Costly Verification in Collective Decisions*

Albin Erlanson Andreas Kleiner

November 2015

Abstract

We study how a principal should optimally choose between implementing a new policy and keeping 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. This mechanism can be implemented as a weighted majority voting rule, where agents are given additional weight if they provide evidence about 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.

Keywords: Collective decision; Costly verification

JEL classification: D82, D71

*For the latest draft, see https://sites.google.com/site/andykleiner. 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 Francesc Dilme, Bart Lipman and seminar participants at Bonn University and at various conferences for insightful comments. Albin Erlanson gratefully acknowledges financial support from the European Research Council and the Jan Wallander and Tom Hedelius Foundation. University of Bonn, E-mail: Albin Erlanson aerlanso@uni-bonn.de and Andreas Kleiner akleiner@uni-bonn.de

1 Introduction

The decision on whether a newly approved pharmaceutical drug should be subsidized in Sweden is determined by the Dental and Pharmaceutical Benefits Board (TLV). The producer of the drug can apply for a subsidy by providing arguments for clinical and cost-effectiveness of the drug. Other stakeholders are also given an opportunity to participate in the deliberations by contributing with relevant information for 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, evidence for cost-effectiveness should be provided through analysis in a health economic model. TLV can verify the information provided, but it is costly to do so. For example, TLV occasionally has to build their own health-economic models or hire external experts to evaluate the evidence that was provided, which induces significant costs. When should TLV invest effort and money to verify the evidence? What decision rule should 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. To learn about costly verification we consider a setting with a principal that decides between introducing a new policy and maintaining status quo. The principal's optimal choice depends on agents' private information. Agents can be in favor or against the new policy, and they are strategic in revealing their information since it influences the decision taken 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 principal's 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 by the principal 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.

To analyze our model, 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: Firstly, which reports will be verified (*verification rules*)? Secondly, what is the decision regarding the new policy (*decision rule*)? Lastly, what is the punishment when someone is revealed of lying? Because we can focus on incentive compatible mechanisms, punishments will be imposed only off the equilibrium path. The principal can therefore always choose the severest possible punishment, as this weakens incentive constraints but does not affect the decision taken 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 or against the new policy. The decision rule compares the sum of weighted votes in favor 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 corresponding to the importance of his information. We say that an agent provides *decisive* evidence if the decision on the policy changes if the agent merely 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. 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. We show that this is not a coincidence, but a general feature of our model. The principal can obtain this robustness of any Bayesian incentive compatible mechanism without requiring 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. As a technical tool to establish this equivalence we show that for any measurable function there exists a function with the same marginals and for which the expectation operator commutes with the infimum/supremum operator.

For purposes of practical applications, there are three main features to be learned for the design of real-world mechanisms. 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 costs of verification 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.

Related Literature

There is a large literature on collective choice problems with two alternatives when monetary transfers are not possible. One particular strand of this literature, going back to the seminal work by 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 or against the policy (Schmitz and Tröger 2012, Azreli and Kim 2014).[1] Introducing a technology to learn the agents' information allows for a much richer class of decision rules that can be implemented. Our main interest lies in understanding how this additional possibility opens up for other implementable mechanisms, and changes the optimal decision rule.

Townsend (1979) introduces costly verification in a principal-agent model. 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. The closest paper to ours is the seminal work by Ben-Porath, Dekel and Lipman (2014, henceforth BDL). They consider a principal that wishes to allocate an indivisible good among a group of agents where 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 optimal mechanism, i.e., the mechanism that maximizes the expected utility of the principal subject to the incentive constraints. While we also study a model with costly verification and without transfers, we are interested in optimal mechanisms in a collective choice problem, where voting rules are optimal in the absence of a verification technology. Having derived the optimal mechanism in our model allows us to analyze and understand which features of the optimal mechanism found by BDL carry over to other models with costly verification. We discuss this question in more detail in Section 6. Mylovanov and Zapechelnyuk (2014) also study the allocation of an indivisible good, though in contrast to BDL the principal always learns the private information of the agents, but only after having made the allocation decision and the principal has only limited punishment at disposal. Glazer and Rubinstein (2004, 2006) consider a situation when 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.[2]

[1] See also Gershkov, Moldovanu and Shi (2014) for a recent extension to settings with more than two alternatives.

[2] For 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).

Our result on the equivalence between Bayesian and ex-post incentive compatible mechanisms relates our work to several papers establishing 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 have been introduced to 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 cutoff mechanisms and discuss their optimality, while Section 5 contains the proof of the optimality of the cutoff mechanisms. We establish an equivalence of Bayesian and ex-post incentive compatible mechanisms in Section 4. In Section 6 we discuss in detail the relation of our paper and BDL. Section 7 concludes the paper.

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 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 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 perfectly learns the true type of agent i. For an agent it induces no costs to be verified. 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. 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 $T \equiv \prod_i T_i$.

We show in Appendix A.1 that it is without loss of generality to focus on direct mechanisms with truth-telling as a Bayesian 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 nobody 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

[3]With a slight abuse of notation, we will drop the realization of the randomization device as an

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. To simplify notation we assume $T_i = T_i^+ \cup T_i^-$.

Truth-telling is a Bayesian equilibrium for the mechanism (d, a, ℓ) if and only if the mechanism (d, a, ℓ) is Bayesian incentive compatible, 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 reports his type t_i' truthfully and all others report truthfully as well. 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

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

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

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

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

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

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 $a_i(m)$.

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

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

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 wlog 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 Cutoff mechanisms

In this section we introduce and illustrate the class of cutoff mechanisms. To describe any cutoff mechanism it is enough to specify for each agent i a pair of scalars, one for supporting the new policy, α_i^+, and one for opposing the new policy, α_i^-. We will show that a cutoff mechanism is optimal. Therefore, the complex optimization problem of maximizing the expected utility of the principal subject to incentive compatibility reduces to optimizing over a profile of cutoffs, a significantly simpler problem.

In a cutoff mechanism each agent reports his type and this report is altered by letting its absolute value being reduced by the verification cost (we call the result *net type*), and reports in T_i^+ below α_i^+ (in T_i^- above α_i^-) are replaced by constants (which we call *baseline reports*). The decision rule d then implements the decision that would be efficient if the altered reports were the true types.

To formally define a *cutoff mechanism* with cutoffs $\{\alpha_i^+, \alpha_i^-\}_{i \in \mathcal{I}}$, where $\alpha_i^- \leqslant \alpha_i^+$ we first formalize the concept of altered reports. For each report t_i, the *altered report* is

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

Figure 1 illustrates how the altered reports are determined.

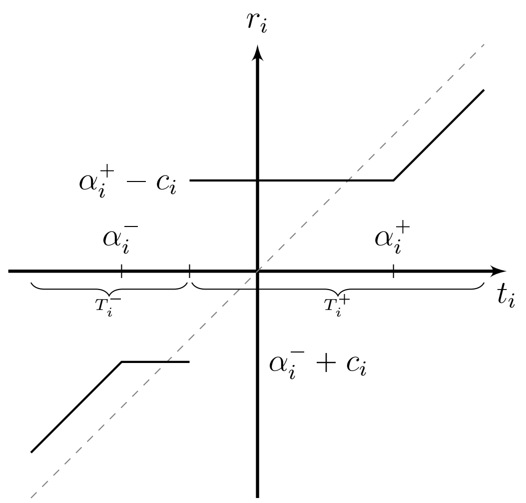


Figure 1: Example illustrating how altered reports are determined.

Given the altered reports, a cutoff mechanism uses the following decision rule:

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

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 is replaced by his relevant cutoff (α_i^+ if he is in favor and α_i^- if he prefers status quo). A cutoff mechanism verifies an agent if and only if he is decisive.[4]

Remark 1 (Incentive compatibility of cutoff mechanisms)**.** We will now show that a cutoff mechanism is incentive compatible. Let $t \in T$ be a profile of types and 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

[4]Our definition of a cutoff mechanism does not specify a decision if the altered reports add up to zero. This is a either a probability zero event, in which case the decision does not affect the principal's expected utility. Or this happens if the baseline reports add up to zero, in which case it is an easy exercise to determine the optimal decision.

by reporting some $t_i' > t_i$ and $t_i' > \alpha_i^+$. But, 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 punishment is to keep 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 argument in fact imply that a cutoff mechanism is *ex-post incentive compatible.*

A cutoff mechanism can be interpreted as a weighted majority voting rule, where agents have the additional option to make specific claims in order to gain additional influence. To see this, consider the following indirect mechanism. Each agent casts a vote either in favor 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 $\alpha_i^+ - c_i$ and $-\alpha_i^- - c_i$ if he votes in favor respectively against the new policy. 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 cutoff mechanism. Any agent 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 on 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.

We are now ready to state our main result.

Theorem 1. *A cutoff mechanism maximizes the expected utility of the principal.*

Section 5 contains the proof of Theorem 1 for finite type spaces, and the proof is extended to infinite type spaces through an approximation argument in Appendix A.2. Before illustrating a cutoff mechanism in a two agent example we will give an intuition for why cutoff mechanisms are optimal.

A cutoff mechanism differs in three respects from the first-best. 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 of all, the principal verifies all decisive agents and incurs the corresponding costs which he need not do if the information was public. Clearly, verifying decisive agents is necessary to satisfy the incentive constraints. Moreover, in a cutoff 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. More precisely any report $t_i \in T_i^+$ and above α_i^+ is replaced by the net type $t_i - c_i$. Similarly are types $t_i \in T_i^-$ and below α_i^- replaced by the net type $t_i + c_i$. The reason why this is part of an optimal mechanism has to do with decisiveness and when the decision on the policy 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. But 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, $t_i - c_i$, 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.

The third inefficiency arises from the fact that all types below the cutoff α_i^+ of an agent in favor of the policy are bunched together and receive the same altered report, the baseline report $\alpha_i^+ - c_i$. Similarly are all types above the cutoff α_i^- and against the policy bunched together into the baseline report $\alpha_i^- + c_i$. Suppose instead that in the optimal mechanism there was 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 cutoff mechanism.

We end this section by explaining a cutoff mechanism in an example with two agents, and we show how to determine the optimal cutoffs in this example. We assume that both agents prefer the new policy compared to status quo, independently of their types. The cutoff 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, since a claim is not decisive when the claim reported by the other agent already induces the principal to implement the 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 by the baseline report the policy would not be implemented.

To determine the optimal cutoffs we use a first-order approach.[5] Consider a slight increase in the cutoff 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 $\alpha_1^+ - c_1 + t_2' - c_2 = 0$. Therefore, suppose that agent 2's type is t_2' and that agent 1's type is below α_1^+. If cutoff α_1^+ is used, the policy will not be implemented.[6] However, if the cutoff is slightly increased, then the new

[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 are more complicated.

[6]Assuming status quo remains if altered reports sum to 0.

policy will be implemented, agent 2 becomes decisive, and therefore agent 2 has to be verified. Hence, the principal's expected utility changes by

$$f_2(t_2') \int_{-\infty}^{\alpha_1^+} t_1 + t_2' - c_2 \; dF_1.$$

Since the new policy will be implemented at type profile (α_1^+, t_2') under the higher cutoff, agent 1 is not decisive at profiles (t_1, t_2') for $t_1 > \alpha_1^+$ (for these profiles he would be decisive if the smaller cutoff was used). Consequently, the principal can save verification costs, which increases his utility by

$$f_2(t_2') \int_{\alpha_1^+}^{\infty} c_1 \; dF_1.$$

At the optimal cutoffs these two effects add up to zero. Using that $t_2' - c_2 = -\alpha_1^+ + c_1$, this yields the following first-order condition for the optimal cutoff for agent 1:

$$\int_{-\infty}^{\alpha_1^+} t_1 - \alpha_1^+ dF_1 = -c_1.$$

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

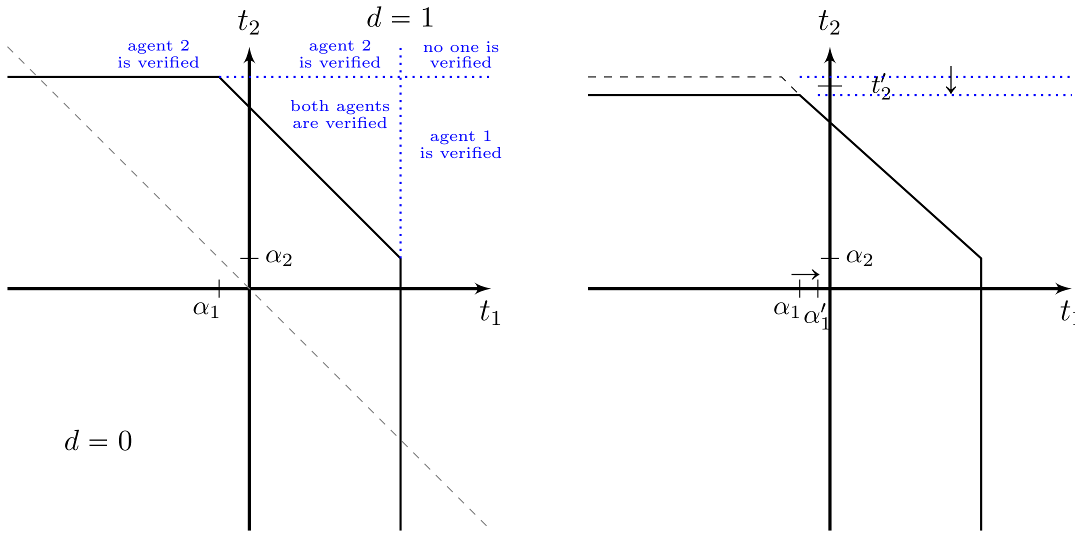
$$\int_{-\infty}^{\alpha_2^+} t_2 - \alpha_2^+ dF_2 = -c_2.$$

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

4 BIC-EPIC equivalence

A cutoff mechanism is not only Bayesian incentive compatible, it satisfies the stronger notion of ex-post incentive compatibility (see Remark 1). This robustness of the cutoff mechanism is a desirable property of any mechanism we wish to use in real-life applications because optimal strategies are independent of beliefs and information structure. By reducing the number of assumptions on common knowledge and weakening the informational requirements the theoretical analysis underpinning the design stands 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 the interim



(a) A cutoff mechanism.

(b) Optimal cutoffs.

Figure 2: Illustration of a cutoff mechanism and optimal cutoffs in a two agent example.

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},

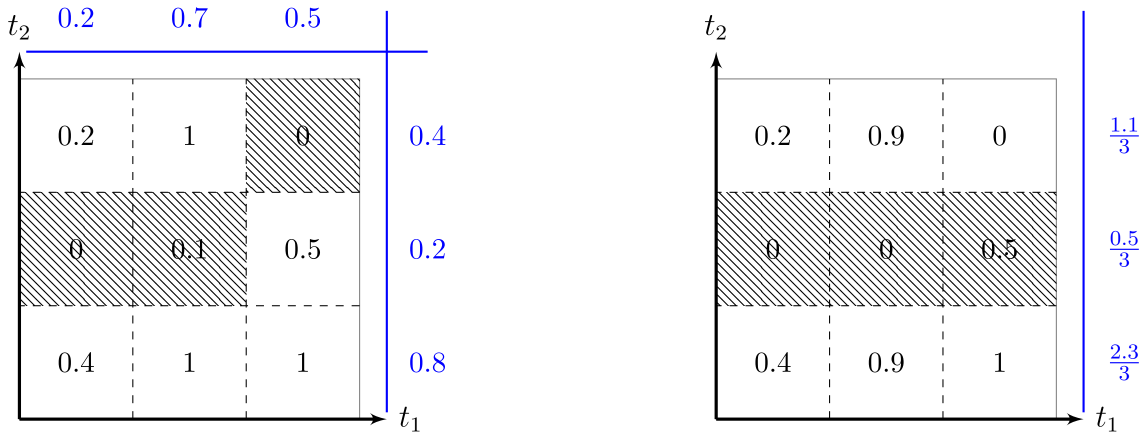
$$\inf_{t_i' \in T_i^+} \mathbb{E}_s[d(t_i', t_{-i}, s)] \geqslant \mathbb{E}_s[d(t_i, t_{-i}, s)[1 - a_i(t_i, t_{-i}, s)]] \quad \text{and} \tag{5}$$

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

Not every BIC mechanism is EPIC. More importantly, 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 enough to ensure incentive compatibility for the "worst-off" types, which are the intermediate types in this example. Since intermediate types are



(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.

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, keep the same decision rule, and expect that the mechanism will also be EPIC without increasing the verification costs. This is in line what to 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 $\mathbb{E}_{t_{-i}} \inf_{t'_i \in T_i^+} \mathbb{E}_s[d(t'_i, t_{-i}, s)]$. 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 which induces the same marginals and for which the expectation operator commutes with the infimum/supremum operator. We will use this resultto 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 $\hat{g}$ which induces the same marginals and which is pointwise increasing. For this function, there is an argument t_i for each i which independently 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.

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.

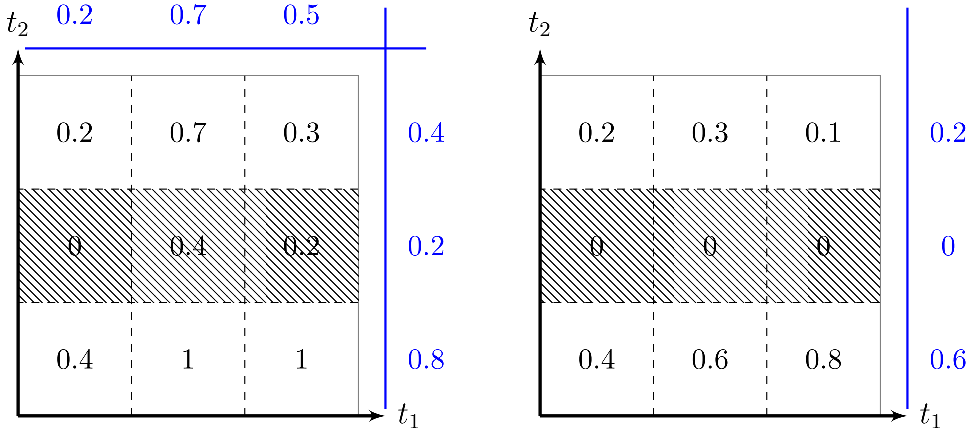
Theorem 3. *For any BIC mechanism (d, a) there exists an equivalent EPIC mechanism $(\hat{d}, \hat{a})$.*

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) resp. (4) are equal to the expected values of the left-hand sides of (5) resp. (6). 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 shows 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, independently 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 that are necessary for implementation in Bayesian equilibrium.



(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.

The economic mechanisms behind our equivalence are different from the ones underlying the BIC-DIC equivalence in a standard social choice setting with transfers (with linear utilities and one-dimensional, private types). 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 $\mathcal{I} = \{1, 2, 3\}$, 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 cutoff mechanism with cutoffs $\alpha_i^+ = 1$ and $\alpha_i^- = -1$ for all i. Let $t = (-5, 2, 2)$. Given truthful reporting the cutoff mechanism specifies $d(t) = 0$. Suppose agent 2 deviates from truth-telling and instead reports to be 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 keep the status quo to induce truthful reporting. But 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 cutoff mechanism specifies $d(t_1, t_2', t_3') = 1$. The cutoff 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 the ones used to proof 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 Proof of Theorem 1

In this section we show that a cutoff mechanism maximizes the expected utility of the principal.

We will study the following problem, and show below that it 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 $\mathbb{1}_{T_i^+}$ denotes the indicator function for T_i^+, $\mathbb{1}_{T_i^-}$ 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 2. *For any Bayesian incentive compatible mechanism* (d, a), $V_P(d, a) \leqslant V_R(d)$.

Proof.

$$\begin{aligned}
V_P(d,a) = {} & \mathbb{E}_t\left[\sum_i (d(t)[t_i - c_i(t_i)] + c_i\mathbb{1}_{T_i^+}[d(t) - a_i(t)] - c_i\mathbb{1}_{T_i^-}[d(t) + a_i(t)]\right] \\
\leqslant & \mathbb{E}_t\left[\sum_i (d(t)[t_i - c_i(t_i)] + c_i\mathbb{1}_{T_i^+}[d(t)(1 - a_i(t))] - c_i\mathbb{1}_{T_i^-}[d(t)(1 - a_i(t)) + a_i(t)]\right] && (7) \\
\leqslant & \mathbb{E}_t\left[\sum_i (d(t)[t_i - c_i(t_i)] + c_i\mathbb{1}_{T_i^+}\inf_{t_i'\in T_i^+}\mathbb{E}_{t_{-i}}[d(t_i',t_{-i})] - c_i\mathbb{1}_{T_i^-}\sup_{t_i'\in T_i^-}\mathbb{E}_{t_{-i}}[d(t_i',t_{-i})]\right] && (8) \\
= & V_R(d).
\end{aligned}$$

The first inequality is obtained by multiplying $a_i(t)$ with $d(t)$ when $t_i \in T_i^+$ and multiplying $d(t)$ with $1 - a_i(t)$ when $t_i \in T_i^-$, and since we multiplied negative terms with terms that are less than or equal to one the first inequality (7) follows. The second inequality (8) 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 a verification rule a such that $V_P(d, a) = V_R(d)$. This implies that d is part of an optimal solution to problem (P).

We now describe an optimal solution to the relaxed problem.

Lemma 3. *Problem* (R) *is solved by a cutoff mechanism.*

Proof. We assume here that T is finite. We extend this proof in Appendix A.2 via an approximation argument to infinite type spaces.

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

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

and consider the following auxiliary maximization problem:

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

Suppose $\{\phi_i^+, \phi_i^-\}_i$ is such that there exists a decision rule which satisfies the constraints in (Aux) as strict inequalities.[7] Clearly, d^* also solves this problem. Let $\alpha_i^+ = \inf_{\alpha \in T_i^+}\{\alpha | \mathbb{E}_{t_{-i}}[d^*(\alpha, t_{-i})] > \varphi_i^+\}$ and $\alpha_i^- = \sup_{\alpha \in T_i^-}\{\alpha | \mathbb{E}_{t_{-i}}[d^*(\alpha, t_{-i})] < \varphi_i^-\}$. The Kuhn-Tucker theorem (see page 217 in Luenberger 1969) implies that there exist Lagrange multipliers $\lambda_i^*(t_i)$, such that d maximizes

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

We can assume that the multipliers λ^* are such that there are constants b_i^+ and b_i^- such that

$$t_i - c_i + \frac{\lambda_i^*(t_i)}{f_i(t_i)} = b_i^+$$

for $t_i \in T_i^+$ such that $t_i < \alpha_i^+$ and

$$t_i + c_i + \frac{\lambda_i^*(t_i)}{f_i(t_i)} = b_i^-$$

for $t_i \in T_i^-$ such that $t_i > \alpha_i^-$. If this were not the case, every solution d that maximizes $\tilde{\mathcal{L}}(\cdot, \lambda^*)$ would have either $\mathbb{E}_{t_{-i}}[d(t_i, t_{-i})] > \varphi_i^+$ or $\mathbb{E}_{t_{-i}}[d(t_i, t_{-i})] < \varphi_i^+$ for some $t_i \in T_i^+$ such that $t_i < \alpha_i^+$. Hence, it is either infeasible or it contradicts the definition of α_i^+. Analogous arguments apply for the second equation.

[7] If this condition is not satisfied, consider the sequence of problems where $\{\phi_i^+, \phi_i^-\}_i$ is replaced by $\{\phi_i^+ - \frac{1}{m}, \phi_i^- + \frac{1}{m}\}_i$ for $m = 1, 2, \ldots$. These problems satisfy the above assumption. Taking $m \to \infty$, the limit of a convergent subsequence of solutions is of the form claimed in Lemma 3.

Moreover, we obtain $\lambda_i^*(t_i) = 0$ for $t_i \in T_i^+$ such that $t_i \geqslant \alpha_i^+$ and for $t_i \in T_i^-$ such that $t_i \leqslant \alpha_i^-$. Indeed, complementary slackness implies $\lambda_i^*(\alpha_i^+) = 0$. Moreover, for every $t_i \in T_i^+$ such that $t_i > \alpha_i^+$, $t_i - c_i \geqslant \alpha_i^+ - c_i$ implies for every optimal solution to the Lagrangian d that $\mathbb{E}_{t_{-i}}[d(t_i, t_{-i})] \geqslant \mathbb{E}_{t_{-i}}[d(\alpha_i^+, t_{-i})] > \varphi_i^+$, which implies $\lambda_i^*(t_i) = 0$ again by complementary slackness. Analogous arguments for $t_i \in T_i^-$ such that $t_i \leqslant \alpha_i^-$ apply.

Feasibility implies that

$$t_i - c_i + \frac{\lambda_i^*(t_i)}{f_i(t_i)} \geqslant b_i^+$$

for all $t_i \in T_i^+$ and

$$t_i + c_i + \frac{\lambda_i^*(t_i)}{f_i(t_i)} \leqslant b_i^-$$

for all $t_i \in T_i^-$. Since $\lambda_i^*(t_i) \geqslant 0$ for all $t_i \in T_i^+$, we can take wlog $b_i^+ = \alpha_i^+ - c_i$. Similarly, since $\lambda_i^*(t_i) \leqslant 0$ for all $t_i \in T_i^-$, we can take wlog $b_i^- = \alpha_i^- + c_i$.

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

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

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

Since d^* maximizes the Lagrangian by assumption, we conclude that it is a cutoff mechanism. □

Now we have all the parts required to establish our main result Theorem 1 that cutoff 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 (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 $V_P(d^*, a^*) = V_R(d^*)$. □

6 Additional equivalence and relation to BDL

In this section we compare and expand on how our paper is related to BDL. The main result in BDL is a characterization of the optimal mechanisms. We compare these mechanisms to the ones that are optimal in our model, and we also discuss briefly an alternative proof to the main step in finding the optimal mechanisms in BDL. In subsection A.4 of the Appendix we describe the formal model in BDL and include our alternative proof.

BDL consider a situation with a principal who wants to allocate one indivisible private good among a group of agents without using money. Each agent has private information regarding the value the principal receives if the good is assigned to him. The principal does not know this value, but can learn it at a given cost. If the principal checks an agent he learns the agent's type perfectly. All agents strictly prefer to receive the object. The principal's objective is to maximize the expected payoff from assigning the object minus the expected cost of verification. The optimal mechanism, i.e., the mechanism that maximizes the expected utility of the principal, is a favored-agent mechanism: the principal chooses one single threshold and a favored agent. If no agent other than the favored agent reports a net type[8] above the threshold, then the object is allocated to the favored agent and no one is verified. If at least one agent different from the favored agent reports a net type above the threshold, the agent with the highest reported net type is verified and obtains the object if he did not lie.

Theorem 4 (Theorem 1 in BDL)**.** *A favored-agent mechanism is optimal. Moreover, every optimal mechanism is essentially a randomization over favored-agent mechanisms.*

The crucial step in BDL to prove Theorem 4 is to establish the optimality of a class of simple mechanisms, called threshold mechanisms.[9] We provide an alternative proof for the optimality of threshold mechanisms. Our proof makes a connection to the literature on reduced form auctions.[10] In the private good environment—the case considered by BDL—the set of feasible reduced form auctions has an explicit description (Border 1991) and a nice combinatorial structure (see e. g. Che, Kim and Mierendorff 2013). A brief sketch of our proof of optimality of the threshold mechanisms is as follows. First we observe that the relevant incentive constraints in the relaxed optimization problem are formulated in terms of reduced form auctions. Thus, we can restate the optimization problem using only reduced forms, and optimize over them instead of ex-post rules. The class of feasible reduced forms rules are readily available due to Border's characterization (Border 1991), and we can show that threshold mechanisms are optimal. Our approach to optimize directly over reduced forms, instead of the more complicated

[8]The net type for agent i with type t_i and verification cost c_i is $t_i - c_i$.

[9]See Section A.4 of the Appendix for the definition of a threshold mechanism and a complete statement of the theorem.

[10]A *reduced form auction* maps the type of an agent into the expected probability of being allocated the object.

ex-post rules, is not viable in our collective choice model. There cannot exist a tractable description of the reduced forms for the model we consider (Gopalan, Nisan and Roughgarden 2015). We had to use other tools and methods for showing that a cutoff mechanism is optimal in the collective choice environment.

Let us now look closer at the cutoff mechanism and the favored-agent mechanism to see which properties of the optimal mechanisms are robust across the two models. Note first that types are replaced by net types in both models: the principal accounts for the costs he incurs in the verification step. This creates generally an inefficiency in our model. In BDL however, the allocation is always efficient if all agents have the same costs of verification and at least one agent reports above the threshold. We conclude that a robust feature of the optimal mechanism is that net types are used to determine the outcome, but that this has different implications in different models.

The second robust feature is that both optimal mechanisms bunch certain types. In BDL, types are bunched as long as they are not too informative to the principal. All types below the threshold are close to each other, and for this reason it does not pay off to separate these types. In our model, the types below the threshold can be very different and therefore have a large impact on the utility of the principal. Instead, the incentive constraints dictate that it is very costly in terms of verifications to separate these types. Another difference between the two optimal mechanisms is that in the favored-agent mechanism there is only one threshold, whereas in the cutoff mechanism individual specific thresholds are optimal in general.

Finally, BDL note that a favored-agent mechanism can be implemented in dominant strategies. The observation that the optimal Bayesian incentive compatible mechanism is dominant strategy incentive compatible (DIC) does not hold in our model, as Example 2 shows. The reason is that in a collective choice setting without private goods there is no possibility to punish an agent without affecting the other agents. It is therefore not possible to induce truth-telling independent of what strategy is used by the others. However, we have seen that the optimal mechanism in our model is EPIC and that this follows from a general BIC-EPIC equivalence. As argued in Section 4, this equivalence can be extended to BDL's setting.

Theorem 5. *In the setting of BDL, there exists for any BIC mechanism an equivalent DIC mechanism.*

7 Conclusion

We have analyzed a collective decision model with costly verification where 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 he takes the decision. We have shown that a cutoff mechanism is optimal for the principal. The cutoff 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. In future work, we plan to model a version of imperfect verification and to study a model with limited commitment.

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. We will show this 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 equlibrium.

Let $(M_1, ..., M_I, \tilde{x}, \tilde{y})$ be an indirect mechanism, and $M = \bigtimes_{i \in \mathcal{I}} M_i$, 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.

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 $\ell_i : T \times T_i \times [0,1] \to \{0,1\}$.

[11] 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.

[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. This implies that a mediator can simulate the strategies in a direct mechanism.

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

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

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 $x(t',t',s) = x(t',t,s)$. Therefore, the decision is the same in both formulations if one agent deviates. Since truth-telling is an equilibrium in the mechanism (x,y), it is an equilibrium in the mechanism (d,a,ℓ), which consequently implements the same decision and verification rules.

A.2 Omitted proofs from Section 5

Proof of Lemma 3 for infinite type spaces.
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 $i \in \mathcal{I}$, $n \geqslant 1$, $l_i = 1,\dots,2^{n+1}$, let

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

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 $\alpha_i^{+n}, \alpha_i^{-n}$:

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

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

Let $\alpha_i^+ := \lim \alpha_i^{+n}$ and $\alpha_i^- := \lim \alpha_i^{-n}$ (by potentially choosing a convergent subsequence). Define

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

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

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 cutoff mechanism solves problem (R). □

A.3 Omitted proofs from Section 4

Proof of Theorem 2.
We first construct a discrete approximation of the type space: For $i \in \mathcal{I}$, $n \geqslant 1$, $l_i = 1, ..., 2^n$, let $S_i(n, l_i) = [F_i^{-1}((l_i - 1)2^{-n}), F_i^{-1}(l_i 2^{-n}))$, which form partitions of T_i such that each partition element has the same likelihood, 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.

Define an averaged function $g(n, l) = 2^{In} \int_{S(n,l)} g(t) dF$, which can be viewed as an I-dimensional tensor. Potentially after a relabeling the partition elements, the marginals of $g(n, l)$ are nondecreasing in l. By Theorem 6 in Gutmann et al. (1991) there exists a 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 $t \in S(n, l)$.

Note that g'_n is nondecreasing and hence satisfies

$$\int \operatorname*{ess\,inf}_{t_i \in B} g'_n(t_i, t_{-i}) dF_{-i} = \operatorname*{ess\,inf}_{t_i \in B} \int g'_n(t_i, t_{-i}) dF_{-i} \tag{9}$$

$$\int \operatorname*{ess\,sup}_{t_i \in B} g'_n(t_i, t_{-i}) dF_{-i} = \operatorname*{ess\,sup}_{t_i \in B} \int g'_n(t_i, t_{-i}) dF_{-i}. \tag{10}$$

Moreover,

$$\int_{S_i(n,l_i)} \int_{T_{-i}} g(t_i, t_{-i}) dF_{-i}\ dF_i = \int_{S_i(n,l_i)} \int_{T_{-i}} g_n'(t_i, t_{-i}) dF_{-i}\ dF_i, \tag{11}$$

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

Draw a weak*-convergent subsequence of g_n' (which is possible by Alaoglu's theorem) and denote its limit by $\hat{g}$. This allocation rule satisfies $0 \leqslant \hat{g} \leqslant 1$ and its marginals equal almost everywhere the marginals of g because of (11).

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, $\operatorname{ess\,inf}_{t_i \in B} \int_{T_{-i}} g_n'(t_i, t_{-i}) dF_{-i} \to \operatorname{ess\,inf}_{t_i \in B} \int_{T_{-i}} \hat{g}(t_i, t_{-i}) dF_{-i}$. Note 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. Suppose now that for some i,

$$\int \operatorname*{ess\,inf}_{t_i \in B} \hat{g}(t_i, t_{-i}) dF_{-i} < \operatorname*{ess\,inf}_{t_i \in B} \int \hat{g}(t_i, t_{-i}) dF_{-i}.$$[13]

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 (9) and thereby proofing 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

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

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

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

[13] 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}:

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

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

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

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

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 the optimality of the threshold mechanisms in BDL

In this section of the Appendix we give an alternative proof for the main result in Ben-Porath et al. (2014), a characterization of the optimal mechanism.

Statement of the problem:

The principal wants to allocate an indivisible object among the agents in $\mathcal{I} \equiv \{1, ..., n\}$. Agents are privately informed about their types $t_i \in \mathcal{T}_i \equiv [\underline{t}_i, \bar{t}_i]$. The principal receives value t_i when the object is allocated to an agent with type t_i. Monetary transfers are not possible, and all agents strictly prefer to receive the object. Types are independently distributed with distribution function F_i and type profiles are denoted by $t \in \mathcal{T}$.

The principal can verify agent i at a given cost of c_i, in which case the type of agent i is perfectly revealed. By invoking a revelation principle, it is enough to consider direct mechanisms. Denote by $p_i : \mathcal{T} \to [0,1]$ the total probability i is assigned the good and by $q_i : \mathcal{T} \to [0,1]$ the probability i is assigned the good and verified.

To solve this problem, BDL consider the following relaxed problem where $\{\varphi_i\}_{i\in\mathcal{I}}$ is taken as given.[14]

$$\begin{aligned} &\max_{0\leqslant p_i} \mathbb{E}_t \left[\textstyle\sum_i p_i(t)t_i\right] && && \text{(R1)} \\ &\text{s.t. } \mathbb{E}_{t_{-i}} p_i(t_i, t_{-i}) \geqslant \varphi_i && \forall\ t_i \in \mathcal{T}_i,\ i \in \mathcal{I} \\ &\qquad \textstyle\sum_i p_i(t) \leqslant 1 && \forall\ t \in \mathcal{T} \end{aligned}$$

Hence, the problem is to find feasible ex-post allocation rules $p_i : \mathcal{T} \to [0,1]$ maximizing the expected value to the principal, subject to an interim incentive constraint that restricts the interim expected value of $p(t_i, t_{-i})$. Problems of this kind are ubiquitous in auction theory and a useful approach is to instead search for optimal reduced forms. Denoting by $\hat{p}_i(t_i) \equiv \mathbb{E}_{t_{-i}} p_i(t_i, t_{-i})$, the objective function and the incentive constraint can be rewritten in terms of $\hat{p}_i$. In addition, one needs to ensure that the optimal $\hat{p}_i$'s are *feasible* in the sense that they can be implemented by a feasible ex-post allocation rule (i.e., p_i's that satisfy $p_i(t) \geqslant 0$ and $\sum_i p_i(t) \leqslant 1$).

The set of feasible reduced forms has been characterized by Border (1991) (for the most general treatment, see Che et al. 2013). In particular, every feasible reduced form must satisfy, for all $(\alpha_1, ..., \alpha_n) \in \mathcal{T}$,

$$\sum_i \int_{\alpha_i}^{\bar{t}_i} \hat{p}_i(t_i) f_i(t_i) dt_i \leqslant 1 - \prod_i F_i(\alpha_i). \tag{Border}$$

This condition is necessary for a reduced form to be feasible: The left-hand side, denoting the probability that an agent i with type above α_i wins the object, must clearly be lower than the probability that there is an agent i with type above α_i, which is written on the right-hand side. This direction is what we use in the proof below. The content of Border's theorem is to show that the above condition is sufficient for a non-decreasing reduced form to be implementable.

We restate (R1) as follows:

$$\begin{aligned} &\max_{0\leqslant \hat{p}_i} \sum_i \mathbb{E}_{t_i} \left[\hat{p}_i(t_i)t_i\right] \\ &\text{s.t. } \hat{p}_i(t_i) \geqslant \varphi_i && \forall\ t_i \in \mathcal{T}_i,\ i \in \mathcal{I} \\ &\qquad \hat{p}_i \text{ is feasible} \end{aligned}$$

Following BDL, we define a threshold mechanism with threshold α to be a mechanism p with the following reduced form: $\hat{p}_i(t_i) = \prod_{j\neq i} F_j(t_i)$ for $t_i > \alpha$ and $\hat{p}_i(t_i) = \varphi_i$ otherwise. Let

[14] In BDL the objective is $\sum_i \mathbb{E}_t[p_i(t)(t_i - c_i)]$. By interpreting t_i as $t_i - c_i$ we have an equivalent formulation of the problem.

$$\alpha^* = \inf\{\alpha \in \mathbb{R}_+ | \sum_i \varphi_i F_i(\alpha) \leqslant \prod_i F_i(\alpha) \text{ and } F_i(\alpha) > 0 \text{ for all } i\}$$

and denote by p^* the threshold mechanism with threshold α^*.[15]

Theorem 6 (Ben-Porath et al. (2014))**.** *The threshold mechanism p^* is the essentially unique solution to problem* (R1)*.*

We show in Step 1 that the reduced form $\hat{p}^*$ is feasible, in Step 2 that it is optimal, and finally in Step 3 that every optimal reduced form must equal $\hat{p}^*$ almost everywhere.

Proof.
Step 1: Feasibility
We will first construct a feasible ex-post rule inducing the interim rule $\hat{p}^*$ and then show that $\hat{p}^*$ satisfies the incentive constraints.

Consider the following ex-post rule p^*. It allocates the object to the agent with the highest type whenever $t_j > \alpha^*$ for some j, and whenever $t_j \leqslant \alpha^*$ for all j it is defined by $p_i^*(t) = \frac{\varphi_i}{\prod_{j \neq i} F_j(\alpha^*)}$.[16] This rule induces the interim rule $\hat{p}_i^*$. Moreover, it is clearly feasible if $t_j > \alpha^*$ for some j. Assuming $t_j \leqslant \alpha^*$ for all j and summing over all agents, we have that $\sum_i p_i^*(t) = \sum_i \frac{\varphi_i}{\prod_{j \neq i} F_j(\alpha^*)}$. By definition of α^* and continuity of the $F_j(\cdot)$'s, $\sum_i \frac{\varphi_i}{\prod_{j \neq i} F_j(\alpha^*)} \leqslant 1$. Thus, p^* is a feasible ex-post rule.

Regarding the incentive constraints, $\hat{p}_i^*(t_i) = \varphi_i$ for all $t_i \leqslant \alpha^*$. Suppose now $t_i > \alpha^*$: By definition of α^*, $F_i(t_i) > 0$ and we obtain $\hat{p}_i^*(t_i) = \frac{\prod_j F_j(t_i)}{F_i(t_i)}$. Again by definition of α^*, $\prod_j F_j(t_i) \geqslant \sum_j \varphi_j F_j(t_i)$. Thus,

$$\hat{p}_i^*(t_i) = \frac{\prod_j F_j(t_i)}{F_i(t_i)} \geqslant \frac{\sum_j \varphi_j F_j(t_i)}{F_i(t_i)} \geqslant \varphi_i.$$

Hence, $\hat{p}^*$ is a feasible solution to (R1).

Step 2: Optimality
We first establish an upper bound for the objective function and then show that the reduced form $\hat{p}^*$ achieves this upper bound.

Let $\tilde{p}_i$ be any feasible reduced form, which therefore satisfies the Border conditions for all $\alpha \in \mathbb{R}$:

$$\sum_i \int_\alpha^{\bar{t}_i} f_i(t_i)\tilde{p}_i(t_i)dt_i \leqslant 1 - \prod_i F_i(\alpha). \tag{12}$$

Since $\tilde{p}_i \geqslant \varphi_i$, the Border conditions also imply that for all α,

$$\sum_i \int_{\underline{t}_i}^{\alpha} \varphi_i f_i(t_i)dt_i + \sum_i \int_\alpha^{\bar{t}_i} f_i(t_i)\tilde{p}_i(t_i)dt_i \leqslant 1,$$

[15]Given $\sum_i \varphi_i \leqslant 1$, the constraint set is nonempty and hence α^* is well-defined.
[16]If $F_j(\alpha^*) = 0$ for some j, the ex-post rule always allocates to the agent with the highest type.

or, equivalently,

$$\sum_i \int_\alpha^{\bar{t}_i} f_i(t_i)\tilde{p}_i(t_i)dt_i \leqslant 1 - \sum_i \varphi_i F_i(\alpha). \tag{13}$$

Note that if $\underline{t}_i < 0$, $\tilde{p}_i(t_i) \geqslant \varphi_i$ implies $\int_{\underline{t}_i}^0 f_i(t_i)\tilde{p}_i(t_i)t_i\ dt_i \leqslant \int_{\underline{t}_i}^0 f_i(t_i)\varphi_i t_i\ dt_i$. Moreover, denoting $\bar{t} = \max_i\{\bar{t}_i\}$ we get:

$$\begin{aligned}
&\sum_i \int_0^{\bar{t}_i} f_i(t_i)\ \tilde{p}_i(t_i)\ t_i\ dt_i \\
=&\sum_i t_i \int_0^{t_i} f_i(s)\ \tilde{p}_i(s)\ ds\Big|_{t_i=0}^{\bar{t}_i} - \sum_i \int_0^{\bar{t}_i}\int_0^{t_i} f_i(s)\ \tilde{p}_i(s)\ ds\ dt_i \\
=&\int_0^{\bar{t}} \sum_i \int_\alpha^{\bar{t}_i} f_i(s)\ \tilde{p}_i(s)\ ds\ d\alpha \\
\leqslant&\int_0^{\alpha^*} [1 - \sum_i \varphi_i F_i(\alpha)]d\alpha + \int_{\alpha^*}^{\bar{t}} [1 - \prod_i F_i(\alpha)]d\alpha,
\end{aligned}$$

where the first equality follows from integration by parts, the second by rearranging terms and the inequality follows from (12) and (13).

We claim that $\hat{p}^*$ satisfies the above inequalities as equalities:
First, for $\alpha \geqslant \alpha^*$, $\sum_i \int_\alpha^{\bar{t}_i} f_i(s)\ \hat{p}_i^*(s)\ ds = \sum_i \int_\alpha^{\bar{t}_i} f_i(s) \prod_{j\neq i} F_j(s)\ ds = 1 - \prod_i F_i(\alpha)$. Moreover, for $\alpha < \alpha^*$,

$$\begin{aligned}
&\sum_i \int_\alpha^{\bar{t}_i} f_i(s)\ \hat{p}_i^*(s)\ ds = \sum_i \int_\alpha^{\alpha^*} f_i(s)\ \varphi_i\ ds + 1 - \prod_i F_i(\alpha^*) \\
=&\sum_i \varphi_i[F_i(\alpha^*) - F_i(\alpha)] + 1 - \prod_i F_i(\alpha^*) = 1 - \sum_i \varphi_i F_i(\alpha)
\end{aligned}$$

since, by definition of α^*, $\sum_i \varphi_i F_i(\alpha^*) = \prod_i F_i(\alpha^*)$. Therefore $\hat{p}_i^*$ is an optimal solution.

Step 3: Uniqueness
Note that any feasible reduced form $\tilde{p}$ satisfies the following inequality:

$$G(\alpha_1, ..., \alpha_n) := \sum_i \int_{\alpha_i}^{\bar{t}_i} f_i(s)\tilde{p}_i(s)ds \leqslant 1 - \prod_i F_i(\alpha_i) =: H(\alpha_1, ..., \alpha_n).$$

Since G is monotone, it is differentiable almost everywhere, and H is differentiable by assumption. For any optimal reduced form, the above arguments imply, for almost every $\alpha \geqslant \alpha^*$, that $G(\alpha, ..., \alpha) = H(\alpha, ..., \alpha)$ and that G and H are differentiable in α_i for all i at $(\alpha, ..., \alpha)$. Since H is an upper bound for G, this implies that their derivatives must coincide at $(\alpha, ..., \alpha)$:

$$-\tilde{p}_i(\alpha)f_i(\alpha) = -\prod_{j\neq i} F_j(\alpha)f_i(\alpha).$$

Moreover, by (12) and $\tilde{p}_i \geqslant \varphi_i$, $\tilde{p}_i(t_i) = \varphi_i$ for $t_i < \alpha^*$. We conclude that $\tilde{p}$ equals $\hat{p}^*$ almost everywhere. □

References

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. 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. (1991). Implementation of reduced form auctions: A geometric approach, *Econometrica* **59**: 1175–1187.

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.

Che, Y.-K., Kim, J. and Mierendorff, K. (2013). Generalized reduced-form auctions: A network-flow approach, *Econometrica* **81**: 2487–2520.

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. (2014). Optimal voting rules, *Working paper* .

Gopalan, P., Nisan, N. and Roughgarden, T. (2015). Public projects, boolean functions, and the borders of border's theorem, *Working paper* .

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

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.

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. (2014). Mechanism design with ex-post verification and limited punishments, *Working paper* .

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.

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.