that the tools we used in this paper can prove useful in similar settings with verification. In fact, we can use arguments paralleling those used to prove Theorem 2 (but using Theorem 1 in Gershkov et al. (2013) instead of the result by Gutmann et al. (1991)) to show that there is an equivalence of Bayesian and dominant-strategy incentive compatible mechanisms in BDL.

5 Imperfect verification

Thus far, we have assumed that the verification technology works perfectly, that is, whenever the principal audits an agent, she will learn the true type with probability one. We now explore the extent to which the above results are robust to imperfect verification. We will 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 only knows 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.

In this case of imperfect verification, we have to adjust Lemma 1 to get the following characterization of Bayesian incentive compatibility:

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

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

We omit the proof which closely follows the proof of Lemma 1.

The imperfectness of the verification technology implies that that it is harder to satisfy the incentive constraints. In particular, it puts an upper bound on how much influence an agent can have in expectation: because $a_i(t,s) \leqslant 1$, Lemma 2 implies that any Bayesian incentive compatible mechanism satisfies

$$\forall t_i \in T_i^+ : \ \mathbb{E}_{t_{-i},s}[d(t_i, t_{-i}, s)] \leqslant \frac{1}{1-p} \inf_{t_i' \in T_i^+} \mathbb{E}_{t_{-i},s}[d(t_i', t_{-i}, s)] \tag{7}$$

$$\forall t_i \in T_i^- : \ \mathbb{E}_{t_{-i},s}[d(t_i, t_{-i}, s)] \geqslant \frac{1}{1-p} \left[\sup_{t_i' \in T_i^+} \mathbb{E}_{t_{-i},s}[d(t_i', t_{-i}, s)] - p \right]. \tag{8}$$

This adds an additional constraint to the relaxed problem that essentially restricts the maximal influence an agent could have on the decision rule in any incentive compatible mechanism.

Theorem 4. *With imperfect verification described as above, an optimal mechanism sets $d(t) = 1$ if and only if $\sum_i w_i(t_i) > 0$, where*

$$w_i(t_i) = \begin{cases} \omega_i^+ & \text{if } t_i \in T_i^+ \text{ and } t_i \leqslant \omega_i^+ + c_i \\ \omega_i^- & \text{if } t_i \in T_i^- \text{ and } t_i \geqslant \omega_i^- - c_i \\ t_i - \frac{c_i}{p} & \text{if } t_i \in T_i^+ \text{ and } \beta_i^+ > t_i > \alpha_i^+ \\ t_i + \frac{c_i}{p} & \text{if } t_i \in T_i^- \text{ and } \beta_i^- < t_i < \alpha_i^- \\ \beta_i^+ - \frac{c_i}{p} & \text{if } t_i \in T_i^+ \text{ and } t_i \geqslant \beta_i^+ \\ \beta_i^- + \frac{c_i}{p} & \text{if } t_i \in T_i^- \text{ and } t_i \leqslant \beta_i^- \end{cases}$$

for some constants $\{\omega_i^+, \omega_i^-, \beta_i^+, \beta_i^-\}$ satisfying $\omega_i^- \leqslant \omega_i^+$.

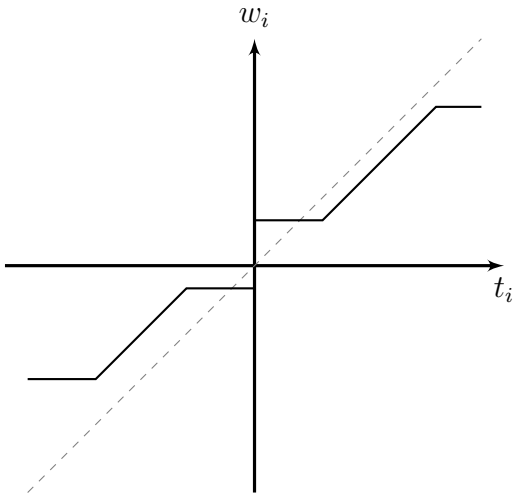


Figure 5: Example illustrating weights for imperfect verification.

Compared to the optimal mechanism in the benchmark model with perfect verification, the weight function adjusts the correction term: instead of reducing the absolute value of the weight by c_i as before, it is now reduced by the expected verification cost to detect a deviation, $\frac{c_i}{p}$. In addition, the weight function in the optimal mechanism with imperfect verification potentially restricts the maximal influence of an agent by putting a bound on the weights. This ensures that the mechanism is incentive compatible even if verification is imperfect. 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 shortlisted and will not win the prize with certainty.

This additional bound on the influence an agent can have on the decision rule is the only potential qualitative difference of the optimal decision rule compared to the model with perfect information. However, in many cases this difference will not even arise: if an optimal decision rule d (as described in Theorem 4) satisfies, for each i,

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

then $\beta_i^+ = \infty$ and $\beta_i^- = -\infty$ (see the proof of Lemma 7). Therefore, the weight function looks qualitatively as 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 $\inf_{t_i\in T_i^+}\mathbb{E}_{t_{-i}}d(t_i,t_{-i}) \geqslant \frac{1}{2}$ and $\sup_{t_i\in T_i^-}\mathbb{E}_{t_{-i}}d(t_i,t_{-i}) \leqslant \frac{1}{2}$.

Note that in an optimal mechanism with imperfect verification, agents are potentially verified although they are not decisive. Auditing an agent only when he changes the

outcome would not be sufficient because the agent would have an incentive to misreport. If he is caught, he will obtain the less-preferred outcome that he would have received in any case, but if the verification did not work he would change the outcome. Moreover, the optimal mechanism with imperfect verification is in general not ex-post incentive compatible: suppose, for example, that given the types of all other agents, an agent can change the outcome from 0 to 1. Then, he will have an incentive to do so if he knows the types of the other agents, thereby violating ex-post incentive compatibility while all Bayesian incentive constraints are satisfied.

6 Conclusion

We have analyzed a collective decision model with costly verification in which a principal decides between introducing a new policy and maintaining status quo. Agents' have private information relevant for the collective choice, and their information can be verified by the principal before she takes the decision. We have shown that a voting-with-evidence mechanism is optimal for the principal. This mechanism is not only Bayesian incentive compatible but ex-post incentive compatible. We show that this feature of robust implementation is not only valid for the optimal mechanism, but it is a general phenomenon.

For practical applications it might be important to have a model of imperfect verification. We have formulated one version of imperfect verification as a robustness check and found that the optimal mechanism is of similar character as in the case of perfect verification. For future work, it would be desirable and interesting to consider other types of departure from a perfect verification technology and to analyze how the mechanism should be adjusted. Analyzing a model of limited commitment could be another interesting direction for future research.

A 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 have been obtained in Townsend (1988) and Ben-Porath et al. (2014). We will 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 $(M_1, ..., M_I, \tilde{x}, \tilde{y})$ be an indirect mechanism, and $M = \times_{i \in \mathcal{I}} M_i$, 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.[7] Fix a Bayes-Nash equilibrium σ of the game induced by the indirect mechanism.[8]

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 or not 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

$$\begin{aligned} d(t, s) &= x(t, t, s) \\ a_i(t, s) &= y(t, t, i, s) \text{ and} \\ \ell_i(t_i', t_{-i}, t_i, s) &= x(t_i', t_{-i}, t_i, t_{-i}, s). \end{aligned}$$

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

$$\begin{aligned} & [1 - a_i(t', s)] d(t', s) + a_i(t', s)\ \ell_i(t_i', t_i, t_{-i}, s) \\ = & [1 - y(t', t', i, s)] x(t', t', s) + y(t', t', i, s)\ x(t', t, s) \\ = & \begin{cases} x(t', t, s) \text{ if } y(t', t', i, s) = 1 \\ x(t', t', s) \text{ if } y(t', t', i, s) = 0, \end{cases} \end{aligned}$$

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

[8] 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.

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

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

where $\mathbb{1}_{T_i^+}(t_i)$ denotes the indicator function for T_i^+, $\mathbb{1}_{T_i^-}(t_i)$ the indicator function for T_i^-, and $c_i(t_i) = c_i$ if $t_i \in T_i^+$ and $c_i(t_i) = -c_i$ 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), $V_P(d, a) \leqslant V_R(d)$.*

Proof.

$$\begin{aligned}
V_P(d,a) &= \mathbb{E}_t \left[\sum_i d(t)[t_i - c_i(t_i)] + c_i \mathbb{1}_{T_i^+}(t_i)[d(t) - a_i(t)] - c_i \mathbb{1}_{T_i^-}(t_i)[d(t) + a_i(t)] \right] \\
&\leqslant \mathbb{E}_t \left[\sum_i d(t)[t_i - c_i(t_i)] + c_i \mathbb{1}_{T_i^+}(t_i)[d(t)(1 - a_i(t))] - c_i \mathbb{1}_{T_i^-}(t_i)[d(t)(1 - a_i(t)) + a_i(t)] \right] \qquad (9) \\
&\leqslant \mathbb{E}_t \left[\sum_i d(t)[t_i - c_i(t_i)] + c_i \mathbb{1}_{T_i^+}(t_i) \inf_{t'_i \in T_i^+} \mathbb{E}_{t_{-i}}[d(t'_i, t_{-i})] - c_i \mathbb{1}_{T_i^-}(t_i) \sup_{t'_i \in T_i^-} \mathbb{E}_{t_{-i}}[d(t'_i, t_{-i})] \right] \qquad (10) \\
&= V_R(d).
\end{aligned}$$

The first inequality holds because $-a_i(t) \leqslant -d(t)a_i(t)$ and $d(t)a_i(t) \geqslant 0$. 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) and let $\varphi_i^+ \equiv \inf_{t_i' \in T_i^+} \mathbb{E}_{t_{-i}}[d^*(t_i', t_{-i})]$ and $\varphi_i^- \equiv \sup_{t_i' \in T_i^-} \mathbb{E}_{t_{-i}}[d^*(t_i', t_{-i})]$.

Consider the following auxiliary maximization problem:

$$\begin{aligned} \max_{0 \leqslant d \leqslant 1} \ & \mathbb{E}_t\big[\textstyle\sum_i d(t)[t_i - c_i(t_i)]\big] \qquad \text{(Aux)} \\ & \text{s.t. for all } i \in \mathcal{I}: \\ & \mathbb{E}_{t_{-i}}[d(t)] \geqslant \varphi_i^+ \text{ for all } t_i \in T_i^+, \text{ and} \\ & \mathbb{E}_{t_{-i}}[d(t)] \leqslant \varphi_i^- \text{ for all } t_i \in T_i^-, \end{aligned}$$

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

$$\begin{aligned} \mathcal{L}(d, \lambda^*) &= \mathbb{E}_t\Big[\sum_i d(t)(t_i - c_i(t_i))\Big] + \sum_i \sum_{t_i \in T_i} \Big(\lambda_i^*(t_i)\big(\mathbb{E}_{t_{-i}}[d(t_i, t_{-i})] - \varphi_i(t_i)\big)\Big) \\ &= \sum_{t \in T} d(t) \sum_i \Big(t_i - c_i(t_i) + \frac{\lambda_i^*(t_i)}{f_i(t_i)}\Big) f(t) + \textit{constant}, \end{aligned}$$

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

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

$$g(d, h^*) = \sum_{t \in T} \sum_i d(t) f(t) h_i^*(t_i). \tag{11}$$

Let

$$\alpha_i^+ = \inf_{t_i \in T_i^+} \{t_i | \mathbb{E}_{t_{-i}}[d^*(t_i, t_{-i})] > \varphi_i^+\} - c_i \tag{12}$$

$$\alpha_i^- = \sup_{t_i \in T_i^-} \{t_i | \mathbb{E}_{t_{-i}}[d^*(t_i, t_{-i})] < \varphi_i^-\} + c_i \tag{13}$$

and define

$$\bar{h}_i(t_i) := \begin{cases} \frac{1}{\mu_i(A_i^+)} \sum_{t_i \in A_i^+} f_i(t_i) h_i^*(t_i) & \text{if } t_i \in T_i^+ \text{ and } t_i \leqslant \alpha_i^+ + c_i \\ \frac{1}{\mu_i(A_i^-)} \sum_{t_i \in A_i^-} f_i(t_i) h_i^*(t_i) & \text{if } t_i \in T_i^- \text{ and } t_i \geqslant \alpha_i^- - c_i \\ t_i - c_i(t_i) & \text{otherwise,} \end{cases}$$

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 $A_i^c = T_i \backslash (A_i^+ \cup A_i^-)$ and $A_i = A_i^+ \cup A_i^-$.

Claim 1. *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 $t_i - c_i + \frac{\lambda_i^*(t_i)}{f_i(t_i)} \geqslant \alpha_i^+$ and hence for every optimal solution to the Lagrangian d that $\mathbb{E}_{t_{-i}}[d(t_i, t_{-i})] \geqslant \mathbb{E}_{t_{-i}}[d(\alpha_i^+ + c_i, t_{-i})] > \varphi_i^+$. 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

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

Step 3: $g(d^*, \bar{h}) = g(d^*, h^*) = \max_{0 \leqslant d \leqslant 1} g(d, h^*) \geqslant \max_{0 \leqslant d \leqslant 1} g(d, \bar{h})$.

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 by $H_i \equiv (H_i(t_i))_{t_i \in T_i}$. Fix an agent $i \in \mathcal{I}$, and define a function $\Psi : \mathbb{R}^{|T_i|} \to \mathbb{R}$, as $\Psi(H_i) := \max_{0 \leqslant d \leqslant 1} \sum_{t \in T} d(t) \big[f_{-i}(t_{-i}) H_i(t_i) + \sum_{j \in \mathcal{I}_{-i}} f(t) h_j^*(t_j) \big]$. 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,

$$\Psi(H_i^*) \geqslant \Psi(\bar{H}_i)$$

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 $\max_{0\leqslant d\leqslant 1} g(d, h^*_{\mathcal{I}-i}h_i)$. By repeating this argument agent by agent we can conclude that,

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

This proves the Claim 1.

Hence, every solution to the Lagrangian can be described as follows:

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

where

$$w_i(t_i) = \begin{cases} \omega_i^+ & \text{if } t_i \in T_i^+ \text{ and } t_i \leqslant \alpha_i^+ + c_i \\ \omega_i^- & \text{if } t_i \in T_i^- \text{ and } t_i \geqslant \alpha_i^- - c_i \\ t_i - c_i(t_i) & \text{otherwise} \end{cases} \tag{14}$$

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 $\omega_i^+ \geqslant \sup_{t_i\in A_i^+}\{t_i - c_i\}$ since $\lambda_i^*(t_i) \geqslant 0$ for $t_i \in A_i^+$. Also, $\omega_i^+ \leqslant \alpha_i^+$, since otherwise we would get, for $t_i \in A_i^+$, $\mathbb{E}_{t_{-i}}[d^*(t_i, t_{-i})] \geqslant \mathbb{E}_{t_{-i}}[d^*(\alpha_i^+ - c_i, t_{-i})] > \varphi_i^+$, contradicting the definition of A_i^+. Analogous arguments imply $\inf_{t_i\in A_i^-}\{t_i + c_i\} \leqslant \omega^- \leqslant \alpha_i^-$. 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 (14) 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 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 $i \in \mathcal{I}$, $n \geqslant 1$, $l_i = 1, \ldots, 2^{n+1}$, let

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

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 $l = (l_1, ..., l_n)$ 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 non-empty 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 $\omega_i^{+,n}, \omega_i^{-,n}$:

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

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

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

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

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

Then, for all i and t_i, $\mathbb{E}_{t_{-i}}[d^n(t_i, t_{-i})] = \text{Prob}[\sum_{j\neq i} r_j^n(t_j) \geqslant -r_i^n(t_i)]$ 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 achieving a strictly higher objective value, there would exist $\mathcal{F}^n$-measurable solutions achieving 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 $V_P(d, a) \leqslant V_R(d)$ 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 (9) holds as an equality for (d^*, a^*).

Note that in mechanism (d^*, a^*), all incentive constraints are binding and therefore inequality (10) holds as an equality as well. We therefore conclude $V_P(d^*, a^*) = V_R(d^*)$. □

A.3 Omitted proofs from Section 4

Proof of Theorem 2.
The proof applies Theorem 6 in (Gutmann et al. 1991) to a discrete approximation of A and by taking limits we establish Theorem 2.

Let $S_i(n, l_i)$ denote the interval,

$$S_i(n, l_i) := [F_i^{-1}((l_i - 1)2^{-n}), F_i^{-1}(l_i 2^{-n})), \qquad i \in \mathcal{I},\ n \geqslant 1 \text{ and } l_i = 1, ..., 2^n.$$

For a given n the function $S_i(n, \cdot)$ form a partition of A_i such that each partition element $S_i(n, k)$ has the same likelihood. Let $\mathcal{F}_i^n$ denote the set consisting of all possible unions of the $S_i(n, l_i)$. Note further that $\mathcal{F}_i^n \subset \mathcal{F}_i^{n+1}$. Let $l = (l_1, \ldots, l_I)$ and $S(n, l) := \prod_{i \in \mathcal{I}} S_i(n, l_i)$. Thus, for a given n the function $S(n, \cdot)$ defines a partition of A such that each partition element $S(n, l)$ has the same likelihood.

Define the following averaged function,

$$g(n, l) := 2^{In} \int_{S(n,l)} g(t) dF.$$

The function $g(n, l)$ is an I-dimensional tensor. Now consider the marginals of $g(n, l)$ with respect to l_{-i}, i.e., $\mathbb{E}_{l_{-i}}[g(n, l_i, l_{-i})]$, each such marginal in dimension i is nondecreasing in l_i. By Theorem 6 in (Gutmann et al. 1991) there exists another tensor $g'(n, l)$ with the same marginals as $g(n, l)$ such that $g'(n, l)$ is nondecreasing in l. Now define $g_n' : T \to [0, 1]$ by letting $g_n'(t) := g'(n, l)$ for all $t \in S(n, l)$.

Note that g_n' is nondecreasing in each coordinate and hence satisfies

$$\int \operatorname*{ess\,inf}_{t_i \in B} g_n'(t_i, t_{-i}) dF_{-i} = \operatorname*{ess\,inf}_{t_i \in B} \int g_n'(t_i, t_{-i}) dF_{-i} \tag{15}$$

$$\int \operatorname*{ess\,sup}_{t_i \in B} g_n'(t_i, t_{-i}) dF_{-i} = \operatorname*{ess\,sup}_{t_i \in B} \int g_n'(t_i, t_{-i}) dF_{-i}. \tag{16}$$

Moreover,

$$\int_{S_i(n,l_i)} \int_{A_{-i}} g(t_i, t_{-i}) dF_{-i}\ dF_i = \int_{S_i(n,l_i)} \int_{A_{-i}} g_n'(t_i, t_{-i}) dF_{-i}\ dF_i, \tag{17}$$

and hence $g(t) - g_n'(t)$ integrates to zero over sets of the form $S_i(n, l_i) \times A_{-i}$ for every $S_i(n, l_i) \in \mathcal{F}_i^n$.

Draw a weak*-convergent subsequence from the sequence $\{g'_n\}$ (which is possible by Alaoglu's theorem) and denote its limit by $\hat{g}$. This function rule satisfies $0 \leqslant \hat{g} \leqslant 1$ and its marginals are equal almost everywhere to the marginals of g because of (17).

Since $g'_n \to^* \hat{g}$, we get
$\operatorname{ess\,inf}_{t_i \in B} g'_n(t_i, t_{-i}) \to \operatorname{ess\,inf}_{t_i \in B} \hat{g}(t_i, t_{-i})$ for almost every t_{-i}. Moreover,
$\operatorname{ess\,inf}_{t_i \in B} \int_{A_{-i}} g'_n(t_i, t_{-i}) dF_{-i} \to \operatorname{ess\,inf}_{t_i \in B} \int_{A_{-i}} \hat{g}(t_i, t_{-i}) dF_{-i}$. Note further that,
$\mathbb{E}_{t_{-i}}[\inf_{t_i \in T_i^+} \hat{g}(t_i, t_{-i})] \leqslant \inf_{t_i \in T_i^+} \mathbb{E}_{t_{-i}}[\hat{g}(t_i, t_{-i})]$ always holds. By way of contradiction suppose now that for some i,

$$\int \operatorname*{ess\,inf}_{t_i \in B} \hat{g}(t_i, t_{-i}) dF_{-i} < \operatorname*{ess\,inf}_{t_i \in B} \int \hat{g}(t_i, t_{-i}) dF_{-i}.^{9}$$

This implies

$$\int \operatorname*{ess\,inf}_{t_i \in B} g'_n(t_i, t_{-i}) dF_{-i} < \operatorname*{ess\,inf}_{t_i \in B} \int g'_n(t_i, t_{-i}) dF_{-i}$$

for n large enough, contradicting (15) and thereby proving the first equality in the theorem. Analogous arguments apply for the second equality in the theorem, thus establishing our claim. □

Proof of Theorem 3.
It follows from Theorem 2 that there exists a decision rule $\hat{d} : T \times [0,1] \to \{0,1\}$ that induces the same marginals almost everywhere and for which

$$\inf_{t_i \in T_i^+} \mathbb{E}_{t_{-i},s}[\hat{d}(t_i, t_{-i}, s)] = \mathbb{E}_{t_{-i}}[\inf_{t_i \in T_i^+} \mathbb{E}_s \hat{d}(t_i, t_{-i}, s)] \text{ and}$$
$$\sup_{t_i \in T_i^-} \mathbb{E}_{t_{-i},s}[\hat{d}(t_i, t_{-i}, s)] = \mathbb{E}_{t_{-i}}[\sup_{t_i \in T_i^-} \mathbb{E}_s \hat{d}(t_i, t_{-i}, s)].$$

We now construct a verification rule $\hat{a}$ such that the mechanism $(\hat{d}, \hat{a})$ satisfies the claim. By setting

$$\hat{a}_i(t,s) := \begin{cases} \frac{1}{\text{Prob}_s(\hat{d}(t,s)=1)} \left(\mathbb{E}_{s'}[\hat{d}(t,s')] - \inf_{t'_i \in T_i^+} \mathbb{E}_{s'}[\hat{d}(t'_i, t_{-i}, s')] \right) & \text{if } \hat{d}(t,s) = 1 \\ \frac{1}{\text{Prob}_s(\hat{d}(t,s)=0)} \left(\sup_{t'_i \in T_i^-} \mathbb{E}_{s'}[\hat{d}(t'_i, t_{-i}, s')] - \mathbb{E}_{s'}[\hat{d}(t,s')] \right) & \text{if } \hat{d}(t,s) = 0, \end{cases}$$

[9] If the inequality only holds for the infimum but not for the essential infimum, we can adjust $\hat{g}$ on a set of measure zero such that our claim holds.

the mechanism $(\hat{d}, \hat{a})$ satisfies (5) as an equality for all t_i, t_{-i}:

$$\mathbb{E}_s[\hat{d}(t,s)(1-\hat{a}_i(t,s))] = \int_{s:\hat{d}(t,s)=1} 1 - \frac{1}{\text{Prob}_s(\hat{d}(t,s)=1)} \left[\mathbb{E}_{s'}[\hat{d}(t,s')] - \inf_{t_i' \in T_i^+} \mathbb{E}_{s'}[\hat{d}(t_i', t_{-i}, s')]\right] ds$$

$$= \int_{s:\hat{d}(t,s)=1} 1 - \frac{1}{\text{Prob}_s(\hat{d}(t,s)=1)} \left[\int_{s':\hat{d}(t,s')=1} \text{Prob}_{s'}(\hat{d}(t,s')=1) ds' - \inf_{t_i' \in T_i^+} \mathbb{E}_{s'}[\hat{d}(t_i', t_{-i}, s')]\right] ds$$

$$= \int_{s:\hat{d}(t,s)=1} \frac{1}{\text{Prob}_s(\hat{d}(t,s)=1)} \left[\inf_{t_i' \in T_i^+} \mathbb{E}_{s'}[\hat{d}(t_i', t_{-i}, s')]\right] ds$$

$$= \inf_{t_i' \in T_i^+} \mathbb{E}_s[\hat{d}(t_i', t_{-i}, s)].$$

Similarly, the mechanism satisfies (6) as an equality and hence it is EPIC.

Moreover,

$$\begin{aligned}
\mathbb{E}_{t_{-i},s}[\hat{a}_i(t,s)] &= \mathbb{E}_{t_{-i},s}\Big[\hat{a}_i(t,s) + \hat{d}(t,s)[1-\hat{a}_i(t,s)] - \hat{d}(t,s)[1-\hat{a}_i(t,s)]\Big] \\
&= \mathbb{E}_{t_{-i}}\left[\sup_{t_i' \in T_i^-} \mathbb{E}_s \hat{d}(t_i', t_{-i}, s) - \inf_{t_i' \in T_i^+} \mathbb{E}_s \hat{d}(t_i', t_{-i}, s)\right] \\
&= \sup_{t_i' \in T_i^-} \mathbb{E}_{t_{-i},s}[d(t_i', t_{-i}, s)] - \inf_{t_i' \in T_i^+} \mathbb{E}_{t_{-i},s}[d(t_i', t_{-i}, s)] \\
&\leqslant \mathbb{E}_{t_{-i},s}[a_i(t,s)],
\end{aligned}$$

where the second equality follows from the fact that (5) and (6) are binding, the third equality follows from Step 1 and the fact that d and $\hat{d}$ induce the same marginals, and the inequality follows from the fact that (d, a) is BIC. Hence, by potentially adding additional verifications one obtains an EPIC mechanism that induces the same interim decision and verification probabilities. □

A.4 Proof of Theorem 4

Consider the relaxed problem

$$\max_{0 \leqslant d \leqslant 1} \mathbb{E}_t\Big[\sum_i d(t)[t_i - \tfrac{c_i(t_i)}{p}] + \tfrac{c_i}{p}\Big(\mathbb{1}_{T_i^+}(t_i) \inf_{t_i' \in T_i^+} \mathbb{E}_{t_{-i}}[d(t_i', t_{-i})] - \mathbb{1}_{T_i^-}(t_i) \sup_{t_i' \in T_i^-} \mathbb{E}_{t_{-i}}[d(t_i', t_{-i})]\Big)\Big] \quad \text{s.t. (7) and (8)} \tag{$\tilde{R}$}$$

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, $V_{\tilde{P}}(d, a) \leqslant V_{\tilde{R}}(d)$.*

Proof. Note that Lemma 3 implies that

$$\forall t_i \in T_i^+ : \ \mathbb{E}_{t_{-i},s}[a_i(t_i,t_{-i},s)d(t_i,t_{-i},s)] \geqslant \frac{1}{p}\left[\mathbb{E}_{t_{-i},s}[d(t_i,t_{-i})] - \inf_{t_i \in T_i^+} \mathbb{E}_{t_{-i},s}[d(t_i',t_{-i},s)]\right] \text{ and} \tag{18}$$

$$\forall t_i \in T_i^- : \ \mathbb{E}_{t_{-i},s}[a_i(t_i,t_{-i},s)[1-d(t_i,t_{-i},s)]] \geqslant \frac{1}{p}\left[\sup_{t_i \in T_i^+} \mathbb{E}_{t_{-i},s}[d(t_i',t_{-i},s)] - \mathbb{E}_{t_{-i},s}[d(t_i,t_{-i})]\right] \tag{19}$$

Hence,

$$\begin{aligned}
V_{\tilde{P}}(d,a) &= \mathbb{E}_t\left[\sum_i d(t)t_i - a_i(t)c_i\right] \\
&\leqslant \mathbb{E}_t\left[\sum_i d(t)t_i - \mathbb{1}_{T_i^+}(t_i)d(t)a_i(t)c_i - \mathbb{1}_{T_i^-}(t_i)[1-d(t)]a_i(t)c_i\right] \qquad (20)\\
&\leqslant \mathbb{E}_{t_i}\left[\sum_i \mathbb{E}_{t_{-i}}[d(t)]t_i - \mathbb{1}_{T_i^+}(t_i)\frac{1}{p}\left[\mathbb{E}_{t_{-i},s}[d(t_i,t_{-i})] - \inf_{t_i \in T_i^+} \mathbb{E}_{t_{-i},s}[d(t_i',t_{-i},s)\right]c_i\right. \\
&\qquad\qquad \left. - \mathbb{1}_{T_i^-}(t_i)\frac{1}{p}\left[\sup_{t_i \in T_i^+} \mathbb{E}_{t_{-i},s}[d(t_i',t_{-i},s)] - \mathbb{E}_{t_{-i},s}[d(t_i,t_{-i})]\right]c_i\right] \qquad (21)\\
&= V_{\tilde{R}}(d)
\end{aligned}$$

□

Lemma 7. *Suppose T is finite. The decision rule stated in Theorem 4 solves problem $(\tilde{R})$.*

Proof. Let d^* denote an optimal solution to the relaxed problem $(\tilde{R})$ 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 following problem:

$$\begin{aligned}
&\max_{0 \leqslant d \leqslant 1} \mathbb{E}_t\big[\textstyle\sum_i d(t)[t_i - c_i(t_i)]\big] && \text{(Aux)}\\
&\qquad \text{s.t. for all } i \in \mathcal{I}: \\
&\qquad \varphi_i^+ \leqslant \mathbb{E}_{t_{-i}} d(t) \leqslant \frac{\varphi_i^+}{1-p} && \text{for all } t_i \in T_i^+, \text{ and}\\
&\qquad \frac{\varphi_i^- - p}{1-p} \leqslant \mathbb{E}_{t_{-i}} d(t) \leqslant \varphi_i^- && \text{for all } t_i \in T_i^-.
\end{aligned}$$

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:

$$\begin{aligned}\mathcal{L}(d,\lambda,\mu) = \mathbb{E}_t\Big[\sum_i d(t)(t_i - c_i(t_i))\Big] \\ + \sum_i \sum_{t_i \in T_i^+} \Big(\lambda_i(t_i)\big(\mathbb{E}_{t_{-i}}[d(t_i,t_{-i})] - \varphi_i^+\big) + \mu_i(t_i)\big(\frac{\varphi_i^+}{1-p} - \mathbb{E}_{t_{-i}}[d(t_i,t_{-i})]\big)\Big) \\ + \sum_i \sum_{t_i \in T_i^-} \Big(\lambda_i(t_i)\big(\mathbb{E}_{t_{-i}}[d(t_i,t_{-i})] - \varphi_i^-\big) + \mu_i(t_i)\big(\frac{\varphi_i^- - p}{1-p} - \mathbb{E}_{t_{-i}}[d(t_i,t_{-i})]\big)\Big)\end{aligned}$$

Define $h_i(t_i) := t_i - c_i(t_i) + \frac{\lambda_i(t_i)+\mu_i(t_i)}{f_i(t_i)}$ and let

$$\begin{aligned}\alpha_i^+ &= \inf_{t_i \in T_i^+}\{t_i | \mathbb{E}_{t_{-i}}[d^*(t_i,t_{-i})] > \varphi_i^+\} \\ \alpha_i^- &= \sup_{t_i \in T_i^-}\{t_i | \mathbb{E}_{t_{-i}}[d^*(t_i,t_{-i})] < \varphi_i^-\} \\ \beta_i^+ &= \sup_{t_i \in T_i^+}\{t_i | \mathbb{E}_{t_{-i}}[d^*(t_i,t_{-i})] < \tfrac{\varphi_i^+}{1-p}\} \\ \beta_i^- &= \inf_{t_i \in T_i^-}\{t_i | \mathbb{E}_{t_{-i}}[d^*(t_i,t_{-i})] > \tfrac{\varphi_i^- - p}{1-p}\}.\end{aligned}$$

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

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

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 4 solves problem $(\tilde{R})$.*

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

Proof of Theorem 4. Denote by d^* the solution to problem $\tilde{R}$. For each i, define $q_i : T_i \to [0,1]$ as the solution to

$$\begin{aligned}\mathbb{E}_{t_{-i}}[d^*(t_i,t_{-i})\ [1 - p \cdot q_i(t_i)]] &= \inf_{t_i' \in T_i^+} \mathbb{E}_{t_{-i}}[d^*(t_i',t_{-i})] && \text{, for } t_i \in T_i^+ \text{ and} \\ \mathbb{E}_{t_{-i}}[d^*(t_i,t_{-i})[1 - p \cdot q_i(t_i)] &= \sup_{t_i' \in T_i^-} \mathbb{E}_{t_{-i}}[d^*(t_i',t_{-i})] - p \cdot q_i(t_i)] && \text{, for } t_i \in T_i^-.\end{aligned}$$

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

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

and setting $q_i(t_i) = 1$ yields

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

where the inequality follows from (7). 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 $t_i \in T_i^+$,

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

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

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

Hence, (d^*, a^*) is Bayesian incentive compatible by Lemma 2 and inequality (21) 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 (20) also holds as an equality and we conclude $V_{\tilde{P}}(d^*, a^*) = V_{\tilde{R}}(d^*)$. Hence, (d^*, a^*) is optimal. □

References

Arrow, K. J., Hurwicz, L. and Uzawa, H. (1961). Constraint qualifications in maximization problems, *Naval Research Logistics (NRL)* **8**(2): 175–191.

Azreli, Y. and Kim, S. (2014). Pareto efficiency and weighted majority rules, *International Economic Review* **55 No.4**: 1067–1088.

Ben-Porath, E., Dekel, E. and Lipman, B. L. (2014). Optimal allocation with costly verification, *American Economic Review* **104**: 3779–3813.

Ben-Porath, E., Dekel, E. and Lipman, B. L. (2017). Mechanisms with evidence: Commitment and robustness, *Working paper* .

Ben-Porath, E. and Lipman, B. L. (2012). Implementation with partial provability, *Journal of Economic Theory* **147**: 1689–1724.

Bergemann, D. and Morris, S. (2005). Robust mechanism design, *Econometrica* **73**(6): 1771–1813.

Border, K. C. and Sobel, J. (1987). Samurai accountant: A theory of auditing and plunder, *Review of Economic Studies* **54 (4)**: 1175–1187.

Bull, J. and Watson, J. (2007). Hard evidence and mechanism design, *Games and Economic Behavior* **58**: 75–93.

Deneckere, R. and Severinov, S. (2008). Mechanism design with partial state verifiability, *Games and Economic Behavior* **64**: 487–513.

Gale, D. and Hellwig, M. (1985). Incentive-compatible debt contracts: the one-period problem, *Review of Economic Studies* **52 (4)**: 647–663.

Gershkov, A., Goeree, J. K., Kushnir, A., Moldovanu, B. and Shi, X. (2013). On the equivalence of bayesian and dominant strategy implementation, *Econometrica* **81 No.1**.

Gershkov, A., Moldovanu, B. and Shi, X. (2016). Optimal voting rules, *Review of Economic Studies* **84 No. 2**: 688–717.

Glazer, J. and Rubinstein, A. (2004). On optimal rules of persuasion, *Econometrica* **72 No. 6**: 1715–1736.

Glazer, J. and Rubinstein, A. (2006). On optimal rules of persuasion, *Theoretical Economics* **1**: 395–410.

Green, J. R. and Laffont, J.-J. (1986). Partially verifiable information and mechanism design, *Review of Economic Studies* **53 No.3**: 447–456.

Gutmann, S., Kemperman, J. H. B., Reeds, J. A. and Shepp, L. A. (1991). Existence of probability measures with given marginals, *The Annals of Probability* **19**(4): 1781–1797.

Halac, M. and Yared, P. (2017). Commitment vs. flexibility with costly verification, *Working paper* .

Luenberger, D. G. (1969). *Optimization by Vector Space Methods*, John Wiley & Sons, New York.

Manelli, A. M. and Vincent, D. R. (2010). Bayesian and dominant-strategy implementation in the independent private-values model, *Econometrica* **78 No.6**.

Mylovanov, T. and Zapechelnyuk, A. (2017). Optimal allocation with ex post verification and limited penalties, *American Economic Review* **107**(9): 2666–94.

Rae, D. W. (1969). Decision-rules and individual values in constitutional choice, *The American Political Scinece Review* **63**: 40–56.

Schmitz, P. W. and Tröger, T. (2012). The (sub-)optimality of the majority rule, *Games and Economic Behavior* **74**: 651–665.

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

Townsend, R. M. (1988). Information constrained insurance, *Journal of Monetary Economics* **21**: 411–450.

Wilson, R. (1987). Game-theoretic analysis of trading, *in* T. Bewley (ed.), *Advances in Economic Theory: Invited papers for the Sixth World Congress of the Econometric Society*, Cambridge University Press, pp. 33–70.