Essays in Economic Theory

Inaugural-Dissertation

zur Erlangung des Grades eines Doktors
der Wirtschafts- und Gesellschaftswissenschaften

durch die

Rechts- und Staatswissenschaftliche Fakultät
der Rheinischen Friedrich-Wilhelms-Universität
Bonn

vorgelegt von

Andreas Kleiner

aus Sigmaringen

Bonn 2016

Dekan:	Prof. Dr. Rainer Hüttemann
Erstreferent:	Prof. Dr. Benny Moldovanu
Zweitreferent:	Prof. Dr. Daniel Krähmer

Tag der mündlichen Prüfung: 15. Juli 2016

Diese Dissertation ist auf dem Hochschulschriftenserver der ULB Bonn
(http://hss.ulb.uni-bonn.de/diss_online) elektronisch publiziert.

Acknowledgements

I am very grateful to my supervisors, Benny Moldovanu and Daniel Krähmer, for their extraordinary support during the past years. For comments and many joyful discussions I am particularly thankful to Moritz Drexl, Tymon Tatur, Sven Rady, Deszö Szalay, Albin Erlanson, Frank Rosar, and Cedric Wasser. My cohort made my studies a delightful experience. Finally, I am greatly indebted to my family for their continuing support.

I received material support from the Bonn Graduate School of Economics, the DFG (SFB-TR 15), and the Hausdorff Center for Mathematics.

Contents

Introduction

This thesis consists of five chapters on topics in mechanism design. In Chapters 1 to 4, we consider collective decision problems, where groups of agents decide jointly which alternative is to be implemented. In Chapter 5, we study the allocation of a private good. Moritz Drexl coauthored Chapters 1, 2, and 5, Chapter 3 is joint work with Benny Moldovanu, and Chapter 4 is joint work with Albin Erlanson.

The general problem we study is how to make joint decisions if the optimal decision depends on information that is privately held by strategic agents. Specifically, we consider agents that are privately informed about their preferences. A designer chooses a social choice function, a function mapping the preferences of the agents into social outcomes, to maximize some objective function. A prime example is a designer acting as a social planner, who maximizes utilitarian welfare (that is, the expected aggregate utility). Because preferences are private information, agents might misreport their preferences if simply asked for the information, and the designer has to be careful to provide the right incentives. She does so by choosing a mechanism, for example an auction format or a voting procedure, which induces a game. If agents play an equilibrium of this game, their preferences are mapped into social outcomes, thereby inducing an incentive-compatible social choice function. We study which mechanisms the designer should choose to achieve her objectives.

In Chapter 1, we study a group of agents deciding whether to implement a given reform or to keep status quo. Each agent is privately informed about his willingness-to-pay for the reform, and the designer wants to maximize utilitarian welfare. The first-best decision rule implements the reform whenever the average willingness-to-pay is positive. However, simply using this decision rule is not incentive-compatible, because any strategic agent would overstate his preference intensity to get his preferred decision. Instead, simple voting rules are commonly used, where each agent votes “yes” or “no” and the reform is implemented if there are sufficiently many “yes”-votes. Voting rules are criticized for being inefficient as the median willingness-to-pay decides, whereas using the average willingness-to-pay would yield better decisions. We ask whether one can design better decision rules, potentially using transfers, that induce robust incentives for the agents to tell the truth. We show that preference intensities can only be elicited if agents have to make transfers. Moreover, we prove that these transfers cannot be given to other members of the group without distorting incentives. The designer therefore faces a trade-off: she can use a "good" decision rule that takes detailed information into account, but then agents have to make transfers that leave the group. We solve for the decision rule that solves this trade-off optimally and find that qualified majority voting is optimal in a broad range of settings. This implies that although transfers could be used, it is optimal to rely on simple procedures that do not use transfers. Our results thereby shed new light on the inefficiency of voting: while voting indeed sometimes leads to inefficient decisions, in many circumstances there are simply no better procedures available.

We consider a dynamic version of the above framework in Chapter 2. Two agents face in each period a policy proposal that they can jointly accept or reject. While we do not allow for monetary transfers in this chapter, we consider general dynamic mechanisms. These mechanisms can be used to model institutions that allow, for example, the storage of votes, vote trading, or the use of budgeted veto rights. We argue that such dynamic mechanisms can, to some extent, be used to replicate monetary transfers by changing the expected continuation values of the agents. The flexibility of dynamic mechanisms thereby allows the construction of procedures that take cardinal preferences finely into account. However, we show that dynamic mechanisms can induce similar trade-offs as the ones we studied in Chapter 1: while the induced fine-tuning can improve the decision rule that is used in the current period, it requires distortions in future periods. We show that this detrimental effect often dominates the efficiency gains in the current period and that it is therefore optimal not to use the dynamic structure and instead to repeatedly employ a static voting procedure.

In Chapter 3, we study sequential, binary voting procedures. These procedures are commonly used to select one out of several alternatives by repeatedly making binary decisions until only one alternative remains, which is then formally elected. Such procedures are used, for example, in US Congress (the so-called amendment procedure) and in many European parliaments (in Germany, for example, the successive voting procedure). We analyze the incentives for agents with single-peaked preferences to strategically manipulate their votes, that is, to vote against their most preferred alternative. We characterize those voting procedures for which sincere voting is a robust equilibrium and for which there is consequently no space for manipulations. These procedures are easy to use for the voters and always select the Condorcet winner, both of which are desirable properties of voting procedures. For procedures that do not satisfy our characterization, we show via case studies that manipulations actually do occur and undesirable outcomes do arise in legislative decision-making.

We consider a model of costly verification in Chapter 4. In a collective decision problem without transfers, the principal can learn the agents' information using a verification technology, but it is costly for him to do so. When should the principal incur the cost to learn an agent's information in order to make better decisions? Although the principal could learn all the information to make the efficient decision, it would be excessively costly to do so. We show that it is better, and indeed optimal, to use a procedure we call voting-with-evidence. In this procedure, agents cast votes in favor or against a new proposal; in addition, they can provide evidence about information they have. If they do so, they get additional weight in the voting procedure, and their evidence will be verified whenever it is pivotal for the outcome. The new policy is implemented whenever there are sufficiently many weighted votes favoring it. This procedure achieves the optimal trade-off between making correct decisions on the one hand and saving costs of verification on the other. Additionally, we derive an equivalence between Bayesian and robust implementation for this model.

We study the allocation of a private good among two agents in Chapter 5. Agents are privately informed about their valuation for an object, and the designer is looking for a robustly incentive-compatible mechanism to maximize utilitarian welfare. We consider two settings: either the designer owns the object initially, or one of the agents owns the good initially (bilateral trade setting). We show that the optimal mechanism is either a posted price mechanism (one of the agents gets the object unless they agree to trade at a prespecified price) or an option mechanism (one of the agents gets the object, the other agent has the option to buy it at a prespecified price). Although mechanisms that lead to better allocations could be implemented, they would require money burning and are therefore inferior to the budget-balanced mechanisms that we find to be optimal. This trade-off and its solution relate this part to Chapters 1 and 2, where, in different settings and using different methods, we also found budget-balanced (respectively, static) mechanisms to be optimal.

Chapter 1

Why Voting? A Welfare Analysis

A committee decides collectively whether to accept a given proposal or to maintain the status quo. Committee members are privately informed about their valuations and monetary transfers are possible. According to which rule should the committee make its decision? We consider strategy-proof and anonymous social choice functions and solve for the decision rule that maximizes utilitarian welfare, which takes monetary transfers to an external agency explicitly into account. For regular distributions of preferences, we find that it is optimal to exclude monetary transfers and to decide by qualified majority voting. This sheds new light on the common objection that criticizes voting for its inefficiency.

1. Introduction

Majority voting is inefficient from a utilitarian perspective because the decision rule does not condition on preference intensities. Consider, for example, a municipality that decides whether to adopt a new law. Suppose a majority has a weak preference against the law, but there is a minority that strongly prefers the proposed law. Even if implementing the law maximizes utilitarian welfare, the law will be defeated if the decision is made by majority voting. Can this municipality benefit from a more complex decision rule that enables voters to signal their preference intensities, for example by allowing lobbying activities or even direct payments?

A specific suggestion is to implement the efficient decision rule (i. e. the one that always chooses the decision that is best from a utilitarian perspective) by using a Vickrey-Clarke-Groves (VCG) mechanism. However, if agents are privately informed this requires that agents make payments and it is well-known that these payments cannot balance the budget. Do agents prefer the efficient rule if they have to make payments in turn? More generally, we show that whenever a decision rule conditions on preference intensities money is lost, which introduces a trade-off for the agents: they can choose a "good" decision rule, but then they will lose money. We study this trade-off and solve for the decision rule that maximizes utilitarian welfare in a class of plausible decision rules. Majority voting turns out to be optimal in our model.

Our analysis follows standard models of collective decision making: A finite population of voters decides collectively whether to accept a given proposal or to maintain the status quo. Agents are privately informed about their valuations and have quasi-linear utilities. Monetary transfers are feasible as long as they create no budget deficit and as long as agents are willing to participate in the decision process. In contrast to much of the literature, we consider a utilitarian welfare function that takes monetary transfers to an external agency into account. We investigate which strategy-proof and anonymous social choice function maximizes expected utilitarian welfare. Strategy-proofness ensures that the rule can be robustly implemented, while anonymity seems to be a reasonable fairness requirement for a society of equals.

Our main result is that the optimal anonymous social choice function is implementable by qualified majority voting. Under such schemes, agents simply indicate whether they are in favor or against the proposal, and the proposal is accepted if the number of agents being in favor is above a predetermined threshold. This implies that, even though it is possible to use monetary transfers, it is optimal not to use them. Specifically, we show that any anonymous decision rule that relies on monetary transfers wastes money to such an extent that, for regular distributions of types, it is inferior to voting. In our model it is therefore not possible to improve upon voting without giving up reasonable properties of the social choice function.

Our finding that voting performs well from a welfare perspective stands in sharp contrast to parts of the previous literature, which suggest to implement the value-maximizing public decision. However, this does not achieve the first-best because it induces budget imbalances (see, e.g., **?**). While it is traditionally assumed that money wasting has no welfare effects, we consider a social planner that cares about aggregate transfers.[1] Our approach seems reasonable for at least two reasons: First, a utilitarian planner is interested in implementing the decision rule that maximizes the agents' expected utility. Since agents care about money, the planner in turn cares about aggregate transfers. Second, groups often choose the rule by which they decide themselves, and when making this choice they take the payments they have to make into account. Hence, our approach characterizes decision rules that are likely to prevail in practice.

Our derivation that transfer-free voting schemes dominate more complex decision rules can be summarized as follows. To prevent agents from overstating their preference intensities, one has to impose incentive payments whenever a decision rule conditions on preference intensities. Strategy-proof implementation severely restricts how these payments can be redistributed to the agents: the redistribution payment an agent receives must not depend on his own reported preference. For anonymous decision rules, we show that this restriction prevents any redistribution; hence, all incentive payments that are collected to induce truthful reports have to be wasted. This implies as a corollary that an anonymous social choice function is implementable with a balanced budget if and only if it can be implemented by qualified majority voting. The result that no money can be redistributed fixes the trade-off between increasing efficiency of the public decision and reducing the waste of monetary resources. For regular distribution

[1] An early exception is **?**, who argue that the budget imbalances of VCG mechanisms are not important because they vanish as the populations grow and hence are quantitatively negligible in many practical applications. We review their argument in Section 4. Note that the nature of the objective function would be irrelevant if we considered social choice functions that are implementable in Bayesian equilibrium, as the expected externality mechanism (**?**, **?**) achieves the first best in our setting.

functions, we show that this trade-off is solved optimally by not using money at all. This implies that the optimal social choice function is implementable by qualified majority voting. We characterize the minimum number of votes that is optimally required for the adoption of the proposal.

Related Literature

The literature evaluating public decision rules by the utilitarian criterion was initiated by **?**, who compares expected welfare of different voting rules and shows that simple majority voting (where a proposal is accepted if at least half of the population votes for it) is optimal if preferences are symmetric across outcomes. Recently, this approach was generalized using insights from the mechanism design literature to include more general decision rules (**?**), to allow for correlated valuations (**?**) and to consider environments with more than two alternatives (**?**). While we extend this approach to allow for monetary transfers, the resulting optimal decision rules relate our study to this literature.

Our results contribute to studies that try to explain why voting rules are used instead of mechanisms that rely on transfers: **?** argue that voting rules are unique in being robust and coalition-proof. **?** argue that voting rules are easy to implement and show that they approximate the efficient decision rule in large populations. In contrast, our results also apply for fixed finite populations; together with the limit results in **?** for VCG mechanisms our results also imply that voting approximates the efficient rule.

In independent and contemporaneous work, **?** analyze the optimal mechanism for the provision of a costly public good. They show that equal cost sharing mechanisms are optimal (in expectation) among dominant-strategy incentive compatible and feasible mechanisms that satisfy an additional kindness axiom. Our results are not logically related (anonymous mechanisms do not necessarily satisfy their kindness axiom, the pivot mechanism being a counterexample; a "kind" mechanism need not be anonymous on the other hand). However, our results are related in spirit: both papers show that if there is a trade-off between balancing the budget or having a more efficient decision rule, it is often preferable to balance the budget. We provide in addition results for general (and potentially correlated) type distributions and on ex-post dominance.

Our modeling approach is related to a small part of the literature, which evaluates allocation rules for the allocation of a private good according to an average efficiency criterion and considers money burning to be welfare reducing (**?**, **?**, **?**). An alternative criterion to evaluate allocation rules is to rank them in terms of their worst-case efficiency (see, for example, **?**, **?**, **?**)

The paper is structured as follows: We present the model in Section 2, derive our main result in Section 3 and discuss the role of the assumptions in Section 4.

2. Model

A population of n agents, $N = \{1, ..., n\}$, decides collectively on a binary outcome $X \in \{0, 1\}$. We interpret this as agents deciding whether they accept a proposal (in which case $X = 1$) or reject it and maintain the status quo ($X = 0$). Given a collective decision X, the utility of agent i is given by $\theta_i \cdot X + t_i$, where θ_i is the agent's valuation for the proposal and t_i is a transfer to agent i.[2] Each agent is privately informed about his valuation, which is drawn independently from a type space $\Theta := \left[\underline{\theta}, \overline{\theta}\right]$ according to a distribution function F with positive density f. To make the problem interesting we assume that $\underline{\theta} < 0 < \overline{\theta}$.[3] Both, type space and distribution function, are common knowledge. Let Θ^n denote the product type space consisting of complete type profiles with typical element $\theta = (\theta_i, \theta_{-i})$.

A social choice function in this setting determines for which preference profiles the proposal is accepted and which transfers are made to the agents. Formally, a *social choice function* is a pair (x, t) consisting of a *decision rule*

$$x \; : \; \Theta^n \; \rightarrow \; \{0, 1\}$$

and a *transfer rule*

$$t : \Theta^n \rightarrow \mathbb{R}^n$$

such that, for any realized preference profile θ, $x(\theta)$ is the decision on the public outcome and $t_i(\theta)$ is the transfer received by agent i.

Not all social choice functions can plausibly be implemented. In the following, we describe several restrictions on the social choice functions that we consider.

We require that social choice functions are *feasible* in the sense that for any realization of preferences no injection of money from an external agency is necessary, i.e.,

$$\sum_{i \in N} t_i(\theta) \leq 0. \tag{F}$$

Given that preferences are observed privately by the agents, a social choice function must induce the agents to report their types truthfully. We are interested in social choice functions that are *strategy-proof*, i.e., for which there exists a mechanism and an equilibrium in dominant strategies for the strategic game induced by this mechanism such that, for any realized type profile, the equilibrium outcome corresponds to the outcome that the social choice function stipulates. This is a strong equilibrium concept that ensures that the social choice function can be implemented irrespective of the exact information structures. Requiring social choice functions to be strategy-proof is a standard approach in social choice theory (see, e.g., **?**) and is equivalent to robust implementation in the spirit of **?** in our model (see **?**).

In many situations agents have the outside option to abstain from the decision process and leave the decision to the other agents. In this case, they cannot be forced to make payments that relate to the decision process. However, they are nonetheless affected by the public outcome, and the choice of the outcome if agent i does not par-

[2]Our analysis applies to costless projects as well as to costly projects with a given payment plan, in which case the valuation of agent i is interpreted as her net valuation taking her contribution into account. Also note that the analysis accommodates more general utility functions: Take any quasi-linear utility function such that the utility difference between $X = 1$ and $X = 0$ is continuous and strictly increasing in θ_i. Redefining the type to equal the utility difference, we can proceed with our analysis without change.

[3]The analysis directly extends to cases where $\underline{\theta} = -\infty$ and/or $\overline{\theta} = \infty$.

ticipate in the decision process, $\underline{x}_i(\theta_{-i}) \in \{0,1\}$, becomes part of the design problem.[4] Given the outside option of abstaining from the decision process, it is without loss of generality to consider social choice functions that ensure *universal participation* (see, e.g., **?**) in the sense that all agents prefer to participate in the decision process:

$$\theta_i x(\theta) + t_i(\theta) \geq \theta_i \underline{x}_i(\theta_{-i}). \tag{UP}$$

This constraint is weaker than the requirement that every agent derive utility of at least zero (often called individual rationality) and better suited for public good environments. For instance, majority voting and the Pivotal mechanism satisfy universal participation but in general are not individually rational.[5]

We also require social choice functions to be anonymous:

Definition 1. *We call a social choice function* (x,t) anonymous *if the decision rule is independent of the agents' identities, i.e. if, for each permutation* $\pi : N \to N$ *and corresponding function* $\hat{\pi}(\theta) = (\theta_{\pi(1)}, \ldots, \theta_{\pi(n)})$*, it holds that* $x(\theta) = x(\hat{\pi}(\theta))$ *for all* θ*.*

This is a weak notion of anonymity, requiring only that the names of the agents do not affect the public decision. One reason to impose anonymity is that many fairness concepts build on this assumption (e. g., equal treatment of equals). This requirement has a long tradition in social choice theory; see for example **?**.[6]

Throughout the paper we focus on feasible and strategy-proof social choice functions that are anonymous and satisfy universal participation.

3. Results

In this section, we characterize incentive-compatible mechanisms and derive an important auxiliary result about redistribution payments. Using this result, we show that simple majority voting is the ex-post dominant mechanism and show that qualified majority voting maximizes ex-ante expected welfare.

To implement a given social choice function, we invoke the revelation principle (**?**). It follows that we can focus without loss of generality on direct revelation mechanisms in which it is a dominant strategy for agents to report their valuations truthfully. Hence, we denote a mechanism simply by a tuple (x,t), where $x : \Theta^n \to \{0,1\}$ maps reported types into a collective decision and, for each agent i, $t_i : \Theta^n \to \mathbb{R}$ maps reported types

[4]It turns out that the participation constraint is binding only for agents with valuation 0. Therefore, the design of the outside option is irrelevant for our results.

[5]The unique voting rule that is individually rational is unanimity voting where the proposal is accepted if and only if all agents vote for the proposal. Our analysis implies that it is optimal to decide by unanimity voting if we impose individual rationality instead.

[6]Note that this assumption would be without loss of generality if we allowed for stochastic decision rules. Given any social choice function (x,t), apply this function after randomly permuting the agents. This defines a new social choice function $(\tilde{x},\tilde{t})$ that is anonymous and achieves the same utilitarian welfare. While this new rule treats all agents equally ex-ante, it is possible that agents with the same valuations are treated very differently after the uncertainty about the randomization is resolved. The important restriction by focusing on deterministic rules is to prevent that the anonymity requirement can by circumvented this way.

into the payment received by that agent. The requirement that a social choice function be strategy-proof translates to

$$\theta_i x(\theta_i, \theta_{-i}) + t_i(\theta_i, \theta_{-i}) \geq \theta_i x(\hat{\theta}_i, \theta_{-i}) + t_i(\hat{\theta}_i, \theta_{-i}) \tag{IC}$$

for all realized preferences θ_i and all reports $\hat{\theta}_i$ and θ_{-i}.

An indirect mechanism is called *qualified majority voting (with threshold k)*, if each agent has the message set {*yes, no*} and the proposal is implemented if and only if at least k agents send message *yes* and no monetary transfers are made, i.e., $t_i(\theta) = 0$ for all i and θ.

The following lemma is a standard characterization of strategy-proof mechanisms.

Lemma 1. *A mechanism is strategy-proof if and only if, for each agent i,*

1. *$x(\theta_i, \theta_{-i})$ is non-decreasing in θ_i for all θ_{-i} and*
2. *there exists a function $h_i(\theta_{-i})$, such that for all θ,*

$$t_i(\theta_i, \theta_{-i}) = -\theta_i x(\theta_i, \theta_{-i}) + \int_0^{\theta_i} x(\beta, \theta_{-i}) d\beta + h_i(\theta_{-i}). \tag{1}$$

We call the first two terms on the right-hand side the "incentive payment" and the last term "redistribution payment".

Equation (1) suggests the following definition:

Definition 2. *Agent i is* pivotal at profile θ, *if* $\theta_i x(\theta) \neq \int_0^{\theta_i} x(\beta, \theta_{-i}) d\beta$.

A necessary condition for agent i to be pivotal at θ is that $x(\theta) \neq x(0, \theta_{-i})$. If agent i is not pivotal at a given profile (θ_i, θ_{-i}) then her payment equals $h_i(\theta_{-i})$. If she is pivotal at this profile, her transfer is reduced by $\theta_i x(\theta) - \int_0^{\theta_i} x(\beta, \theta_{-i}) d\beta$.

The following lemma shows that redistribution payments are fixed for any anonymous and strategy-proof mechanism.

Lemma 2. *Suppose a strategy-proof mechanism (x, t) is anonymous and satisfies universal participation. Then $h_i(\theta_{-i}) = 0$ for all i and θ_{-i}.*

To obtain Lemma 2, first we use anonymity to argue that for any profile of reports θ_{-i}, there is a report by agent i such that no agent has to make an incentive payment. This implies that all redistribution payments are zero: Suppose for a contradiction that there is a report profile θ_{-i} such that agent i receives a strictly positive redistribution. From the previous step we know that there is a profile (θ_i, θ_{-i}) such that no agent makes an incentive payment, and since the redistribution payment of agent i does not depend on θ_i (Lemma 1), it is strictly positive at this profile. However, due to universal participation each redistribution payment has to be weakly positive and feasibility requires that their sum is weakly negative, which yields the desired contradiction.

Proof. The proof consists of two steps.

Step 1: For all i and θ_{-i}, there exists θ_i such that no agent is pivotal at (θ_i, θ_{-i}).

Note first that all agents that are pivotal at profile θ submit reports of the same sign: If $x(\theta) = 1$ then monotonicity implies that $x(0, \theta_{-i}) = 1$ for all agents i with $\theta_i < 0$ and hence only agents with positive reports can be pivotal (and similarly for $x(\theta) = 0$).

Fix an arbitrary agent i and a report profile $\theta_{-i} \in \Theta^{n-1}$. Suppose $x(0, \theta_{-i}) = 1$ and hence that all pivotal agents submit positive reports (if no agent is pivotal at this profile, we are done; if $x(0, \theta_{-i}) = 0$ analogous arguments hold). We show that no agent is pivotal at profile $\theta := (\theta_{j^*}, \theta_{-i})$, where $j^* \in \arg\max_j \theta_j$. Monotonicity implies that $x(\theta) = x(0, \theta_{-i}) = 1$ and hence agent i is not pivotal. Anonymity implies that agent j^* is not pivotal.

It remains to show that if j is not pivotal at θ and $\theta_{j'} \leq \theta_j$, then j' is not pivotal at θ. Assume to the contrary that j' is pivotal at θ; hence, $x(\theta) = 1$ and $x(\varepsilon, \theta_{-j'}) = 0$ for some $\varepsilon > 0$. If $\hat{\pi}_{j,j'} : \Theta^n \to \Theta^n$ is the function permuting the j-th and j'-th component, then $\hat{\pi}_{j,j'}[(\varepsilon, \theta_{-j})] \leq (\varepsilon, \theta_{-j'})$. From monotonicity it follows that $x\left(\hat{\pi}_{j,j'}[(\varepsilon, \theta_{-j})]\right) = 0$ and anonymity implies that $x(\varepsilon, \theta_{-j}) = 0$, contradicting the assumption that j is not pivotal at θ.

Step 2: For all i and θ_{-i} we have $h_i(\theta_{-i}) = 0$.

Universal participation immediately implies that an agent with valuation 0 gets a weakly positive utility: $0 \cdot x(0, \theta_{-i}) + t_i(0, \theta_{-i}) \geq 0 \cdot \underline{x}(\theta_{-i})$. From (1) it follows that $h_i(\theta_{-i}) \geq 0$ for all i, θ_{-i}. To obtain a contradiction, suppose that there exists an agent j and a report profile $\theta_{-j} \in \Theta^{n-1}$ such that $h_j(\theta_{-j}) > 0$. By step one, we can choose θ_j such that no agent is pivotal at $\theta := (\theta_j, \theta_{-j})$, implying by (1) that $\sum_i t_i(\theta) = \sum_i h_i(\theta_{-i}) > 0$, which contradicts feasibility. □

As an easy consequence, Lemma 1 and Lemma 2 permit a characterization of the set of strategy-proof social choice functions that have a balanced budget.

Corollary 1. *A feasible and anonymous social choice function satisfying universal participation has a balanced budget if and only if it is implementable by qualified majority voting.*

Proof. By Lemmas 1 and 2, the budget is balanced if and only if no agent is pivotal. This implies that the decision rule is constant in the interior of each orthant. Consequently, the decision rule can be implemented via qualified majority voting. □

A related result has been obtained by **?**, who in addition require weak Pareto efficiency but do not impose participation constraints.

Definition 3. *A mechanism (x, t)* dominates *another mechanism (x', t') if, for every θ, a majority of the agents prefers the outcome of mechanism (x, t) compared to the outcome of mechanism (x', t'). Formally, for all θ,*

$$\#\{i | \theta_i x(\theta) + t_i(\theta) \geq \theta_i x'(\theta) + t'_i(\theta)\} \geq \frac{n}{2}.$$

? showed that simple majority voting dominates the pivotal mechanism in a public good setting with logarithmic utilities and **?** obtained the same result for the model we study. Lemma 2 also allows us to extend this result to a much larger class of mechanisms.

Proposition 1. *Let (x, t) be any feasible and strategy-proof mechanism that is anonymous and satisfies universal participation. Then simple majority voting dominates (x, t).*

Proof. Lemma 2 implies that $t_i(\theta) \leq 0$ for all i and θ. Under simple majority voting, there is always a majority that gets its preferred alternative. These agents are weakly worse off under mechanism (x, t), because they make weakly positive payments and potentially get their less preferred alternative. $\square$

Utilitarian Social Planner

The above result takes an ex-post dominance perspective and therefore does not take preference intensities into account. Even though there will always be a majority that ex-post prefers the outcome of simple majority voting, this does not imply that agents would choose majority voting from an ex-ante viewpoint (because the minority preferring a different mechanism might have stronger preferences). Therefore we take an ex-ante perspective in this section and study a utilitarian planner who chooses a social choice function to maximize expected utilitarian welfare given by

$$U(x, t) := \mathbb{E}_\theta \left[\sum_{i=1}^{N} \Big[\theta_i x(\theta) + t_i(\theta) \Big] \right].$$

The expectation is taken with respect to the prior distribution of θ; hence, the planner uses prior information on the distribution of types to evaluate decision rules. Note that the specification of this information does not affect the incentives of the agents (as we focus on robust implementation), but only how different rules are compared.[7] A social choice function is *optimal* if it maximizes this expression.

We concentrate first on distribution functions that satisfy the following condition:

Definition 4. *A distribution function F has* increasing hazard rates *if $\frac{f(\theta_i)}{1-F(\theta_i)}$ is non-decreasing for $\theta_i \geq 0$ and $-\frac{f(\theta_i)}{F(\theta_i)}$ is non-decreasing for $\theta_i \leq 0$.*

This assumption is well-known from the literature on optimal auctions and procurement auction design; it is satisfied by many commonly employed distribution functions, for example by the uniform, (truncated) normal, and exponential distributions.

We are now ready to state our main result.

Theorem 1. *Suppose F has increasing hazard rates and consider the class of feasible and strategy-proof social choice functions that are anonymous and satisfy universal participation. The optimal social choice function in this class is implementable by qualified majority voting with threshold $\lceil k \rceil$, where*

$$k := \frac{-n \; \mathbb{E}[\theta_i | \, \theta_i \leq 0]}{\mathbb{E}[\theta_i | \, \theta_i \geq 0] - \mathbb{E}[\theta_i | \, \theta_i \leq 0]}.$$

[7] We relax the assumption that the planner perfectly knows the type distribution below.

That is, the optimal decision rule does not rely on monetary transfers at all and can be implemented using a simple indirect mechanism where each agent indicates whether she is in favor of or against the proposal. The proposal is accepted if more than $\lceil k \rceil$ voters are in favor.[8]

The following example illustrates how voting mechanisms compare to the first-best and the pivotal mechanism.

Example 1. *Let $n = 2$ and θ_i be independently and uniformly distributed on $[-3, 3]$ for $i = 1, 2$.*

If valuations were publicly observable, the first-best could be implemented which would yield welfare $U_{FB} = \frac{1}{2}\mathbb{E}[\theta_1 + \theta_2 \mid \theta_1 + \theta_2 \geq 0] = 1$.

The best mechanism that decides efficiently is the pivotal mechanism, where each agent pays the externality she creates on other agents. It gives a welfare of $U_{VCG} = \frac{1}{2}$ (see the Appendix).

Unanimity voting, that is, accepting the proposal if and only if both agents have a positive valuation, is an optimal voting rule (together with the voting rule that rejects the proposal if and only if both agents have a negative valuation). These rules yield welfare $U_{UV} = \frac{1}{4}\mathbb{E}[\theta_1 + \theta_2 \mid \theta_1 \geq 0, \theta_2 \geq 0] = \frac{3}{4}$.

Hence, the welfare loss due to private information is twice as large under the best VCG mechanism as compared to unanimity voting.

The role of the underlying assumptions is discussed in Section 4 and a formal proof for Theorem 1 is provided in the Appendix. In the following, we build some intuition for this result.

Lemma 2 shows that money cannot be redistributed in anonymous social choice functions, and hence there is a direct trade-off between improving the decision rule and reducing the outflow of money. We show that under increasing hazard rates, this conflict is resolved optimally in favor of no money burning. To gain some intuition, fix a type profile of the other agents, θ_{-i}. Strategy-proofness implies that there is a cutoff θ_i^* such that the proposal will be accepted if the type of agent i is above θ_i^*. To solve for the optimal decision rule we need to find the optimal cutoff. Assume that the sum of valuations $\sum_{j \neq i} \theta_j + \theta_i^*$ is negative. Marginally increasing the cutoff leads to a rejection of the proposal which in this case increases efficiency (with a positive effect on welfare proportional to $f(\theta_i^*)$). On the other hand, strategy-proofness implies that agents with a type above the cutoff make a payment equal to the cutoff. Increasing the cutoff increases these payments (with a corresponding negative effect on welfare proportional to $1 - F(\theta_i^*)$). Monotone hazard rates imply that if the positive effect outweighs the negative effect at θ_i^* and if it is therefore beneficial to marginally increase the cutoff, then it is optimal to set the cutoff to the highest possible value. Symmetric arguments imply that it is optimal to set all cutoffs either equal to zero or to the boundary of the type space, and hence that the optimal mechanism can be implemented by a voting rule.

The optimal number of votes required in favor of a proposal is given by the smallest integer number k such that the expected aggregate welfare of a proposal, given that

[8]This indirect implementation also alleviates the commitment problem of the planner: Given the information she obtains in this mechanism, the decision rule promised to the agents is optimal.

k out of n voters have a positive valuation, is positive. Hence, the optimal threshold required for qualified majority voting depends on the conditional expected values given that the valuation is either positive or negative. Simple majority voting is optimal if valuations are distributed symmetrically around 0. If, however, opponents of a proposal are expected to have a stronger preference intensity, then it is optimal to require a qualified majority that is larger than simple majority.

General distributions and correlated types

In this section we generalize our analysis in two directions. First, we allow for more general distribution functions, not only those having increasing hazard rates. Second, we relax the assumption that the planner knows perfectly the type distribution. If types are drawn independently conditionally on some unknown state of the world, but the distribution depends on the state of the world, this potentially creates correlated types from the planner's perspective. Therefore, we first state the general optimization problem allowing for correlated types. Lemma 2 still applies and shows that all redistribution payments are equal to zero. Hence, in analogy to Lemma 3, we can state the problem as

$$\max_{0 \le x \le 1} \int \Big[\sum_i \Psi(\theta_i|\theta_{-i}) \Big] x(\theta) dG(\theta)$$
$$\text{s. t. } x \text{ being point-wise non-decreasing,}$$

where G (g) denotes the joint cdf (pdf) and

$$\Psi(\theta_i|\theta_{-i}) = \begin{cases} -\frac{G(\theta_i|\theta_{-i})}{g(\theta_i|\theta_{-i})} & \text{if } \theta_i < 0 \\ \frac{1-G(\theta_i|\theta_{-i})}{g(\theta_i|\theta_{-i})} & \text{if } \theta_i \ge 0. \end{cases}$$

If types are independently distributed (that is, if $G(\theta) = \prod_i F(\theta_i)$), and if F has decreasing hazard rates, then the monotonicity constraint is not binding and consequently the efficient decision rule is optimal and can be implemented by the pivot mechanism. More generally, as long as types are independently distributed, a standard ironing procedure can be used to determine the optimal decision rule.

While standard procedures cannot be applied if types are not independently distributed, we can immediately deduce the optimal mechanism in two special cases. First, if the conditional hazard rates are decreasing ($-\frac{G(\theta_i|\theta_{-i})}{g(\theta_i|\theta_{-i})}$ and $\frac{1-G(\theta_i|\theta_{-i})}{g(\theta_i|\theta_{-i})}$ are non-decreasing for each θ_{-i}) and types are affiliated, then the pivot mechanism is optimal. Affiliation implies that an increase in θ_i increases $\Psi(\theta_j|\theta_{-j})$ for all j (see, for example, **?**), and hence that the monotonicity constraint is not binding. Analogously, one can show that if the conditional hazard rates are increasing and types are negatively affiliated, then a voting mechanism is optimal.

These results are not fully satisfying because both, negatively affiliated types and decreasing hazard rates, are strong assumptions. If types are positively affiliated and the conditional distributions have increasing hazard rates, then the optimal mechanism usually depends on the details of the distribution function. To gain additional insights, we analyze the problem that arises if the planner has imperfect information about the distribution from which types are drawn. Specifically, we assume that there are finitely many states of the world, $\omega \in \Omega$, and that types are drawn independently from a distribution function F_ω.

Corollary 2. *Suppose that F_ω has increasing hazard rates and that $\mathbb{E}_{F_\omega}[t_i|t_i > 0] = \mathbb{E}_{F_{\omega'}}[t_i|t_i > 0]$ and $\mathbb{E}_{F_\omega}[t_i|t_i < 0] = \mathbb{E}_{F_{\omega'}}[t_i|t_i < 0]$ for all $\omega, \omega' \in \Omega$. Then qualified majority voting is optimal among all feasible and strategy-proof social choice functions that are anonymous and satisfy universal participation.*

Under the assumptions made in the corollary, uncertainty about the state of the world only affects the expected number of supporters of status quo, but not the expected type of a supporter of status quo. Consequently, in each state of the world the same qualified majority rule is optimal and therefore it is optimal ex-ante.

Proposition 2. *Suppose there are two states of the world, ω_1 and ω_2, each occurring with strictly positive probability. Suppose that $\mathbb{E}_{F_{\omega_1}}[t_i] > 0 > \mathbb{E}_{F_{\omega_2}}[t_i]$, and that the expected number of supporters of status quo is the same in both states. Then, for large enough populations, the pivot mechanism achieves a higher expected welfare than any qualified majority voting rule.*

Proof. As the population grows, the aggregate payments in the pivot mechanism converge to 0, and consequently aggregate welfare converges to the first-best (see Theorem 6 in **?**). For a voting procedure to approach the first-best, it must implement reform with probability approaching 1 in state ω_1 and with probability approaching 0 in state ω_2. However, since the expected number of supporters of status quo is the same in both states, no voting procedure can differentiate between the two states. Consequently, if the population is large enough, the pivot mechanism outperforms any voting mechanism. □

If the state of the world influences the type distributions, but not the expected number of supporters of reform, the optimal majority requirement differs between the states of the world. Because the planner cannot choose the correct majority requirements, welfare under any voting procedure does not converge to first-best welfare. Because the welfare under the pivot mechanism converges to the first-best, it outperforms any voting rule for large populations.

4. Discussion

The fact that the efficient decision rule cannot be implemented with a balanced budget introduces a trade-off for a utilitarian planner: Should she choose a more efficient decision rule or one that requires less payments by the agents? We model this trade-off explicitly and solve for the welfare maximizing social choice function. For regular type distributions, it is optimal not to waste any monetary resources and to decide by majority voting.

The characterization of majority voting as the optimal social choice function relies on the specifics of our environment. In particular, the result that no money can be redistributed hinges on the anonymity requirement. If we relax this requirement (or allow for stochastic decision rules), majority voting is no longer optimal. The best budget-balanced social choice function for distributions that are symmetric around 0 is a "sampling Groves approach": pick a default agent, implement the efficient decision for the remaining agents, and award all incentive payments to the default agent (**?**).

Example 2 (cont.). *Let $n = 4$ and θ_i be independently and uniformly distributed on $[-3, 3]$.*

By Theorem 1, the optimal deterministic and anonymous mechanism is given by an optimal voting rule (which accepts the proposal if either at least 2 or at least 3 agents are in favor). This yields a welfare of $U_{MV} = \frac{36}{32}$.

A sampling Groves scheme would, for example, implement the decision that is jointly optimal for agents 1,2, and 3. All incentive payments that are collected from these agents would then be awarded to agent 4. This yields welfare $U_{sGroves} = \frac{39}{32}$. If stochastic mechanisms are allowed, the same welfare can be achieved via an anonymous mechanism, that permutes the names of the agents at random and then applies the mechanism described above.

In contrast to the anonymity (respectively, non-randomness) requirement, the participation constraint seems not to be a driving force of our results. Imposing universal participation simplifies the analysis and allows for the clear-cut result that no redistribution is possible for anonymous decision rules. Without it, a characterization of the optimal redistribution payments is hard; our numerical results suggest nonetheless that voting is often optimal even if one does not impose participation constraints.

Considering a richer set of possible alternatives, the results depend on the specification of agents' preferences. While in many cases similar trade-offs as in our model are present, our results do not extend in general; for example, in an environment with quadratic utilities and a continuum of alternatives, the efficient decision rule can be implemented with a balanced budget (**?**).

In an early contribution, **?** argued that implementing the pivotal mechanism is a welfare-superior way to decide on public projects, even if the payments that accrue in the decision process are wasted. They suggest that aggregate payments vanish as the number of agents gets large and argue that we should implement the efficient decision rule instead of relying on inefficient voting procedures. The same argument has more recently been put forward by **?**, who showed formally that aggregate payments in the pivot mechanism vanish and concluded that this mechanism produces higher welfare than voting. Our result contrasts with this suggestion: we show that voting can be welfare-superior to the pivot mechanism for any number of agents. More generally, we show that the problem of money wasting is not specific to the efficient decision rule, but is present in any anonymous decision rule that conditions on preference intensities. Our results thereby shed a new light on the widespread criticism that voting is inefficient: despite sometimes imposing the "wrong" decision, it can be optimal to ban monetary transfers and decide by majority voting.

Appendix

Verification of Example 1. Welfare of the pivot mechanism can be expressed as the difference between the welfare of the first-best and the transfers needed to implement the efficient decision:

$$U_{VCG} = U_{FB} - \frac{4}{36} \int_{-3}^{0} \int_{0}^{-\theta_1} (-\theta_2)\ d\theta_2\ d\theta_1 = \frac{1}{2}$$

Here, we used the fact that transfers are symmetric in the four regions $\{\theta \mid \theta_i \geq 0, \theta_j \leq 0, \theta_i + \theta_j \lesseqqgtr 0\}$ and zero everywhere else. $\square$

The following lemma shows how utilitarian welfare of a social choice function can be expressed as the sum of two terms. The first only depends on the allocation rule, and the second consists of the redistribution payments.

Lemma 3. *Let (x, t) be an incentive compatible direct mechanism for social choice rule $G = (X^G, T^G)$ and define*

$$\psi(\theta_i) = \begin{cases} \frac{-F(\theta_i)}{f(\theta_i)} & \text{if } \theta_i \leq 0, \\ \frac{1-F(\theta_i)}{f(\theta_i)} & \text{otherwise.} \end{cases} \tag{2}$$

Then we have

$$U(X^G, T^G) = \int_{\Theta^N} \left[\sum_{i \in N} \psi(\theta_i) \right] x(\theta) dF^N(\theta) + \sum_{i \in N} \int_{\Theta^{N-1}} h_i(\theta_{-i}) dF^{N-1}(\theta_{-i}).$$

Proof. Note that for all θ_{-i},

$$\begin{aligned}
&\int_{\underline{\theta}}^{\overline{\theta}} \left[\int_0^{\theta_i} x(\beta, \theta_{-i}) d\beta \right] f(\theta_i) d\theta_i \\
&= \left[\int_0^{\overline{\theta}} x(\beta, \theta_{-i}) d\beta \underbrace{F(\overline{\theta})}_{=1} - \int_0^{\underline{\theta}} x(\beta, \theta_{-i}) d\beta \underbrace{F(\underline{\theta})}_{=0} \right] - \int_{\underline{\theta}}^{\overline{\theta}} x(\theta_i, \theta_{-i}) F(\theta_i) d\theta_i \\
&= \int_0^{\overline{\theta}} \frac{1 - F(\theta_i)}{f(\theta_i)} x(\theta_i, \theta_{-i}) dF(\theta_i) + \int_{\underline{\theta}}^{0} \frac{-F(\theta_i)}{f(\theta_i)} x(\theta_i, \theta_{-i}) dF(\theta_i) \\
&= \int_{\underline{\theta}}^{\overline{\theta}} \psi(\theta_i)\ x(\theta_i, \theta_{-i}) dF(\theta_i), \qquad (3)
\end{aligned}$$

where the first equality follows from integrating by parts, the second from rearranging terms and the third from the definition of Ψ.

Now rewrite

$$\begin{aligned}
U(X^G, T^G) &= \int_{\Theta^N} \sum_{i \in N} \Big[\theta_i x(\theta) + t_i(\theta) \Big] dF^N(\theta) \\
&= \sum_{i \in N} \int_{\Theta^{N-1}} \int_{\underline{\theta}}^{\overline{\theta}} \left[\int_0^{\theta_i} x(\beta, \theta_{-i}) d\beta + h_i(\theta_{-i}) \right] dF(\theta_i) dF^{N-1}(\theta_{-i})
\end{aligned}$$

$$= \int_{\Theta^N} \left[\sum_{i \in N} \psi(\theta_i) \right] x(\theta) dF^N(\theta) + \sum_{i \in N} \int_{\Theta^{N-1}} h_i(\theta_{-i}) dF^{N-1}(\theta_{-i}),$$

where the first equality follows by definition, the second from Lemma 1 and the third by plugging in equation (3). □

For any subset $S \subseteq N$ of the agents, define the corresponding *orthant* as $\mathcal{O}_S = \{\theta \in \Theta^N \mid \theta_i \geq 0 \text{ if } i \in S, \theta_i \leq 0 \text{ if } i \notin S\}$.

Lemma 4. *Suppose that $\psi(\theta)$ is non-increasing in θ and $\int \psi(\theta) dF^N(\theta) < \infty$. Let $\mathcal{O}_S$ be the orthant corresponding to some subset of agents S. Then the problem*

$$\begin{aligned} \max_x \; & \int_{\mathcal{O}_S} \psi(\theta) \cdot x(\theta) dF^N(\theta) \\ \text{s. t. } & x \textit{ is non-decreasing in } \theta \\ & 0 \leq x(\theta) \leq 1 \end{aligned}$$

is solved optimally either by setting $x^*(\theta) = 1$ or $x^*(\theta) = 0$.

The objective is to find a non-decreasing function that maximizes the integral over the product of this function with a non-increasing function. Extending Chebyshev's inequality to multiple dimensions yields that the objective function is maximized by choosing the non-decreasing function to be constant.

Proof. Suppose to the contrary that there exists a function $\hat{x}(\theta)$ that achieves a strictly higher value. Let $a_i := \inf\{\theta_i \mid (\theta_i, 0_{-i}) \in \mathcal{O}_S\}$, $b_i := \sup\{\theta_i \mid (\theta_i, 0_{-i}) \in \mathcal{O}_S\}$ and define $x^{(1)}(\theta_1, \theta_{-1}) := \frac{1}{F(b_1) - F(a_1)} \int_{a_1}^{b_1} \hat{x}(\beta, \theta_{-1}) dF(\beta)$. This function is constant in θ_1, feasible for the above problem given that $\hat{x}$ is feasible and, by Chebyshev's inequality, for all θ_{-1},

$$\begin{aligned} &\int_{a_1}^{b_1} \psi(\theta_1, \theta_{-1}) \hat{x}(\theta_1, \theta_{-1}) dF(\theta_1) \\ &\qquad \leq \int_{a_1}^{b_1} \psi(\theta_1, \theta_{-1}) dF(\theta_1) \frac{1}{F(b_1) - F(a_1)} \int_{a_1}^{b_1} \hat{x}(\theta_1, \theta_{-1}) dF(\theta_1) \\ &\qquad = \int_{a_1}^{b_1} \psi(\theta_1, \theta_{-1}) x^{(1)}(\theta_1, \theta_{-1}) dF(\theta_1). \end{aligned}$$

Since this inequality holds point-wise, we also have

$$\int_{\mathcal{O}_S} \psi(\theta) \hat{x}(\theta) dF^N(\theta) \leq \int_{\mathcal{O}_S} \psi(\theta) x^{(1)}(\theta) dF^N(\theta).$$

Iteratively defining $x^{(j)}(\theta_j, \theta_{-j}) = \frac{1}{F(b_j) - F(a_j)} \int_{a_j}^{b_j} x^{(j-1)}(\beta, \theta_{-j}) dF(\beta)$ for $j = 2, \ldots, N$ we get a function $x^{(N)}(\theta)$ that is constant in θ. Repeatedly applying Chebyshev's inequality along every dimension, we get

$$\int_{\mathcal{O}_S} \psi(\theta) \hat{x}(\theta) dF^N(\theta) \leq \int_{\mathcal{O}_S} \psi(\theta) x^{(N)}(\theta) dF^N(\theta).$$

Since the objective function is linear in x, the constant function $x^{(N)}$ is weakly dominated by either $x^* \equiv 1$ or $x^* \equiv 0$, contradicting the initial claim. □

Proof of Theorem 1. Lemma 2 and Lemma 3 together imply that for any anonymous social choice function $G = (X^G, T^G)$ it holds that

$$U(X^G, T^G) = \int_{\Theta^N} \left[\sum_{i \in N} \psi(\theta_i) \right] x(\theta) dF^N(\theta),$$

where ψ is defined in (2) and x is the decision rule of the corresponding strategy-proof direct revelation mechanism. Lemma 4 then implies that the optimal allocation rule is constant and equal to 0 or 1 in each orthant. Symmetry of the problem implies that the optimal choice depends only on the number of agents with positive types.

Hence, it remains to determine the optimal cutoff for qualified majority voting. Let k solve

$$k\mathbb{E}[\theta_i \mid \theta_i \geq 0] + (N-k)\mathbb{E}[\theta_i \mid \theta_i \leq 0] = 0.$$

Then the expected aggregate valuation, given that $k' < k$ agents are in favor of the proposal, is negative. Therefore, it is optimal to accept the proposal if and only if at least $\lceil k \rceil$ agents have a positive valuation. □

Proof of Corollary 1. Lemma 2 implies that for any social choice function satisfying the requirements of the corollary, one cannot redistribute money back to the agents. Lemma 1 then implies that any budget balanced social choice function must be constant in each orthant. Monotonicity and anonymity then imply that these social choice functions can be implemented by qualified majority voting. □

Chapter 2

Preference Intensities in Repeated Collective Decision-Making

We study welfare-optimal decision rules for committees that repeatedly take a binary decision. Committee members are privately informed about their payoffs and monetary transfers are not feasible. In static environments, the only strategy-proof mechanisms are voting rules which are inefficient as they do not condition on preference intensities. The dynamic structure of repeated decision-making allows for richer decision rules that overcome this inefficiency. Nonetheless, we show that often simple voting is optimal for two-person committees. This holds for many prior type distributions and irrespective of the agents' patience.

1. Introduction

Simple voting rules are known to be inefficient when a majority with weak preferences outvotes a minority with strong preferences. For instance, if ten out of one hundred citizens of a village are willing to pay $20 for changing a law, but the rest has a willingness-to-pay of $1 for keeping the old one, votes would be 90 to 10 against the new law, although it would be efficient to pass it.

Money could be used as a tool to elicit preference intensities and thereby to implement the efficient allocation, but in many situations there are moral or other considerations that prevent the use of monetary means. Instead, this chapter examines the possibilities of using the dynamic structure of environments where group decisions have to be made repeatedly in order to provide incentives for truthful preference revelation. In fact, repeated decision problems are ubiquitous in everyday life, ranging from examples in parliament to hiring committees. In these environments, it is sensible to assume that agents will not proceed myopically from period to period and therefore will not vote sincerely. As **?** emphasize, "any rule must be analyzed in terms of the results it will produce, not on a single issue, but on the whole set of issues." Consequently, it is not only reasonable to look at equilibrium behavior under a specific decision rule, but to search for rules that maximize a given objective like, for example, the welfare of the agents.

Consider the following example, which illustrates the possibility of increasing sensitivity to preference intensities: Assume that the decision rule prescribes to accept if at least one of two agents is in favor of the project, unless the other agent uses one of his limited possibilities to exercise a veto. In this situation, agents are faced with a trade-off between the current and future periods. If an agent exercises a veto now, the decision rule decides in her favor, but at the cost of fewer possibilities to use a veto in the future, which reduces the agent's continuation value. Intuitively, agents will use their veto right only if their preference against the proposed project exceeds some threshold. This has the effect that more refined information about the agents' preferences is elicited and potentially a more efficient allocation can be implemented.

Given these ideas, the question is why we see so many decision rules that use simple majority voting in every period, and, more generally, which decision rule is the best in terms of providing the highest welfare to the agents. In this chapter, we tackle the latter question and show that, surprisingly, voting rules are optimal among many reasonable decision rules. This provides a hint to the answer for the former question on why voting is used so universally.

More specifically, we analyze a model with two agents who are repeatedly presented a proposal that they need to either accept or reject. Each agent has a positive or negative willingness-to-pay for accepting the proposal, which is private information and drawn from a distribution function. Due to the revelation principle, we focus on direct mechanisms that simply map past preferences and decisions, and preferences in the current period, into a probability of accepting the current proposal. This allows for the modeling of many conceivable decision rules. We require that decision rules be incentive compatible, so that reporting preferences truthfully is a *periodic ex-post equilibrium*. This means that in any period, given any history, it is a dominant strategy to report the preference truthfully. This requirement renders incentives robust to uncontrolled changes in the information structure as well as deviations of the other player.

We provide a characterization of incentive compatible decision rules in terms of the allocation in a given period and the continuation values the rule promises. Viewing the continuation values as a substitute for money enables us to treat any given decision rule as a static mechanism which can then be improved upon while preserving incentives. The new continuation values of the improved static mechanism can then be implemented by specifying a new dynamic decision rule. As a result, we are able to show that if the preference distributions satisfy an increasing hazard rate condition, then voting rules are optimal within two classes of mechanisms. First, they are optimal among decision rules that satisfy *unanimity*, i.e., rules that never contradict the decision that both agents would unanimously agree on. This is a reasonable robustness requirement since one could expect that the agents will not adhere to the decision rule if they unanimously agree to do something else. Second, if the type distributions are *neutral across alternatives*, i.e., the density is symmetric around zero, then voting rules are also optimal among all deterministic decision rules.

Therefore, if the type distributions are neutral across alternatives, we get the summarizing result that any decision rule yielding higher welfare than every voting rule has both weaknesses of not satisfying unanimity and not being deterministic. This provides a strong rationale for the use of voting rules in the setting we consider and also provides hints on why rules other than voting are not considered in settings with more agents either.

Relation to the Literature

We build upon literature studying decision rules for dynamic settings. **?** note that

> much of the traditional discussion about the operation of voting rules seems to have been based on the implicit assumption that the positive and negative preferences of voters for and against alternatives of collective choice are of approximately equal intensities. Only on an assumption such as this can the failure to introduce a more careful analysis of vote-trading through logrolling be explained.

? proceed to analyze vote trading. They argue that agents can benefit if they trade their vote on a decision for which they have a weak preference intensity, and in turn get a vote for a future decision. However, it has early been noted that a trade in votes, while being beneficial for the agents involved, might actually reduce aggregate welfare of the whole committee, a fact sometimes called "the paradox of vote trading" (**?**). A formal analysis of vote trading has been missing until recently, when **?** examined in a competitive equilibrium spirit a model of vote trading. They show that vote trading can actually increase welfare in small committees, but is certain to reduce welfare for committees that are large enough.

Instead of relying on agents playing an equilibrium with non-sincere voting so that they can express their preference intensities, one can design specific decision rules that explicitly take intensities into account. **?** is the first to take this approach in a dynamic setting, in which agents repeatedly decide on a binary choice. He proposes the concept of storable votes: In each period, each agent receives an additional vote and can use some of his votes for the current decision or, alternatively, he can store his additional vote for future usage. By shifting their votes inter-temporally, agents can concentrate their votes on decisions for which they have a strong preference. **?** shows that this procedure increases welfare of the committee if there are two members and conjectures that in many circumstances this also holds for larger committees. **?** analyzes a similar proposal for a static setting (meaning that agents are completely informed about their preferences in all decision problems when making the first decision), in which agents face a number of binary decisions.

Going one step further, one can systematically look for the "best" decision rule. **?** take a mechanism design approach and show that for a static setting the efficient outcome can be approximated even in the absence of money, by linking a large number of independent copies of the decision problem. This result extends to dynamic settings, as long as individuals are arbitrarily patient. This surprising result hinges critically on a number of strong assumptions: each decision problem has to be an identical copy, the designer is required to have the correct prior belief, agents need to be arbitrarily patient and their beliefs about other agents have to be identical to the common prior. In an attempt to find more robust decision rules, **?** characterizes the set of strategy-proof decision rules for a static problem. Given that strategy-proofness is a strong requirement in multi-dimensional settings, it is not too surprising that voting rules are the only decision rules that satisfy this restriction.

In contrast, our focus on periodic ex-post equilibrium implies that on the one hand, the set of implementable decision rules is very rich, but on the other hand our results are robust and the optimal mechanism is bounded away from attaining the first-best.

The chapter is structured as follows: In Section 2 we present our model in detail. The results are presented in Section 3 and discussed in Section 4. Some proofs are omitted from the main text and relegated to the appendix.

2. Model

There are two agents who are repeatedly faced with a proposal and have to accept or reject each proposal. Periods are indexed by $t = 0, 1, ... \in T = \mathbb{N}$. The type of an agent i in a given period t is denoted by θ_{it} and indicates his willingness-to-pay for the proposal. Type spaces and distribution functions are the same for each period and each agent, denoted by Θ_i and F respectively, and types are drawn independently across time and agents. We denote by $\tilde{\theta}_{it}$ the random variable corresponding to the type of agent i, and by θ_t a type profile which is an element of the product type space Θ.

In each period, a decision $x_t \in \{0, 1\}$ has to be made. We denote the sequence of decisions up to period t by x^t, and similarly for a sequence of types θ_i^t. Accordingly, for an infinite sequence we write x^T.

Mechanisms

In this model a dynamic version of the revelation principle holds (**?**), hence we can focus on truthfully implementable direct revelation mechanisms.

Definition 1. *A* mechanism χ *is a sequence of decision rules* $\{\chi_t\}_{t \in T}$ *that map past decisions and type profiles into a distribution over decisions in the current period:*

$$\chi_t : \Theta^t \times \{0, 1\}^{t-1} \to [0, 1].$$

Preferences

Agents have linear von-Neumann-Morgenstern utility functions and there are no monetary payments. Given a period t and a decision x_t for this period, the utility of agent i with type θ_{it} is $v_{it}(\theta_{it}, x_t) = \theta_{it} x_t$. Agents discount the future with the common discount factor $\delta \in [0, 1)$. Consequently, utility of agent i with type sequence θ_i^T is

$$V_i(\theta_i^T, x^T) = \sum_{t \in T} \delta^t \theta_{it} x_t$$

for the decision sequence x^T.

Equilibrium Concept and Incentive Compatibility

In every period t, agent i learns about his preference type θ_{it}, which is his private information, and then sends a report r_{it}. The history known to the designer in period t, $h^t = (x^{t-1}, r^{t-1})$, consists of past decisions and past reports.

Given a mechanism χ, we can write the value function for agent i:

$$W_i(h^t, \theta_t) = \sup_{r_{it} \in \Theta_i} \theta_{it} \chi_t(h^t, r_{it}, \theta_{-it}) + \delta \mathbb{E}_{\Theta_{t+1}} W_i(h^{t+1}, \tilde{\theta}_{t+1}) \tag{1}$$

Here, h^{t+1} is the history in the next period, consisting of $\chi_t(h^t, r_{it}, \theta_{-it})$ and (r_{it}, θ_{-it}) appended to h^t. The valuation function specifies, given any history h^t, and the current type profile θ_t, the highest utility the agent can possibly obtain for some report r_{it}, assuming that she reports optimally in the future and the other agents report truthfully. Given a specific history h^t, the mechanism χ induces an allocation rule and continuation functions which we will denote

$$\begin{aligned} x_t(\theta_t) &= \chi_t(h^t, \theta_t) \quad \text{and} \\ w_{it}(\theta_t) &= \delta \mathbb{E}_{\Theta_{t+1}} W_i(h^{t+1}, \tilde{\theta}_{t+1}). \end{aligned}$$

If the current period is clear from the context, we will also drop the subscript t. The pair (x_t, w_t) is called the *stage mechanism after history* h_t and we say that w_t is *generated* by the mechanism χ. A stage mechanism is *admissible* if it is generated by some mechanism χ.

Definition 2. *A mechanism is* periodic ex-post incentive compatible (IC) *if for every period t and for all histories h^t the following holds: For every θ_{-i} and every θ_i we have that*

$$\theta_{it} x(\theta_{it}, \theta_{-it}) + w_{it}(\theta_{it}, \theta_{-it}) \geq \theta_{it} x(r_{it}, \theta_{-it}) + w_{it}(r_{it}, \theta_{-it}) \tag{2}$$

for all reports $r_i \in \Theta_i$.

See, e.g., **?**, or **?**. The definition in particular states that if a mechanism is incentive compatible, then every stage mechanism for all histories is incentive compatible. The following lemma can be proved using the Envelope Theorem (which is a standard exercise in mechanism design).

Lemma 1. *A mechanism is IC if and only if for each agent i the following two conditions hold:*

1. *Monotonicity of x: $x(\theta_i, \theta_{-i}) \leq x(\theta'_i, \theta_{-i})$ for $\theta_i \leq \theta'_i$.*
2. *Payoff equivalence: Fix $\hat{\theta}_i \in \Theta_i$. Then for all θ*

$$\theta_i x(\theta_i, \theta_{-i}) + w_i(\theta_i, \theta_{-i}) = \hat{\theta}_i x(\hat{\theta}_i, \theta_{-i}) + w_i(\hat{\theta}_i, \theta_{-i}) + \int_{\hat{\theta}_i}^{\theta_i} x(\beta, \theta_{-i}) d\beta. \tag{3}$$

Since the term $\hat{\theta}_i x(\hat{\theta}_i, \theta_{-i}) + w_i(\hat{\theta}_i, \theta_{-i})$ is independent of θ_i, we will write $h_i(\theta_{-i})$ for it. Note, however, that $h_i(\theta_{-i})$ does depend on the particular choice of $\hat{\theta}_i$.

Objective

For a given stage mechanism we can write down the expected welfare going forward from period t as

$$U_{h^t}(\chi) = U_{h^t}(x, w) := \mathbb{E}_{\Theta_t}\Big[(\theta_1 + \theta_2) x_t(\theta) + w_{1t}(\theta) + w_{2t}(\theta)\Big].$$

This is the period-t expected discounted welfare that the agents receive after history h^t. The aim of this chapter is to identify welfare-optimal mechanisms, that is, mechanisms χ that solve

$$\max_{\chi} \; U(\chi) := U_{h^0}(\chi), \quad \text{s. t.} \quad \chi \text{ is IC.}$$

Lemma 2 in the appendix provides a useful way to rewrite the objective function in terms of the allocation rule x and $h_i(\theta_{-i})$.

3. Results

The aim of this section is to identify mechanisms that are optimal in the above stated sense. The following conditions on F which we need to derive our results are standard in the mechanism design literature.

Condition 1 (Monotone Hazard Rates)**.** *The hazard rate $\frac{f(\theta_i)}{1-F(\theta_i)}$ is non-decreasing in θ_i and the reversed hazard rate $\frac{f(\theta_i)}{F(\theta_i)}$ is non-increasing in θ_i.*

A *voting rule x* is a rule where $x(\theta)$ only depends on $\{sgn(\theta_i)\}_{i=1,2}$. A voting mechanism is a mechanism where the allocation rule after all histories is a voting rule. In each of the two subsections below we will present a setting in which the welfare-maximizing dynamic decision rule is a voting mechanism.

The proofs in each part will proceed as follows: First, we show that under the appropriate assumptions stage mechanisms consisting of a voting rule and promising the same continuation payoffs for all type profiles are weakly welfare-superior to all other stage mechanisms. Then we make use of the following proposition to deduce that also the best dynamic mechanism uses a voting rule in every period. For this step to work it is helpful that optimal stage mechanisms are of as simple a form as voting mechanisms.

Proposition 1. *Assume that for every history h^t and admissible stage mechanism (x_t, w_t) in period t, there exists an admissible stage mechanism $(\hat{x}_t, \hat{w}_t)$, where $\hat{x}_t$ is a voting rule and $\hat{w}_t$ is constant, and such that*

$$U_{h^t}(x_t, w_t) \leq U_{h^t}(\hat{x}_t, \hat{w}_t).$$

Then a voting mechanism is among the optimal mechanisms.

Proof. We start with any dynamic mechanism χ and transform it into a mechanism that uses a voting rule in every period and such that U weakly increases. Start with $t = 0$. The assumption states that there exists a voting stage mechanism $(\hat{x}_0, \hat{w}_0)$ with constant $\hat{w}_0$ and such that $U(\hat{x}_0, \hat{w}_0) \geq U(x_0, w_0)$. Since the voting stage mechanism is admissible and promises constant continuations, these continuations can be generated by a mechanism that is independent of h^1. Denote by χ' this new dynamic mechanism. Since x'_1 and w'_1 are independent of h^1, we know (again by the assumption) that there exists a voting stage mechanism $(\hat{x}_1, \hat{w}_1)$ with constant $\hat{w}_1$ and such that $U_{h^1}(\hat{x}_1, \hat{w}_1) \geq U_{h^1}(x'_1, w'_1)$ for all h^1. Again, $\hat{w}_1$ can be generated by a mechanism that does not condition on histories h^2. Now if we let χ'' be the mechanism that arises if one exchanges the stage mechanism (x_1', w_1') in χ' for $(\hat{x}_1, \hat{w}_1)$, we know that χ'' is still incentive compatible: All promised continuations in period 0 change by the same amount, independent of the history h^1 and in particular independent of θ_0. Repeating this argument inductively for $t \geq 2$ completes the proof. □

Unanimity

Unanimity requires the mechanism to always adhere to a decision to which both agents agree. For example, if both types in some period are positive the mechanism has to choose $x_t = 1$ for sure. Formally, the condition is defined as follows:

Definition 3. *A mechanism is called* unanimous *if, for every period and all possible histories,* $x(\theta) = 1$ *if* $\theta > 0$ *and* $x(\theta) = 0$ *if* $\theta < 0$.

Note that mechanisms not satisfying this requirement will probably have legitimacy problems: Although all parties involved in the decision process opt in favor of the proposal, the mechanism forces its rejection. Furthermore, if agents are not able to collectively commit to the decision prescribed by the mechanism, then mechanisms satisfying unanimity are the only feasible mechanisms. Also note that mechanisms proposed in the literature are not excluded by this assumption (see, e. g., **?**, **?**). In the next subsection we will see that even when relaxing this restriction, for certain distribution functions only non-deterministic decision rules can yield a higher expected welfare than voting rules.

Theorem 1. *Suppose that* F *satisfies Condition 1. Then a voting mechanism is optimal among all unanimous mechanisms.*

Proof. The proof consists of establishing the preconditions of Proposition 1. So let (x, w) be a stage mechanism after some history h^t (since we are only concerned with unanimous mechanisms, x satisfies unanimity). Set $(\hat{\theta}_1, \hat{\theta}_2) = (0, 0)$ and let h_i be the resulting redistribution functions implied by Lemma 1. Let $\theta^* \in \arg\max_{\theta \in \Theta_i} h_1(\theta) + h_2(\theta)$. We first show that setting $h_1(\theta_2) = h_1(\theta^*)$ for all θ_2 and $h_2(\theta_1) = h_2(\theta^*)$ for all θ_1 does not decrease $U_{h^t}(x, w)$.

Since so far we have not changed x, by Lemma 2 it is enough to show that the terms involving the redistribution functions do not decrease in this step. But this follows from

$$\begin{aligned}\int_{\Theta_1} h_2(\theta_1) dF(\theta_1) + \int_{\Theta_2} h_1(\theta_2) dF(\theta_2) &= \int_{\underline{\theta}}^{\overline{\theta}} \Big[h_2(\beta) + h_1(\beta)\Big] dF(\beta) \\ &\leq \int_{\underline{\theta}}^{\overline{\theta}} \Big[h_2(\theta^*) + h_1(\theta^*)\Big] dF(\theta^*).\end{aligned}$$

Next we show that changing x to a voting rule does not decrease welfare. It is enough to consider the regions where $\theta_1 \leq 0, \theta_2 \geq 0$ and $\theta_1 \geq 0, \theta_2 \leq 0$ because the mechanism is unanimous. By Lemma 3 and the choice of $(\hat{\theta}_1, \hat{\theta}_2)$, we know that the first term in (4), which for the region $\theta_1 \leq 0, \theta_2 \geq 0$ amounts to

$$\int_{\underline{\theta}}^{0} \int_{0}^{\overline{\theta}} \left[\frac{-F(\theta_1)}{f(\theta_1)} + \frac{1 - F(\theta_2)}{f(\theta_2)}\right] x(\theta_1, \theta_2) dF(\theta_2) dF(\theta_1),$$

is maximized by setting x to 1, as soon as Condition 1 holds. Since the same is true for the region where $\theta_1 \geq 0, \theta_2 \leq 0$, we have constructed a voting stage mechanism that is weakly welfare superior to the old stage mechanism.

Let (x', w') denote the new stage mechanism. The proof is complete if we can show that w' is constant and can be generated. Constancy of w' holds for any stage mechanism where x' is a voting rule and the functions h'_i are constant. More specifically, w'_i is equal to $h_i(\theta^*)$. Since the old mechanism was unanimous, $w_i(\theta^*, \theta^*) = h_i(\theta^*)$. Because $w_i(\theta^*, \theta^*)$ could be generated, it follows that w' can be generated. □

Neutrality of Alternatives

In this section, we show that in some situations we can derive optimality of voting mechanisms even if unanimity does not hold. This shows that the restriction imposed in the previous section does in many cases not reduce welfare.

We assume that the distribution of types is *neutral across alternatives*, i.e., it is symmetric around 0. This is an important special case of our general model and has been analyzed, among others, by **?**. For instance, this assumption is satisfied if a committee has to decide among two proposals that are valued equally ex ante. Specifying one alternative as the default, the distribution of valuations for changing from the default to the alternative proposal is symmetric around 0.

Theorem 2. *Suppose F satisfies Condition 1 and is neutral across alternatives. Then a voting mechanism is optimal among all deterministic mechanisms.*

The proof of Theorem 2 is presented in the appendix. Similar arguments as in the last subsection can be given for restricting attention to deterministic mechanisms: First, stochastic mechanisms are difficult to implement and face legitimacy problems in practice. It is barely conceivable that a parliament would introduce decision protocols that involve random elements. Second, all proposed mechanisms in the literature and mechanisms observed in practice are usually deterministic and therefore not excluded from our analysis. Numerical simulation also suggests that expected welfare can be improved only slightly using stochastic mechanisms. The following corollary combines Theorem 1 and Theorem 2 and summarizes all properties one has to give up in order to improve upon voting rules.

Corollary 1. *Assume F satisfies Condition 1 and is neutral across alternatives. Then every decision rule that is strictly welfare-superior to any voting rule is stochastic and does not satisfy unanimity.*

4. Discussion

We have seen that despite the absence of money as a means for implementing rules other than majority voting, the possibility to condition decision rules on the past gives us the possibility to design dynamic decision rules that take preference intensities into account. However, we have shown that for committees consisting of two players the welfare maximizing dynamic decision rule nonetheless consists of simple majority voting in every period. This holds unless desirable properties of the decision rule are given up. We therefore provide a possible explanation for why majority voting is used almost universally in practice.

One extension of our model is to allow for correlation of agent types over time. However, this restricts the class of incentive compatible mechanisms since the quasi-linear separation of continuation payoffs from the payoff in the current period disappears. While voting rules would still be optimal in this restricted class, our model without correlation shows that voting rules are also optimal in the larger class.

A major open problem is the question as to what extent our results generalize to more than two agents. We believe that a substantial difficulty towards progress in this direction is to understand in how far continuation values can be redistributed among the agents.

Appendix

Helpful Lemmata

The following shows how the welfare of every incentive compatible mechanism can be expressed in terms of the allocation function and the functions h_i defined following Lemma 1.

Lemma 2. *Let χ be an incentive compatible mechanism and define*

$$\psi(\theta_i) = \begin{cases} \frac{-F(\theta_i)}{f(\theta_i)} & \text{if } \theta_i \leq \hat{\theta}_i, \\ \frac{1-F(\theta_i)}{f(\theta_i)} & \text{otherwise.} \end{cases}$$

Then for every history h^t we have

$$U_{h^t}(\chi) = \int_{\Theta} \Big[\psi(\theta_1) + \psi(\theta_2)\Big] x(\theta) dF(\theta) + \int_{\Theta_1} h_2(\theta_1) dF(\theta_1) + \int_{\Theta_2} h_1(\theta_2) dF(\theta_2). \quad (4)$$

Proof. First note that

$$U_{h^t}(\chi) = \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{\theta}}^{\overline{\theta}} \Big[\theta_1 x(\theta) + \theta_2 x(\theta) + w_1(\theta) + w_2(\theta)\Big] dF(\theta_2) dF(\theta_1), \quad (5)$$

and by Lemma 1

$$w_i(\theta) = \int_{\hat{\theta}_i}^{\theta_i} x(\beta, \theta_{-i}) d\beta - \theta_i x(\theta) + h_i(\theta_{-i}). \quad (6)$$

Using integration by parts, we first rewrite the term

$$\int_{\underline{\theta}}^{\overline{\theta}} \left[\int_{\hat{\theta}_i}^{\theta_i} x(\beta, \theta_{-i}) d\beta\right] f(\theta_i) d\theta_i$$

$$= \left[\int_{\hat{\theta}_i}^{\overline{\theta}} x(\beta, \theta_{-i}) d\beta \underbrace{F(\overline{\theta})}_{=1} - \int_{\hat{\theta}_i}^{\underline{\theta}} x(\beta, \theta_{-i}) d\beta \underbrace{F(\underline{\theta})}_{=0}\right] - \int_{\underline{\theta}}^{\overline{\theta}} x(\theta_i, \theta_{-i}) F(\theta_i) d\theta_i$$

$$= \int_{\hat{\theta}_i}^{\overline{\theta}} \frac{1 - F(\theta_i)}{f(\theta_i)} x(\theta) dF(\theta_i) + \int_{\underline{\theta}}^{\hat{\theta}_i} \frac{-F(\theta_i)}{f(\theta_i)} x(\theta) dF(\theta_i). \tag{7}$$

Now plug (6) into (5) and use (7) to complete the proof. □

The next lemma implies, together with Condition 1, that the first part of (4) is maximized by a constant allocation function whenever only one part of the function ψ is considered.

Lemma 3. *Suppose that $\psi(\theta_1, \theta_2)$ is non-increasing in θ_1 and θ_2, and that $\int \psi(\theta) dF(\theta) < \infty$. Then the problem*

$$\begin{aligned} \max_x \; & \int_a^b \int_c^d \psi(\theta_1, \theta_2) \cdot x(\theta_1, \theta_2) dF(\theta_2) dF(\theta_1) \\ \text{s.t. } & x \text{ is non-decreasing in } \theta \\ & 0 \le x(\theta) \le 1 \end{aligned}$$

is solved optimally either by setting $x^*(\theta) = 1$ or $x^*(\theta) = 0$.

Proof. Suppose to the contrary that there exists a function $\hat{x}(\theta)$ that achieves a strictly higher value. Define $x'(\theta_1, \theta_2) := \frac{1}{F(d) - F(c)} \int_c^d \hat{x}(\theta_1, \beta) dF(\beta)$. This function is feasible for the above problem given that $\hat{x}$ is feasible and, by Chebyshev's inequality, for all θ_1,

$$\begin{aligned} &\int_c^d \psi(\theta_1, \theta_2) \hat{x}(\theta_1, \theta_2) dF(\theta_2) \\ &\qquad \le \int_c^d \psi(\theta_1, \theta_2) dF(\theta_2) \frac{1}{F(d) - F(c)} \int_c^d \hat{x}(\theta_1, \theta_2) dF(\theta_2) \\ &\qquad = \int_c^d \psi(\theta_1, \theta_2) x'(\theta_1, \theta_2) dF(\theta_2). \end{aligned}$$

Since this inequality holds for every θ_1, we also have

$$\int_a^b \int_c^d \psi(\theta_1, \theta_2) \hat{x}(\theta_1, \theta_2) dF(\theta_2) dF(\theta_1) \le \int_a^b \int_c^d \psi(\theta_1, \theta_2) x'(\theta_1, \theta_2) dF(\theta_2) dF(\theta_1).$$

Defining $x''(\theta_1, \theta_2) = \frac{1}{F(b) - F(a)} \int_a^b x'(\theta_1, \theta_2) dF(\theta_1)$ and again applying Chebyshev's inequality as above, we get that

$$\int_a^b \int_c^d \psi(\theta_1, \theta_2) x'(\theta_1, \theta_2) dF(\theta_2) dF(\theta_1) \le \int_a^b \int_c^d \psi(\theta_1, \theta_2) x''(\theta_1, \theta_2) dF(\theta_2) dF(\theta_1).$$

Since the objective function is linear in x, the constant function x'' is weakly dominated by either $x \equiv 1$ or $x \equiv 0$, contradicting the initial claim. □

Proof of Theorem 2

Proof. We establish the preconditions of Proposition 1. Fix an arbitrary history h_t and consider the stage mechanism (x, w) employed after this history. Let $\overline{w} := max_\theta \{w_1(\theta) + w_2(\theta)\}$ and let θ_w be an optimizer. We normalize w such that $w_1(\theta_w) = w_2(\theta_w) = 0$ by

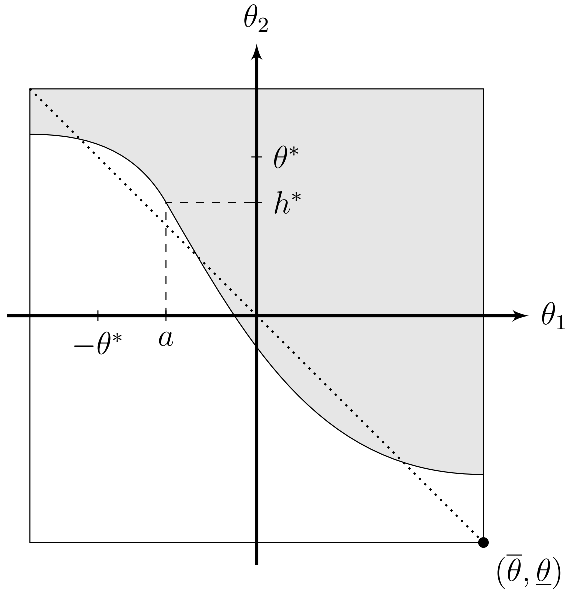


Figure 1: Proof of Theorem 2. The shaded area indicates the profiles θ where $x(\theta) = 1$.

decreasing w_i by $w_i(\theta_w)$ for all i. This does not affect incentive compatibility. After the normalization we have

$$w_1(\theta) + w_2(\theta) \leq 0.$$

We start with some preliminaries where we derive a set of inequalities that are satisfied by every incentive compatible stage mechanism for which the above inequality holds.

Preliminaries:

Set $(\hat{\theta}_1, \hat{\theta}_2) := (\overline{\theta}, \underline{\theta})$, let h_i denote the resulting redistribution functions implied by Lemma 1 and define $g_i(\theta) := \theta_i x(\theta) - \int_{\hat{\theta}_i}^{\theta_i} x(\beta, \theta_{-i}) d\beta$. It follows from Lemma 1 that $w_i(\theta) = -g_i(\theta) + h_i(\theta_{-i})$. Let $h^* := \max_\theta \{h_1(\theta) + h_2(-\theta)\} - \overline{\theta}$ and θ^* be a maximizer. Normalize h such that $h_1(\theta^*) = h^* + \overline{\theta}$ and $h_2(-\theta^*) = 0$ by increasing $h_1(x_2)$ and decreasing $h_2(x_1)$ by $h_2(-\theta^*)$. The definition of h^* implies

$$h_1(\theta) + h_2(-\theta) \leq h^* + \overline{\theta} \qquad \text{for all } \theta, \tag{8}$$

and $w_1(\theta, -\theta) + w_2(\theta, -\theta) \leq 0$ implies

$$\begin{aligned} h_1(\theta) + h_2(-\theta) &\leq g_1(-\theta, \theta) + g_2(-\theta, \theta) \\ &= -\int_{\overline{\theta}}^{-\theta} x(\beta, \theta) d\beta - \int_{\underline{\theta}}^{\theta} x(-\theta, \beta) d\beta \\ &\leq \int_{-\theta}^{\overline{\theta}} x(\beta, \theta) d\beta \leq \overline{\theta} + \theta. \end{aligned} \tag{9}$$

By plugging θ^* into (9) and using the definition of h^*, it follows that $h^* \leq \theta^*$.

Define $a := \inf\{\theta_1 \mid x(\theta_1, h^*) = 1\}$. If there does not exist θ_1 such that $x(\theta_1, h^*) = 1$, set $a := \overline{\theta}$. Without loss we can assume that $a \geq -h^*$, since otherwise we can "mirror" the mechanism on the dotted line shown in Figure 1.[1] Let $\theta_1 \geq a$. Then expanding and rearranging $w_1(\theta_1, \theta^*) + w_2(\theta_1, \theta^*) \leq 0$ yields

$$\begin{aligned} h_2(\theta_1) &\leq -(h^* + \overline{\theta}) + g_1(\theta_1, \theta^*) + g_2(\theta_1, \theta^*) \\ &= -h^* - \overline{\theta} + \theta_1 - \int_{\overline{\theta}}^{\theta_1} x(\beta, \theta^*) d\beta + \theta^* - \int_{\underline{\theta}}^{\theta^*} x(\theta_1, \beta) d\beta \\ &= -h^* + \theta^* - \theta^* + h^* - \int_{\underline{\theta}}^{h^*} x(\theta_1, \beta) d\beta \\ &= -\int_{\underline{\theta}}^{h^*} x(\theta_1, \beta) d\beta, \end{aligned} \tag{10}$$

where in the second equality we made use of the fact that $x(\beta, \theta^*) = 1$ for $\beta \geq a$ and $x(\theta_1, \beta) = 1$ for $\theta_1 \geq a$, $h^* \leq \beta \leq \theta^*$ (see Figure 1). Similar arguments will be used more often in the equalities below.

Define $b := \inf\{\theta_2 \mid x(-h^*, \theta_2) = 1\}$ (if there is no θ_2 such that $x(-h^*, \theta_2) = 1$, set $b := \overline{\theta}$) and let $\theta_2 \leq b$. Then $w_1(-\theta^*, \theta_2) + w_2(-\theta^*, \theta_2) \leq 0$ implies

$$\begin{aligned} h_1(\theta_2) &\leq g_1(\theta^*, \theta_2) + g_2(\theta^*, \theta_2) \\ &= 0 - \int_{\overline{\theta}}^{-\theta^*} x(\beta, \theta_2) d\beta - \int_{\underline{\theta}}^{\theta_2} x(-\theta^*, \beta) d\beta \\ &= \int_{-\theta^*}^{\overline{\theta}} x(\beta, \theta_2) d\beta. \end{aligned} \tag{11}$$

Since by Lemma 1 an incentive compatible stage mechanism is completely determined by x and h, we will in the following change x and h in a number of consecutive steps while making sure that x stays monotone and we never decrease the welfare $U^{h_t}(x, h) := U^{h_t}(x, w)$. At the end of the proof we will make sure that the resulting mechanism is admissible. First, we increase $h_2(\theta_1)$ for $\theta_1 \geq a$ and $h_1(\theta_2)$ for $\theta_2 \leq b$ until (10) and (11) hold with equality since this trivially weakly increases welfare.

Step 1:

In this step we will change the variables $x(\theta)$ with $\theta \in A := \{(\theta_1, \theta_2) \mid \theta_1 \geq a, \theta_2 \leq h^*\}$, $h_2(\theta_1)$ with $\theta_1 \geq a$ and $h_1(\theta_2)$ with $\theta_2 \leq h^*$. If we change h_1 and h_2 such that (11) and (10) continue to hold with equality, we can express changes of all the variables in terms of changes of x. Making use of the fact that for $\theta_2 \leq h^*$, (11) is equivalent to

$$h_1(\theta_2) = \int_a^{\overline{\theta}} x(\beta, \theta_2) d\beta,$$

and by substituting (11) and (10), we can rewrite the the part of U_{h^t} that depends on

[1] Let $(x^{\#}, w^{\#})$ be the mirrored mechanism, then $x^{\#}(\theta_1, \theta_2) = 1 - x(-\theta_2, -\theta_1)$, $w_i^{\#}(\theta_1, \theta_2) = w_{-i}(-\theta_2, -\theta_1)$. The new mechanism is IC iff. the old mechanism is IC and by our symmetry assumptions the mirrored mechanism yields the same welfare. Also, h^* and θ^* will not be changed by this operation.

changes of the variables $x(\theta)$ for $\theta \in A$ as

$$\int_a^{\overline{\theta}} \int_{\underline{\theta}}^{h^*} \left[\frac{-F(\theta_1)}{f(\theta_1)} + \frac{1-F(\theta_2)}{f(\theta_2)} \right] x(\theta_1, \theta_2) dF(\theta_2) dF(\theta_1)$$

$$+ \int_{\underline{\theta}}^{h^*} \int_a^{\overline{\theta}} x(\beta, \theta_2) d\beta \, dF(\theta_2) - \int_a^{\overline{\theta}} \int_{\underline{\theta}}^{h^*} x(\theta_1, \beta) d\beta \, dF(\theta_1)$$

$$= \int_a^{\overline{\theta}} \int_{\underline{\theta}}^{h^*} \left[\frac{1-F(\theta_1)}{f(\theta_1)} + \frac{-F(\theta_2)}{f(\theta_2)} \right] x(\theta_1, \theta_2) dF(\theta_2) dF(\theta_1).$$

Lemma 3 implies that this term is maximized by setting $x(\theta) = 0$ or 1 for $\theta \in A$. To see that we cannot gain by setting $x(\theta) = 1$ we bound

$$\begin{aligned}
U_{h^t}(1, h) &= \int_a^{\overline{\theta}} \int_{\underline{\theta}}^{h^*} \left[\frac{1-F(\theta_1)}{f(\theta_1)} + \frac{-F(\theta_2)}{f(\theta_2)} \right] dF(\theta_2) dF(\theta_1) \\
&= \int_{\underline{\theta}}^{-a} \int_{\underline{\theta}}^{h^*} \left[\frac{F(\theta_1)}{f(\theta_1)} + \frac{-F(\theta_2)}{f(\theta_2)} \right] dF(\theta_2) dF(\theta_1) \\
&= \int_{\underline{\theta}}^{-a} \int_{-a}^{h^*} \left[\frac{F(\theta_1)}{f(\theta_1)} + \frac{-F(\theta_2)}{f(\theta_2)} \right] dF(\theta_2) dF(\theta_1) \\
&\leq 0 = U_{h^t}(0, h).
\end{aligned}$$

Here, the second equality is due to the symmetry of F around zero, the third equality is because the integral over $[\underline{\theta}, -a] \times [\underline{\theta}, -a]$ vanishes, and the inequality is due to log-concavity of F and the fact that $-a \leq h^*$. Hence, we weakly increase welfare by setting $x \equiv 0$ in A and h_1 and h_2 according to (11) and (10), respectively.

Step 2:

For this step define the set $B = \{\theta_1 > -h^*, \theta_2 > h^* \mid x(\theta_1, \theta_2) = 0\}$. Set $x(\theta) = 1$ for $\theta \in B$ and $h_1(\theta_2) = h^* + \overline{\theta}$ for all θ_2 for which there is a θ_1 such that $(\theta_1, \theta_2) \in B$. We claim that this does not decrease U_{h^t}. Since allocative efficiency improved in this step, we only need to check that the sum of promised continuations increased. First, let $(\theta_1, \theta_2) \in B$. Then (11) is equivalent to

$$h_1(\theta_2) = \int_{-h^*}^{\overline{\theta}} x(\beta, \theta_2) d\beta.$$

Continuations before this change are given by

$$h_2(\theta_1) + h_1(\theta_2) + \int_{\overline{\theta}}^{\theta_1} x(\beta, \theta_2) d\beta = h_2(\theta_1) + \int_{-h^*}^{\overline{\theta}} x(\beta, \theta_2) d\beta + \int_{\overline{\theta}}^{\theta_1} x(\beta, \theta_2) d\beta = h_2(\theta_1).$$

After the change we get

$$h_2(\theta_1) + h^* + \overline{\theta} - \theta_1 + \int_{\overline{\theta}}^{\theta_1} \underbrace{x(\beta, \theta_2)}_{=1} d\beta - \theta_2 + \int_{h^*}^{\theta_2} \underbrace{x(\theta_1, \beta)}_{=1} d\beta = h_2(\theta_1).$$

Fixing $(\theta_1, \theta_2) \in B$, the claim can similarly be shown for points of the form (θ_1', θ_2) and

(θ_1, θ_2') where $\theta_2' > \theta_2$.

Step 3:

We claim that setting $x(\theta) = 1$ or $x(\theta) = 0$ for $\theta \in [\underline{\theta}, -h^*] \times [h^*, \overline{\theta}]$ increases U_{h^*}. This follows from the fact that, since, ignoring the part which depends on h_i, the objective function in the area where we change x has the form required by Lemma 3. Symmetry implies that $x(\theta) = 0$ gives the same welfare as $x(\theta) = 1$.

Step 4:

Note that the original mechanism satisfied

$$h_1(-\theta) + h_2(\theta) \leq h^* + \overline{\theta}.$$

Therefore, welfare is not decreased by setting $h_2(\theta) := 0$ and $h_1(-\theta) = h^* + \overline{\theta}$ for $\theta \leq -b$.

Note that the changed mechanism satisfies $w_1(\theta, -\theta) + w_2(\theta, -\theta) \leq 0$: For $a \leq \theta$ this holds as we assumed (10) and (11) to be binding in Step 1, hence $g_1(\theta, -\theta) = g_2(\theta, -\theta) = h_1(-\theta) = h_2(\theta) = 0$. For $-h^* \leq \theta \leq a$, this holds as continuations weren't changed for these values (changed Pivot payments were offset by changes in the h functions, as (11) was assumed to hold with equality in Step 1). For $-b \leq \theta \leq -h^*$ this holds as constraints were assumed to bind in Step 2. For $\underline{\theta} \leq \theta \leq -b$ this holds as $h_1(-\theta) + h_2(\theta) \leq h^* + \overline{\theta} = g_1(\theta, -\theta) + g_2(\theta, -\theta)$.

The fact that $w_1(\theta, -\theta) + w_2(\theta, -\theta) \leq 0$ implies that $h_1(-\theta) + h_2(\theta) \leq g_1(\theta, -\theta) + g_2(\theta, -\theta)$. We can increase h so that equality holds, thereby again improving the mechanism, ending up with the following stage mechanism:

$$x(\theta) = \begin{cases} 1 & \text{if } \theta_2 \geq h^* \\ 0 & \text{else,} \end{cases}$$

$$h_1(\theta_2) = \begin{cases} 0 & \text{if } \theta_2 \leq h^* \\ h^* + \overline{\theta} & \text{else,} \end{cases}$$

$$h_2(\theta_1) = 0.$$

We call this class of mechanisms *phantom dictatorship with parameter* h^*.

Step 5:

So far we have shown that every stage mechanism can be modified until it is a phantom dictatorship while weakly improving welfare. To prove that for every stage mechanism there is a simple voting stage mechanism with weakly higher welfare, we show that simple voting weakly welfare-dominates every phantom dictatorship: Indeed, the optimal phantom dictatorship is given by the parameter $h^* = \mathbb{E}[\theta]$. Therefore, symmetry of F around 0 implies that the optimal phantom dictatorship is characterized by $h^* = 0$, which has the same aggregate welfare as unanimity voting.

The voting stage mechanism we have constructed so far has the continuations profile $w_1(\theta) = w_2(\theta) = 0$ for all θ. It remains to show that this mechanism is admissible. But this follows from the fact that $(0, 0)$ was an implementable continuation profile of the original mechanism (namely, at the type profile θ_w). We therefore established the conditions for Proposition 1, which completes the proof of the theorem. □

Chapter 3

Sophisticated Sincerity under Incomplete Information

We study binary, sequential voting procedures in settings with privately informed agents and single-peaked (or single-crossing) preferences. We identify two conditions on binary voting trees, convexity of divisions and monotonicity of qualified majorities, ensuring that sincere voting at each stage forms an ex-post perfect equilibrium in the associated extensive form game with incomplete information. We illustrate our findings with several case studies: procedures that do not satisfy our two conditions offer ample space for strategic manipulations. Conversely, when the agenda satisfied our conditions, sincere behavior was indeed the most likely outcome.

1. Introduction

Sequential, binary voting procedures are widely used in democratic legislatures and committees. A very large literature in the fields of Economics, Law and Political Science has studied both their theoretical and their institutional/empirical aspects.[1] Almost the entire previous literature assumes that agents are completely informed about the preferences of others. Under complete information and simple majority, a Condorcet winner - which always exists with single-peaked preferences - must be selected by sophisticated voters independently of the particular structure of the binary voting tree and its particular agenda. Thus, at least for the many real life examples where the single-peakedness assumption is justified, the existing explanations of "strategic manipulations" found in the literature are somewhat strained. If preferences are known, why do some manipulations succeed, while others backfire? Why the "victims" of an attempted manipulation cannot anticipate it, and respond by changing their strategy?

We analyze sequential, binary voting in settings where several privately informed agents have single-peaked preferences on a finite set of alternatives, and we focus on robust equilibria that do not depend on assumptions about the players' beliefs about each other. We identify voting procedures in which sincere voting is a robust equilibrium and there is consequently no space for strategic manipulations. We also discuss several

[1] See for example **?**, **?**, **?**, and **?** for theoretical treatments, **?** and **?** for institutional aspects, and **?**, **?**, **?**, **?**, and **?** for case studies and empirical analyses.

case studies from the Legal and Political Science literature in the light of our results and illustrate how manipulations can succeed/backfire in voting procedures violating our conditions.

At each stage of a process described by a binary voting tree, one among two subsets of alternatives is adopted by (possibly qualified) majority. This process is repeated until a unique alternative is singled out, and formally elected. Under complete information, the associated extensive form games are amenable to analysis by backward induction, and the resulting, so called *sophisticated,* strategies and equilibrium outcomes are compared to the sincere ones that would result from *naive* (or myopic) behavior. The empirical strand is mostly concerned with the identification of strategic behavior or its absence, and their respective causes and consequences. Both, the theoretical and the empirical strand, emphasize the crucial role played by the agenda - the particular order in which alternatives appear in the voting tree/get eliminated from further consideration - and its control by interested parties.

The reason behind the focus on complete information is mainly technical: even with restrictions on the set of admissible preferences (such as single-peakedness), the analysis of the extensive form games represented by voting trees can be daunting if agents are privately informed. Optimal strategies generally depend on the specific, cardinal representation of utilities and on the beliefs about others. In turn, these beliefs are influenced by inferences that can be drawn from the ex-ante probabilities attached to the different possible profiles of preferences, and from new information generated by the employed strategies in the respective institutional setting. In particular, *manipulations* may occur also via voting behavior that attempts to influence the beliefs of other voters, and hence their future behavior (signaling effects), an effect that is absent under complete information. A pioneering analysis of strategic, sequential voting under incomplete information is **?** who construct Bayesian equilibria for an *amendment* procedure with three alternatives and with three possible preference profiles that potentially lead to a Condorcet paradox.[2]

Given the technical difficulties, our present analysis takes a different route, focusing instead on the much more robust, *ex-post perfect* equilibrium: in such an equilibrium, the resulting optimal strategies do not depend on ex-ante beliefs, and continue to be optimal irrespective of the information that is revealed during the sequence of votes. This is the strongest form of equilibrium possible in our setting since a dynamic implementation in dominant strategies is generally impossible if there are more than two alternatives. Not surprisingly, it is easy to construct simple examples - even with single-peaked preferences - where robust, dynamic equilibria do not exist.[3]

Our main results identify two intuitive conditions on binary voting trees ensuring that sincere voting at each stage forms an ex-post perfect equilibrium in the associated extensive form game with incomplete information. In other words, in the identified game trees sincere voting constitutes the sophisticated equilibrium for any possible profile of single-peaked preferences, for any ex-ante probabilities governing the agents' private

[2]See also **?**, who study the voting on the school construction bill of 1956 as an incomplete information game.

[3]Ex-post equilibria in settings with cardinal utility and with monetary transfers have been studied in the literature on *robust* mechanism design, e.g., by **?** and **?**.

information, and for any information disclosure policy along the voting sequence. In addition, the equilibrium is *strong Nash,* which means that no coalition of voters can profitably deviate from sincere voting. This last robustness result is particularly important in legislatures where voting is mostly according to party lines, so that coordinated deviations are likely.

The two conditions are:

1) *Convexity of divisions (CONV).* Recall that in procedures governed by binary voting, each vote is taken by (possibly qualified) majority among two, not necessarily disjoint, subsets of alternatives. Convexity says that if two alternatives a and c belong to the left (right) subset at a given node, then any alternative b such that $a < b < c$ (in the order governing single-peakedness) also belongs to the left (right) subset. Intuitively, each of the Yes-No votes in the sequence must be among two options that cover a well-defined, coherent segment of positions in the respective ideological spectrum.

2) *Monotonicity of qualified majorities* (*MON*). This condition says that, after a vote that resulted in the adoption of a left (right) subset of alternatives, a subsequent movement left (right) requires a qualified majority that is at least as large as the one that governed the previous move in the same direction. Intuitively, adopting consecutive and more "extreme" positions should become more and more difficult. The standard case of keeping a constant majority requirement at each vote - such as simple majority - satisfies monotonicity.

In order the understand the role of the above two conditions, note that, under incomplete information, the main determinants of optimality are the decisions at pivotality events: only such instances offer the opportunity to directly influence the outcome, and hence the optimal strategy must recommend a "correct" action whenever an agent is pivotal. In particular, agents should be able to make accurate inferences about future events, conditional on being pivotal.[4] Roughly speaking, the combined effect of CONV and MON is to finely tune this inference: for example, in a voting tree which satisfies these two assumptions, a pivotal agent can infer that a more preferred alternative will be ultimately elected, either because there are anyway enough other supporters for this alternative, or because the agent will continue to remain pivotal (and hence in control of the decision) at future stages. Thus, sincere voting - according to the preference relation restricted to the remaining set of alternatives - is optimal at each and every stage of the voting process.

Many ubiquitous voting procedures satisfy both conditions if simple and intuitive rules of agenda formation are respected. Let us consider two prominent examples:

1) The *amendment* procedure is predominantly used by legislatures in the English speaking world, Switzerland and Scandinavia. Alternatives are paired, and the winner competes against the next alternative on the agenda, until all alternatives are exhausted. The amendment procedure will satisfy convexity if the pairing is such that the most "extreme" alternatives compete against each other at each round of voting. If the single-peakedness order is $1 < 2 < ... < A$, the first vote should be among alternatives 1 and A. If 1 wins this contest then the next vote is among 1 and $A - 1$, whereas if A won at the first round, the second vote should be between A and 2, and so on. Monotonicity holds, for example, if the qualified majority needed to keep alternative 1 in the process

[4]This point has been forcefully made by **?**.

	Sequential Voting	Amendment Procedure
Always Vote on Most Extreme Alternative	Austria, Denmark, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Netherlands, Norway, Poland, Slovenia, Spain, European Parliament	Finland
Other procedural rule	Belgium, Czech Republic, Luxembourg, Portugal, Slovakia	Sweden, Switzerland, UK, (US)

Table 1: Parliamentary Floor Voting Procedures. Source: Rasch (2000)

is not decreasing along the voting tree.

2) The *successive* procedure is predominantly used in continental Europe. Voters consider alternatives one after the other, and the process stops as soon as one alternative garners the required majority. It satisfies convexity if, at each stage, the considered alternative is one of the two most extreme ones. If the single-peakedness order is $1 < 2 < ... < A$, convex agendas are, for example, to vote on the alternatives in the order $1, 2, 3, ..., A$,[5] in the order $A, A-1, ..., 1$ or even in a left-right alternating order such as $1, A, 2, A-1, ..., A/2$. If the successive procedure uses the order $1, 2, 3, ..., A$, monotonicity says, for example, that the qualified majority required to continue the voting process cannot decrease (i.,e, the majority needed to accept the current alternative does not increase).

Hence, for our conditions to be satisfied, the content of proposals (rather than purely procedural considerations) should determine the agenda. Of course, the interpretation of content may be ambiguous - leading to possible manipulations that could be prevented by more rigid, procedural rules. Table 1 (reporting findings by **?**) shows that the basic idea of a content-based agenda is anchored in the rules governing many legislatures. For example, the Standing Orders of the Slovenian parliament prescribes in its Article 192 that

> "if several amendments are proposed to an article of a proposed law, deputies shall vote first on the amendment which departs most from the content of the article in the proposed law, and then, following this criterion, on other amendments" (see **?**).

The present general treatment of varying qualified majorities at each node of the voting tree - generalizing the constant, simple majority rule assumed in almost the entire literature - has both an empirical and a theoretical content.

[5]For this particular procedure and agenda, monotonicity of qualified majorities has been shown to be necessary for a robust dynamic implementation by **?**. These authors were mainly concerned with identification of welfare maximizing mechanisms in settings where monetary transfers are not possible.

Firstly, many legislatures and committees use special, sometimes staggered, supermajorities for the passing of various, special laws. Prominent among these are the supermajorities required for laws that represent constitutional amendments in all democracies, and the Taxes and Expenditure laws (TELs) in most states in the U.S. For example, the legislature of Nebraska can vote to increase property taxes reflecting changes in the Consumer Price Index by simple majority, while larger increases up to 5% require a three-quarters majority. Further increases above 5% require a referendum in the population. Another good example, discussed in more detail below, is the U.S. Supreme court where accepting a case for review (*certiorari)* requires 4 out of 9 votes, but decisions on the accepted cases need to garner a simple majority (5 votes out of 9).

Secondly, varying the qualified majority at each node in the voting tree, allows us to theoretically replicate, via any binary voting procedure that satisfies convexity, the entire set of anonymous, unanimous and dominant strategy implementable social choice functions for the domain of single-peaked preferences. The resulting dynamic implementation in an ex-post perfect and sincere equilibrium necessarily satisfies our monotonicity condition. The proof of these results builds upon the seminal contribution of **?** where he identified the strategy proof social choice functions on the domain of single-peaked preferences as *generalized medians*. These are direct revelation mechanisms that choose the median peak among the reported peaks of real voters and other peaks of "phantom" voters (the phantom peaks are fixed by the mechanism, and do not vary with the reports of the real voters). Our results offer both a realistic implementation - via binary sequential voting procedures - of a rich class of non-dictatorial social choice functions, and a transparent interpretation of the "phantom voters" that play the main role in Moulin's analysis.

We also discuss below several case studies from the Legal and Political Science literature in the light of our results. We show below that, under incomplete information, procedures that do not satisfy our two conditions CONV and MON offer ample space for strategic manipulations that may or may not succeed, depending on the particular beliefs that voters entertain and on the actually realized profile of preferences (which is not common knowledge). As a consequence, the Condorcet winner need not be elected in equilibrium. Conversely, we illustrate examples where the agendas satisfied our conditions and where sincere behavior was indeed the most likely outcome.

A lively debate in the Political Science literature has revolved around the question whether strategic behavior is a common phenomenon in real life voting situations (see, for example, **?**, **?**, **?**, **?**). Several explanations have been advanced for the relative rarity of clear, unambiguous examples where such behavior has been postulated: 1) Legislators may be bound either by party discipline or by the need to fulfill constituents' expectations, and hence cannot act opportunistically at each instance (**?**, **?**); 2) Real life agendas are endogenous, and sincere voting is an equilibrium in the resulting game (of complete information) where the agenda is chosen in a first step (**?**). These, and other explanations, require additional features beyond those captured by the simple, isolated model of voting via a given, sequential binary procedure. In contrast, our results offer a simple explanation for the relative rarity of observed strategic voting based on robustness to the potential presence of private information: sincere voting constitutes the most compelling equilibrium for all situations that can be described by convex and

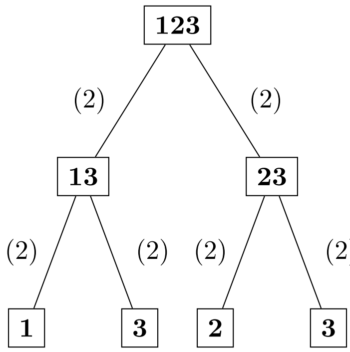


Figure 1: Illustration of an amendment procedure

monotone voting procedures, and hence it cannot be empirically distinguished from sophisticated voting (with which it simply coincides) in those cases. Conversely, our results also delineate the conditions under which strategic behavior may be advantageous for some parties.

The rest of the paper is organized as follows. In Section 2 we present several simple examples that illustrate our results. In Section 3 we present the voting model, and the convexity and monotonicity conditions. In Section 4 we present our main results for voting procedures where sincere voting is a robust equilibrium. In Section 5 we characterize the class of social choice functions that can be robustly and dynamically implemented our setting. In Section 6 we discuss several case studies in light of our findings. Section 7 concludes.

2. Illustrative Examples

The purpose of this section is to illustrate the main ideas and results in the simplest non-trivial setting with three privately informed agents and three alternatives, labeled $\{1, 2, 3\}$. We assume that preferences are strict and single-peaked with respect to the order on natural numbers.[6] This assumption yields four possible individual preferences: $1 \succ 2 \succ 3$, $2 \succ 1 \succ 3$, $2 \succ 3 \succ 1$, and $3 \succ 2 \succ 1$.

2.1 The Role of Convexity

Assume that the voters use an amendment procedure where the first vote is by simple majority between alternative 1 and 2, and the second vote is by simple majority between the winner of the first vote and alternative 3. This can be represented by the voting tree illustrated in Figure 1.

The second (and last vote) is either between 1 and 3, or between 2 and 3. Voting sincerely at the last vote (which is always a simple binary choice between two deterministic outcomes) is clearly optimal for all types of all agents.

Consider now the first vote between alternatives 1 and 2. Sincere voting calls for agents with peaks on 1 to vote for alternative 1, and for agents with peaks on 2 and 3 to vote for alternative 2. But, these actions cannot be part of an ex-post perfect

[6]By relabeling alternatives, this is without loss of generality.

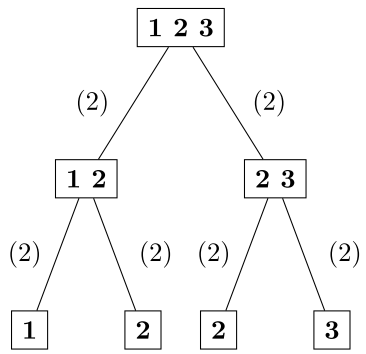


Figure 2: Illustration of an amendment procedure satisfying CONV

equilibrium: consider for example the case where our three agents have the following profile of preferences: $1 \succ 2 \succ 3$, $2 \succ 3 \succ 1$, $3 \succ 2 \succ 1$. Then alternative 2 is ultimately chosen under sincere voting, but a deviation to the left of the agent with peak on alternative 3 at the first vote (vote for 1) would result in a second vote among alternatives 1 and 3, where 3 would win. Hence such a deviation may be profitable, and our agent may regret his first sincere vote. Thus, sincere voting is not an equilibrium, and the reader may easily work out that an ex-post perfect equilibrium does not exist in this game.

Consider now the same amendment procedure, but with a different agenda: the first vote is between alternatives 1 and 3, and the second vote between the winner of the first vote and alternative 2. This procedure can be represented by the voting tree illustrated in Figure 2.

The second vote is either between 1 and 2, or between 2 and 3, and voting sincerely at the last vote is again optimal for all types of all agents. Consider now the first vote between alternatives 1 and 3. Sincere voting calls for agents with peaks on 1 to vote for alternative 1 and this is indeed optimal since every outcome that can be reached following the left branch at the origin is (weakly) preferred by such a player to any outcome than can be reached following the right branch. An analogous reasoning shows that sincere voting is optimal for agents with peaks on alternative 3.

Consider next an agent i with preference profile $2 \succ 1 \succ 3$, for whom sincere voting recommends voting for alternative 1 at the first vote. This vote matters only if such an agent is pivotal, thus only in case there is exactly one other agent that votes for alternative 1, and exactly one agent that votes for alternative 3. But then our agent i can be sure that voting sincerely at the first vote will ultimately lead to a second vote between 1 and 2 where his most preferred alternative is elected. An analogous reasoning for an agent i with preference profile $2 \succ 3 \succ 1$ completes the proof that sincere voting is an ex-post perfect equilibrium in this amendment procedure.

As we shall see below, the crucial difference between the two agendas is a (discrete) convexity idea: given the order under which preferences are single-peaked, the votes in the second example are always among sets of alternatives "without holes": the generated divisions are $\{12|23\}, \{1|2\}$ and $\{2|3\}$. In contrast, in the first example, the divisions are $\{13|23\}, \{1|3\}$ and $\{2|3\}$ and the division $\{13|23\}$ contains the non-convex set $\{13\}$ with a "hole" in place of alternative 2. This creates an uncertainty about the actions of others (and hence about the outcome) that cannot be satisfactorily resolved - i.e.,

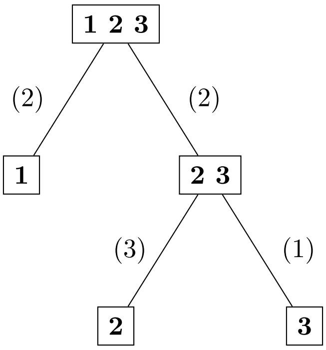


Figure 3: Illustration of a successive voting procedure

without ever experiencing regret - under incomplete information.

2.2 The Role of Monotonicity

In order to explain the role of our second condition, let us look at the successive voting procedure illustrated in Figure 3.

At the first vote the agents decide whether to accept or reject alternative 1. If 1 is accepted voting ends, and otherwise a vote is taken whether to accept or reject alternative 2. If this alternative is accepted voting ends, and otherwise alternative 3 is elected.

Assume first that alternative 1 is adopted by simple majority (so that two votes are sufficient to reject it) while alternative 2 is adopted by unanimity (so that one vote is sufficient to reject it, and hence elect alternative 3). Consider an agent i with preferences $2 \succ 1 \succ 3$. Then sincere voting calls for such an agent to vote against alternative 1 at the first vote. However, in case there is one agent with a peak on 1 and one agent with a peak on 3, sincere voting would lead to alternative 3 being elected. Agent i would then be better off by deviating and voting for alternative 1 at the first vote, which would lead to the implementation of alternative 1. Therefore, sincere voting is not an ex post perfect equilibrium for the above voting procedure.

Note that the partitions generated by this procedure are convex, without "holes": these are $\{1|23\}$ at the first vote and $\{2|3\}$ at the second. So the difficulty lies here elsewhere, namely in the specific thresholds for rejection of consecutive alternatives.

To correct the problem consider the same voting tree, but where alternative 1 is adopted by unanimity (so that one vote is sufficient to reject it) while alternative 2 is adopted by simple majority (so that two votes are required to reject it, and hence to elect alternative 3) (see Figure 4).

Then sincere voting is an ex-post perfect equilibrium: this is obvious for the agents with peaks on alternative 1 and 3 and for an agent with preferences $2 \succ 3 \succ 1$, so consider again an agent i with preferences $2 \succ 1 \succ 3$. If such an agent is pivotal at the first vote (to accept or reject 1) then there are at least two other agents with peaks on alternative 1, so that, at the next vote it must be the case that alternative 2 will be chosen. Hence it is optimal for our agent to also vote sincerely.[7]

[7]Claims in the literature indicate that sincere voting should be a dominant strategy for this proce-

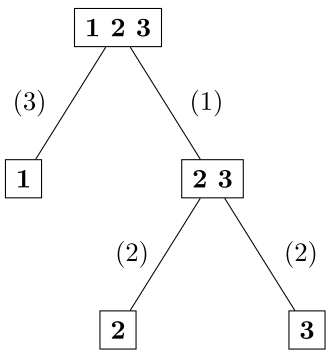


Figure 4: Successive voting procedure satisfying MON

This example shows that the thresholds at consecutive votes must be such that pivotal agents can infer (from the instance of them being pivotal) that, roughly speaking, they can still control the outcome at future stages.

3. The Sequential, Binary Voting Model

There is a finite set of alternatives, $A = \{1, 2, ..., |A|\}$, and a finite set of voters, N. Each voter $i \in N$ has a strict preference ordering $\succ_i$ over the elements of A. A preference ordering $\succ_i$ is single-peaked (with respect to the natural order $\leq$ on A) if there is an alternative $a \in A$ such that $c < b \leq a$ or $a \leq b < c$ imply $b \succ_i c$. We denote by $\mathbf{P}^{SP}$ the set of preference profiles that are single-peaked, and assume below that each voter has single-peaked preferences.

A binary tree is a pair $\mathbf{V} = (V, Q)$ where V is a set whose elements are called *nodes*, and $Q \subseteq V \times V$ is a partial ordering over V whose elements are called *branches*. V is assumed to have a unique least element o called the *origin*, and for any $v, v' \in V$ there is at most one Q-chain between v and v'. Any maximal element of V is called a *terminal node*, and we denote by V' the set of non-terminal nodes. If $v \; Q \; v'$ for $v \neq v'$, we say v' is a *successor* of v. Every non-terminal node has exactly two successors. If two distinct nodes are successors of the same node, we call them *neighbors*. If there are nodes $v_0 = v, v_1, ..., v_l = v'$ such that $v_{k-1} \; Q \; v_k$ for $k = 1, ..., l$, we say that v' *follows* v.

A *division correspondence* $L : V \rightrightarrows A$ associates with each node a subset of alternatives such that

1. $L(o) = A$ for the origin o,

2. $L(w) = L(u) \cup L(v)$ if u and v are successors of w,

3. $L(u) \subset L(w)$ if u follows w, and

4. $|L(t)| = 1$ for each terminal node t.

dure. This is not true as the following example illustrates: Suppose agent 1 has preferences $2 \succ 1 \succ 3$, agent 2 votes left in the first vote and right in the second, and agent 3 always votes right. Then sincere voting by agent 1 yields alternative 3 being selected. However, if agent 1 deviates and votes left in the first vote instead, then alternative 1, which he prefers to alternative 3, will be selected.

A set $L(v)$ denotes the alternatives that are under consideration by the voters at node v. A division L is *partitional* if $L(u) \cap L(v) = \emptyset$ for all neighbors u, v. We denote by $\min_{SP} L(v)$ and $\max_{SP} L(v)$ the smallest and largest alternatives of the set $L(v)$ in the order underlying the single-peaked preferences.

A *voting tree* $(\mathbf{V}, L)$ is a binary tree $\mathbf{V}$ together with a division correspondence L.

Definition 1. *A voting tree* $(\mathbf{V}, L)$ *satisfies* convexity of divisions (CONV) *if,*

$$\text{for all } v \in V,\ a \leq b \leq c \text{ and } a, c \in L(v) \text{ imply } b \in L(v).$$

Given a voting tree $(\mathbf{V}, L)$ and two neighbors $u, v \in V$, we label one of the branches leading to u by ℓ and the other by r. If $(\mathbf{V}, L)$ satisfies CONV, we label the branches leading to u and v as follows: If $\min_{SP} L(u) < \min_{SP} L(v)$, we label the branch leading to node u by ℓ, the *left* branch, and the branch leading to v by r, the *right* branch (and vice versa).[8]

We can identify a node in terms of the branches that lead to this node starting from the origin. Given a path $v \in \{r, \ell\}^k$, we denote by $v \oplus r$ the path of length $k+1$ whose first k entries coincide with the entries of v and whose last entry is r (and similarly for ℓ).

A *system of thresholds* for a voting tree $(\mathbf{V}, L)$ is a tuple of functions $\tau^\ell : V' \to \{1, 2, ..., n\}$ and $\tau^r : V' \to \{1, 2, ..., n\}$ such that for any $v \in V'$,

$$\tau^\ell(v) + \tau^r(v) = n + 1.$$

For any non-terminal node, the thresholds determine the number of votes needed in order to continue to the successor node according to the left and right branch, respectively. The sum of the two thresholds is $n+1$, so that no ties can occur. A *voting procedure* $(\mathbf{V}, L, \tau)$ is a voting tree $(\mathbf{V}, L)$ together with a system of thresholds τ.

Definition 2. *A voting procedure* $(\mathbf{V}, L, \tau)$ *satisfies monotonicity of thresholds (MON) if*

(i) for every $u, v \in V'$ *such that* $\max_{SP} L(u \oplus \ell) = \max_{SP} L(v \oplus \ell) = k$ *and* $\min_{SP} L(u \oplus r) = \min_{SP} L(v \oplus r) = k+1$ *, it holds that* $\tau^\ell(u) = \tau^\ell(v)$*, and*

(ii) for every $u, v \in V'$ *and* $s \in \{\ell, r\}$ *such that* v *follows* $u \oplus s$*, it holds that* $\tau^s(u) \leq \tau^s(v)$*.*

For voting procedures with convex divisions, part (i) is a mild consistency condition requiring that, whenever the vote is among the same set of alternatives, the same threshold is used, irrespective of the previous voting history. This consistency condition allows us to unambiguously define the threshold $\tau(k)$ that is used whenever k is the largest alternative associated with the left branch and $k+1$ is the lowest alternative associated with the right branch:

$$\tau(k) = \tau^\ell(v) \text{ for any } v \text{ such that } \max_{SP} L(v \oplus \ell) = k, \text{ and } \min_{SP} L(v \oplus r) = k+1.$$

[8] This labeling procedure is well-defined given that $(\mathbf{V}, L)$ satisfies CONV.

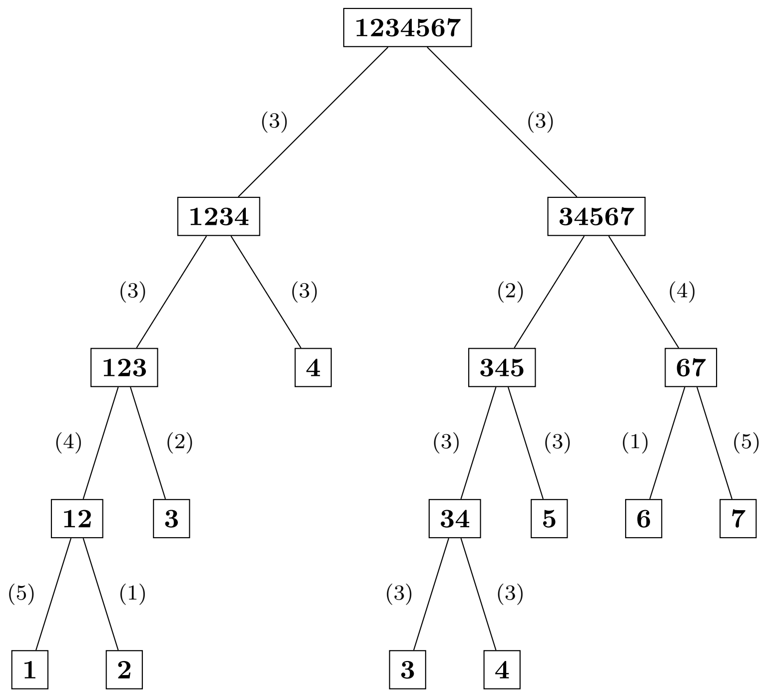


Figure 5: Example of a voting procedure satisfying CONV and MON.

Condition (*ii*) states that after a vote that resulted in the adoption of the left (right) division, a subsequent movement left (right) requires a qualified majority that is at least as large as the one that governed the previous move in the same direction.

Figure 5 shows an example of a general voting procedure satisfying CONV and MON.

Together with a given set of agents and their preferences and beliefs, any voting procedure describes a game of incomplete information. At each non-terminal node, players simultaneously vote either for the left or the right branch. If there are at least $\tau^\ell(v)$ voters voting for the left branch at node v, then the game advances to node $v \oplus \ell$; otherwise the game advances to node $v \oplus r$. If a terminal node v' is reached, then alternative $L(v')$ is implemented.

Let H_i^v denote the part of the history of play that is observable to player i at node v. One sensible specification is that H_i^v consists of the aggregate number of left and right votes at each previous node, and i's own voting behavior at all previous nodes. Another possible specification is that H_i^v includes the individual voting behavior of every player at all previous nodes. A strategy of player i associates to each node and each history a (possibly mixed) action. None of our results below depend on the exact specification of H_i^v.

Definition 3. *A strategy profile constitutes an* ex-post perfect equilibrium *if for every non-terminal node, following any history, and for every realization of preferences, the continuation strategies form a Nash equilibrium of the game in which the realization of preferences is common knowledge.*

The ex-post equilibrium embodies a notion of no-regret: even if all private information is revealed, no voter regrets her sincere strategy, given the sincere strategies of the other voters. It is particularly robust because it does not depend on the beliefs voters entertain.

For any preference ordering $\succ_i$ over A we denote, by the same symbol, its lexicographic extension over sets of alternatives. This allows us to define sincere voting as:

Definition 4. *A voting strategy is* sincere *given preference $\succ_i$ if it prescribes at each node $v \in V'$ to vote ℓ if and only if $L(v \oplus \ell) \succ_i L(v \oplus r)$.*

Under sincere voting, each voter votes for the set that contains his most preferred alternative. If this alternative is contained in both sets, he votes for the set containing his second-most preferred alternative, and so on. This definition of sincere voting goes back to **?** (see also **?** for a recent discussion of this definition).

4. Sophisticated Sincerity

In order to make our arguments as transparent as possible, we first treat the class of partitional voting procedures, where the sets of alternatives associated with two neighbors (i.e., successors of the same node) are disjoint. A well known example is the successive voting procedure: alternatives are considered one at a time, so that, for any non-terminal node, one successor node leads to a single alternative while its neighbor leads to all other remaining alternatives that were not yet eliminated. We will turn next to the somewhat more complex case where the voting tree need not be partitional.

4.1 Partitional Voting Procedures

Theorem 1. *Consider any partitional voting procedure satisfying CONV and MON. Then the profile of strategies where each player votes sincerely constitutes an ex-post perfect equilibrium.*

Proof. Suppose the result is not true. Then there exist a preference profile $\succ$, a voter i, and a node v such that sincere voting prescribes a left vote for i at v, but this is not a best response to the sincere strategies used by the others (the arguments are analogous if sincere voting prescribes a right vote). Because the voting procedure is partitional and satisfies CONV, there exists an alternative k such that $\max_{SP} L(v \oplus \ell) = k$ and $\min_{SP} L(v \oplus r) = k + 1$. Because sincere voting prescribes a left vote for i and because preferences are single-peaked, i prefers k to any alternative in the right branch, $k \succ_i l$ for all $l \in L(v \oplus r)$.

We now show that, if i always votes for the branch containing alternative k, then k will ultimately be selected. This contradicts our initial hypothesis that a right vote at v was a best response.

Since, by assumption, a left vote is not a best response for voter i at v, it must be that i is pivotal at v. Hence,

$$\#\{j \neq i \mid j \text{ votes } \ell \text{ at } v\} = \tau^{\ell}(v) - 1,$$

which implies that there are at most $\tau^{\ell}(v) - 1$ voters with a peak which is situated (weakly) left to of alternative k.

Since the voting procedure satisfies CONV, for every node v' following $v \oplus \ell$ it holds that, for some alternative $l < k$,

$$\max_{SP} L(v' \oplus \ell) = l \text{ and } \min_{SP} L(v' \oplus r) = l + 1$$

Therefore, only agents with a peak (weakly) to the left of l will vote for the left branch at v' and hence at most $\tau^\ell(v) - 1$ of the other voters will vote left at v'. Since the voting procedure satisfies MON, $\tau^\ell(v) \leq \tau^\ell(v')$. Therefore, if i votes right at v', the right branch is chosen. Since this holds for all nodes v' following $v \oplus \ell$, and since the right branch always contains alternative k (which is the largest remaining alternative), this alternative is selected at the final node. □

4.2 Non-Partitional Voting Procedures

We focus first on a particular, very prominent example of a non-partitional voting scheme, the amendment procedure. We then offer our general result about non-partitional voting procedures.

The Amendment Procedure

At each stage of an amendment procedure, a vote is taken among two alternatives, with the winner advancing to the next stage. Thus, for any two neighbors, the intersection of the corresponding sets of alternatives is generally non-empty, and consists of all alternatives that were not yet eliminated, and that are not directly considered at the respective stage.

Definition 5.

1. *An* amendment tree *is a voting tree* $(\mathbf{V}, L)$ *such that, at each non-terminal node, two alternatives are voted on and the losing alternative is eliminated.*
2. *An amendment procedure satisfies* CONV *if and only if for every non-terminal node v there exist alternatives $k < l \in A$ such that*

$$\begin{aligned} \min_{SP} L(v \oplus \ell) &= k < k + 1 = \min_{SP} L(v \oplus \ell) \\ \max_{SP} L(v \oplus \ell) &= l < l + 1 = \max_{SP} L(v \oplus r). \end{aligned}$$

Theorem 2. *Consider an amendment procedure satisfying CONV and MON. Then the profile of strategies where each player votes sincerely constitutes an ex-post perfect equilibrium.*

Proof. Assume, by contradiction, that sincere voting is not an equilibrium. Then, there exist a preference profile $\succ$, an agent i, and a node v such that sincere voting prescribes a left vote for i at v, but this is not a best response to the sincere strategies used by the others voters (the arguments are analogous if sincere voting prescribes a right vote). Suppose that the final outcome is alternative k if i votes right at v and plays best responses thereafter.

Since, by assumption, a left vote is not a best response for voter i at v, it must be the case that i is pivotal at v. If i votes left at v and right at all following nodes, then an alternative in the intersection set $L(v \oplus \ell) \cap L(v \oplus r)$ is chosen (because thresholds are monotone, and because the other voters vote sincerely). Because sincere voting prescribes a left vote at v, it must hold that $L(v \oplus \ell) \succ_i L(v \oplus r)$. Since preferences are single-peaked, all alternatives in the intersection set $L(v \oplus \ell) \cap L(v \oplus r)$ are preferred to the largest available alternative, $\max_{SP} L(v \oplus r)$. Therefore, we can conclude that $k \neq \max_{SP} L(v \oplus r)$ and hence that $k \in L(v \oplus \ell) \cap L(v \oplus r)$.

Since alternative k is selected following the right branch at v, a node v' is reached such that the left branch (which contains alternative k) is chosen, and such that $k+1 \notin L(v' \oplus \ell)$. By Lemma 1 (i) in the Appendix, the fact that the left branch is chosen at v' implies that at least $\tau(k)$ voters voted left; consequently, there are at least $\tau(k)$ voters having a peak weakly left of alternative k. If $k \neq \min L(v \oplus r)$, then a node v'' is reached such that the right branch (which contains alternative k) is chosen and $k - 1 \notin L(v'' \oplus r)$. By Lemma 1 (ii), there are at most $\tau(k-1)$ voters having a peak weakly left of alternative $k - 1$. If $k = \min_{SP} L(v \oplus r)$, the fact that i is pivotal at v implies that there are exactly $\tau(k-1) - 1$ voters with peak weakly left of alternative $k - 1$.

We now show that if i always votes for a branch containing alternative k, then k is chosen also following the left branch at v. Consider any node v' following $v \oplus \ell$ such that k is compared to a larger alternative, that is $k \in L(v' \oplus \ell)$ and $k \notin L(v' \oplus r)$. By Lemma 1 (iii), $\tau^{\ell}(v') \leq \tau(k)$. Since at least $\tau(k) - 1$ of the other voters have a peak weakly left of alternative k, the left branch is chosen at v'.

Analogous arguments imply that the right branch is chosen whenever alternative k is only in the right branch. As a consequence, the branch containing alternative k is chosen at each node, and k is finally elected in the left branch starting at v. This contradicts the initial assumption that a sincere left vote was not a best response for i at node v. □

General, Non-Partitional Voting Procedures

? argue that the theoretical concept of amendment procedures as defined above is too narrow, and that many realistic agendas take a different form. Therefore we now extend our results to a much broader class of voting procedures. Specifically, we show that for any voting procedure satisfying MON and CONV, sincere voting is an ex-post perfect equilibrium. Therefore, our characterization extends, for example, to *asymmetric* and *non-uniform* amendment agendas in Ordeshook and Schwarz's terminology.[9] Note, however, that among amendment procedures, only their *continuous* agendas can satisfy CONV.[10]

Theorem 3. *Consider a voting procedure satisfying CONV and MON. Then sincere voting is a strong ex-post perfect equilibrium.*

[9] An amendment procedure is *symmetric* if the voting tree splits into two subtrees that are alike except for an interchange of alternatives. It is *uniform* if all branches have the same length.

[10] An amendment procedure is *continuous* if, at every stage, the winner of the previous stage is contested at the current stage (i. e., it is contained in only one of the following divisions.

The proof of Theorem 3 is found in the Appendix. Note that this theorem shows that sincere voting is even a strong, ex-post perfect equilibrium, implying that there is no coalition that can achieve a better outcome by deviating from sincere voting. Moreover, it also immediately implies that sincere voting is the sophisticated, strong equilibrium of any procedure that satisfies CONV and MON for any complete information game where agents have single-peaked preferences.

4.3 Single-crossing preferences

It is well known that consistent aggregation of preferences with desirable properties is also possible on the domain of single-crossing preferences. (**?**, **?**, **?**, **?**). We briefly discuss here several implications for this important domain.

Given a linear order $\leq$ on alternatives and a linear order on preferences, we say that a set of preferences is *single-crossing* if $\succ'$being larger than $\succ$ and $a \leq b$ imply $b \succ' a$ whenever $b \succ a$.

Under single-crossing preferences, a Condorcet winner exists and it is the most preferred alternative of the median voter. Our next result provides conditions under which sincere voting is an ex-post perfect equilibrium if preferences are single-crossing. As a consequence, any voting tree satisfying CONV can be used to elect the Condorcet winner under incomplete information.

Theorem 4. *Suppose that n is odd, that preferences are single-crossing, and consider a voting procedure satisfying CONV that uses simple majority thresholds at each node. Then sincere voting is an ex-post perfect equilibrium.*

The proof of the theorem is found in the Appendix. Theorem 4 can be generalized to allow for more general qualified majority requirements. However, sincere voting need not constitute an ex-post perfect equilibrium for every voting procedure satisfying CONV and MON, as the following example illustrates. Consequently, compared to the analysis under single-peaked preferences, there is less freedom in choosing the majority requirements with single-crossing preferences. The additional problem arises because, under single-crossing preferences, an alternative need not be the peak of some admissible profile.

Example 1. *Suppose there are two voters and three possible preferences:* $1 \succ 2 \succ 3$, $1 \succ 3 \succ 2$, *and* $3 \succ 1 \succ 2$. *Note that this set of preferences is single-crossing (if preferences are listed in increasing order).*

Consider a successive voting procedure in which alternative 1 is selected at the first stage if both voters vote in favor of it. Otherwise the game proceeds to the second stage, at which alternative 2 is selected if at least one voter votes for it, and alternative 3 is selected otherwise. Note that this procedure satisfies MON and CONV.

Sincere voting prescribes a right vote at the first stage for voter 1 with preferences $3 \succ 1 \succ 2$. *If voter 1 is pivotal at the first stage, he can conclude that voter 2 prefers alternative* 1 *to the others. If voter 2 has preferences* $1 \succ 2 \succ 3$, *alternative 2 will be selected under sincere voting. But, voting left at the first stage is a profitable deviation for voter 1 since it leads to alternative 1 being implemented. Hence, sincere voting is not an ex-post perfect equilibrium.*

The above voting procedure facilitates the adoption of an alternative that is not one of the voters' peaks. This introduces here a potential incentive to deviate from sincere voting since a pivotal voter cannot be certain about the outcome of sincere voting.

5. Dynamic Implementation

In this section we generally characterize the social choice functions (seen as mappings that associate to each profile of single-peaked preferences a social alternative) that can be implemented by the sincere, ex-post perfect equilibria of sequential, binary voting procedures. To do so, let us first focus on the social choice functions that are anonymous, unanimous and dominant-strategy implementable (DIC) on the domain of single-peaked preferences. Unanimity is a weak form of Pareto-Optimality, and requires that an alternative is selected as the social choice when it is preferred by all agents to all other alternatives. Anonymity requires invariance of the social choice with respect to permutations of the voter's names. Note that we do not require neutrality since sequential, binary voting schemes and their agenda are per-se non-neutral.

The set of anonymous, unanimous and DIC social choice functions has been elegantly characterized by **?** and further refined by **?** and **?**. Moulin showed that any such mechanism can be described as choosing the median peak among n reported real peaks and $n-1$ *phantom* peaks (these are constant for each mechanism, and do not depend on the reports of the real voters).[11] In their study about optimal, utilitarian voting schemes **?** showed that the successive voting procedure with a particular convex agenda can be used to replicate, in ex-post perfect equilibrium, every anonymous, unanimous and DIC mechanism. This is done by varying the adoption threshold associated with each alternative in the successive procedure.[12] Our result below extends their observation to any convex, not necessarily partitional, voting procedure.

Theorem 5. *Consider any unanimous, anonymous, and DIC social choice function $f : \mathbf{P}^{SP} \to A$, and an arbitrary voting tree $(\mathbf{V}, L)$ satisfying CONV. Then there exists a system of thresholds satisfying MON such that sincere voting is an ex-post perfect equilibrium that implements f.*

Proof. It follows from **?** and **?** that f is a generalized median voting rule, with $\rho_k \geq 0$ phantom voters with peak on alternative k, such that $\sum_{k\in A} \rho_k = n-1$. For each node $v \in V'$, set

$$\tau^\ell(v) = n - \sum_{m=1}^{\max_{SP} L(v\oplus\ell)} \rho_m \quad \text{and} \quad \tau^r(v) = \sum_{m=1}^{\max_{SP} L(v\oplus\ell)} \rho_m + 1.$$

By construction, $\tau^\ell(v) + \tau^r(v) = n+1$ for all $v \in V'$, and this system satisfies the consistency condition (i) in the definition of MON. We now show that it also satisfies the monotonicity condition (ii). Suppose that node v follows $u\oplus\ell$. Then $L(v\oplus\ell) \subset L(u\oplus\ell)$ and hence $\tau^\ell(v) \geq \tau^\ell(u)$. Similarly, if v follows $u \oplus r$, then $L(v \oplus r) \subset L(u \oplus r)$.

[11] Moulin assumed that the social choice functions depend only on the reported peaks, while Border and Jordan, among others, showed that the peak-only assumption is not needed.

[12] Note that in that procedure each alternative appears at a unique terminal node.

Hence, $\tau^\ell(v) \le \tau^\ell(u)$, which is equivalent to $\tau^r(v) \ge \tau^r(u)$. Therefore, the system of thresholds τ satisfies MON, and Theorem 3 implies that sincere voting is an ex-post perfect equilibrium.

Fix now a preference profile $\succ$ in $\mathbf{P}^{SP}$ and suppose that $f(\succ) = k$. We show that k is selected in the sincere voting equilibrium. Because $f(\succ) = k$, there are at least $n - \sum_{m=1}^{k} \rho_m$ agents with a peak weakly to the left of k and at most $n - \sum_{m=1}^{k-1} \rho_m - 1$ voters with a peak strictly to the left of k. By construction

$$\tau^\ell(v) = n - \sum_{m=1}^{\max_{SP} L(v \oplus \ell)} \rho_m \le n - \sum_{m=1}^{k} \rho_m$$

for any node $v \in V'$ such that $k \in L(v \oplus \ell)$. Since there are at least $n - \sum_{m=1}^{k} \rho_m$ agents voting for left, the left branch is chosen in this case. If $k \notin L(v \oplus \ell)$, then, by construction,

$$\tau^r(v) = \sum_{m=k}^{\max_{SP} L(v \oplus \ell)} \rho_m + 1 \le \sum_{m=1}^{k-1} \rho_m + 1.$$

By the above argument, there are at least $\sum_{m=1}^{k-1} \rho_m + 1$ voters with a peak weakly to the right of k, and therefore the right branch is chosen. Hence, at each node, a branch that contains alternative k is chosen, and consequently the final choice must be alternative k. □

6. Case studies

In this Section we present several case studies in light of our findings. In particular, we illustrate cases where a probable Condorcet winner has not been chosen under sophisticated voting, a phenomenon that cannot be satisfactorily explained under complete information.

6.1 Pension and Women's Status in Switzerland (Senti, 1998)

This case is about a pension reform (pertaining to the status of married women), debated in the Swiss Chamber of Cantonal Representatives.[13] The alternatives were:

(1) keep the status quo where women's benefits are mainly defined by their husband's contribution

(2) a moderate reform[14]

(3) a radical reform, that would make pension contributions and benefits individual rather than family based.

The likely (but not certain) preferences where:

- $1 \succ 2 \succ 3$ for a group of 18 legislators associated with the ruling, conservative party.

[13]This is the second Chamber of the Swiss Parliament, similar to the U.S. Senate.

[14]In fact there were two moderate proposals, but this does not change the conclusions. We use the simplified version, following Senti.

- $2 \succ 1 \succ 3$ or $2 \succ 3 \succ 1$ for a group of 21 legislators associated with moderate parties

- $3 \succ 2 \succ 1$ for a group of 6 legislators associated to left leaning parties (?).

The first vote was among alternatives 2 and 3. Although 2 is the Condorcet winner, alternative 3, the radical proposal, won 24-19. The only possible explanation is that the conservatives voted strategically for their **worst** alternative in order to eliminate the Condorcet winner and likely outcome of the overall vote from further consideration.[15] In the second vote, alternative 1, the status quo preferred by the Conservatives easily won 30-13 against the radical proposal 3.

This is exactly the kind of manipulation we illustrated in our introductory example (see Figure 1) and is typical of "both extremes against the middle" example of successful strategic voting discussed in the literature. But note that, under complete information such a manipulation should never be successful since a Condorcet winner should be elected irrespective of the agenda! In the present example, the 6 left leaning legislators could have shifted their vote to alternative 2, the moderate reform, causing it to advance to the next stage where it would have most probably won against alternative 1. Thus, the above outcome can be explained by a complete information model only by attributing sophistication to some legislators and naivete to others, a not very satisfactory hypothesis. Instead, under incomplete information and a non-convex agenda, sincere voting is not an equilibrium, and hence various beliefs can lead to maneuvers that succeed in some circumstances while failing in others.

The following example illustrates the strategic deviations from sincerity that are part of many Bayesian equilibria.

Example 2. *There are 3 voters that use the voting procedure illustrated in Figure 1. Each of the four possible single-peaked ordinal preferences is associated with a "type": preference* $1 \succ 2 \succ 3$ *with type* t_1*,* $2 \succ 1 \succ 3$ *with type* t_2*,* $2 \succ 3 \succ 1$ *with type* t_3*, and* $3 \succ 2 \succ 1$ *with type* t_4*. Types are I.I.D. and, for each agent, type* t_i *is realized with probability* q_i*. In addition, for the analysis of Bayesian equilibria, we need a cardinal description of utilities: we assume that each voter values his most preferred alternative by 1, his second-most preferred alternative by* v*, and his least preferred alternative by 0. We show in the Appendix that, if* $v \leq \frac{\min\{q_2,q_3\}}{q_2+q_3}$*, then the following strategies form a Bayes-Nash equilibrium: all types except* t_4 *vote sincerely, and voters of type* t_4 *vote left at the first stage and vote sincerely at the second.*

Even though they are aware of the deviation by type t_4 *voters, it is optimal for type* t_1 *voters to vote sincerely because a deviation would significantly reduce their chances of getting their most preferred alternative. Type* t_4 *voters vote strategically (for their worst alternative!) at the first stage in order to increase the chances of their most preferred alternative at the second stage. This is optimal for type* t_4 *voters as long as* v *is small enough, but this behavior entails a risk that their worst alternative will be finally adopted. The following case study provides exactly such an instance.*

[15] This is analogous to the "killer amendments" discussed in the literature about the U.S. Congress.

6.2 Gun Control in Sweden (Bjurulf and Niemi, 1978)

The Swedish parliament, composed of two chambers, had to decide among three alternatives:

1 : appropriate 500.000 crown for the riflemen's association.

2 : appropriate 470.000 crowns;

3 : no appropriation.

The main difference between 1 and 2, and the reason for controversy was an extra sum for young riflemen aged 12-15. The Social Democrats had clear preferences $3 \succ 2 \succ 1$, while the Conservatives had opposed preferences $1 \succ 2 \succ 3$. The Farmers had preference $2 \succ 1 \succ 3$, while the preferences of the Liberals were more uncertain, but most probably did not have a peak on alternative 1.

In each chamber the procedure was an amendment agenda with 1 and 3 competing against each other at the first stage, and with the winner against 2 at the second stage. This is a convex agenda, so that, according to our theory, we should not observe strategic manipulations, e.g. of the type "both extremes against the middle".

In one Chamber the result was unspectacular, with alternative 3 winning over alternative 2 with a 81-72 margin. But, this case is particularly interesting because of the entire structure of the legislative process: each Chamber decides independently, and if the outcomes are different, a joint vote is taken among the respective winners in the two Chambers. This may lead to a violation of convexity. See Figure 6 for an illustration.

The binary supra-structure can induce the same incentives for manipulation as a non-convex agenda: Indeed, the Social Democrats tried to achieve their preferred outcome by letting alternative 1 - their worst outcome - win in the other Chamber, so that the final contest would be against 1 and 3, where 1 was expected to loose. They were wrong: in the final vote alternative 1 defeated 3 with 197 to 168 votes, and the manipulation replaced the likely Condorcet winner, alternative 2, with the worst outcome for the manipulators (**?**). It should be clear that, under an assumption of complete information, such a failure should never occur.

6.3 Roll Calls in the U.S. Congress (Ladha, 1994)

This study is unusual because it looks at 200 votes selected from more than 8000 roll calls over a period of 8 years (**?**). The criterion for selection - the existence of several comparable votes on the same issue - was driven by the author's aim of carefully testing several direct empirical implications of a model proposed by **?** in their monumental book on the U.S. Congress. It turns out that series of votes that are presented, which cover quite diverse areas of legislation, describe situations where the order of votes on amendments closely follows their position on the Liberal-Conservative ideological spectrum. Let us illustrate these general findings with one of Ladha's examples: this concerns a series of 1977 Senate votes on the level of tip credit enabling hotels and restaurants to pay less than the minimum wage to employees who receive part of their earnings from tips. The Human Resources Committee has proposed a 20% credit, i.e., allowing pay 20% under the minimum wage. Three amendments were considered at levels of 50%, 40% and 30%, respectively. Since the proposals are mutually disjoint and concern the entire issue at hand, the voting tree suggested by the actual order of votes

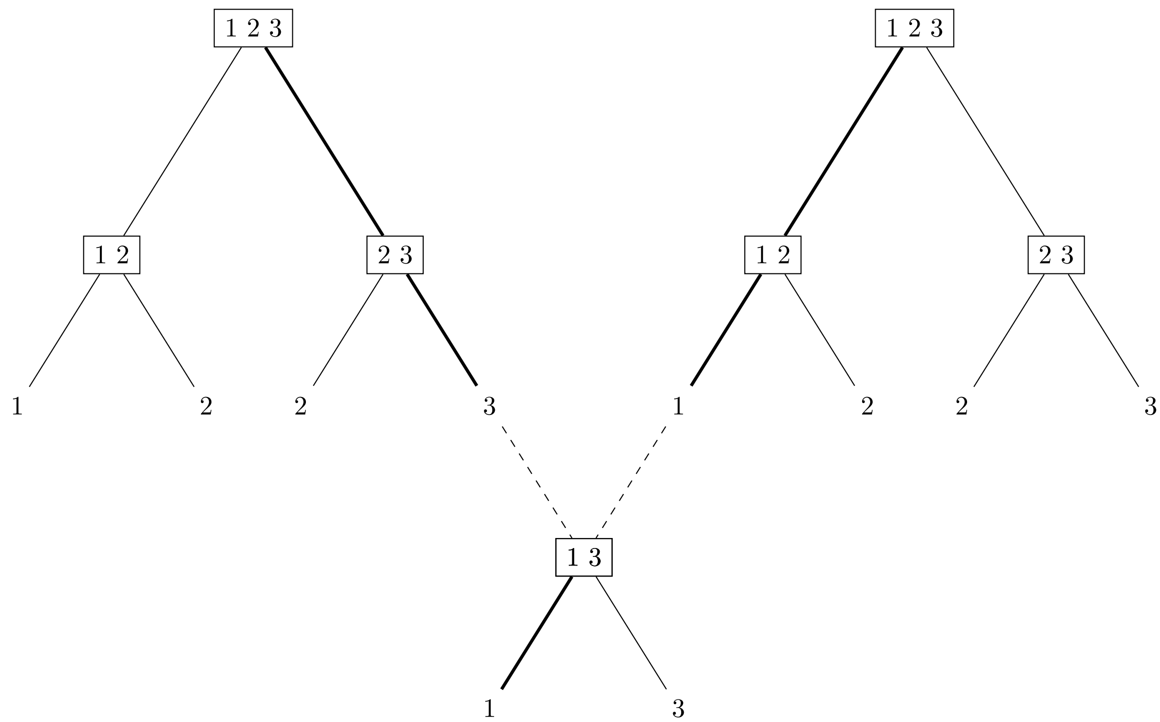


Figure 6: Gun control in Sweden

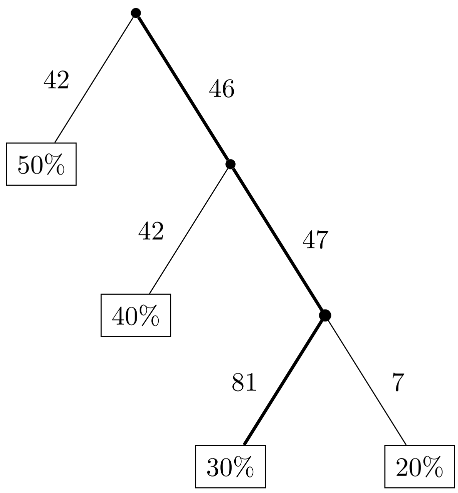


Figure 7: Successive voting procedure satisfying MON

is depicted, together with the voting results, in Figure 7.

Note that sincere voting is an equilibrium (the agenda is convex) and one should expect the number of Yes votes (of legislators who prefer higher levels of credit) to increase and the number of No votes (of legislators who prefer lower levels) to decrease, as voting advances from the high to the low proposal.[16] This is what the actual results suggest for this case, and basically the same picture emerges in all 200 examined cases: smooth increases of the number of Yes votes paralleled by smooth decreases in the number of No votes (or vice versa) as the amendment becomes more and more liberal or more and more conservative. Thus, there are never instances of large swings characteristic of strategic voting, as illustrated above.

Although we do not have precise information about the hypothetical complete agenda in all these cases, Ladha's many examples suggest indeed that trees with a convex agenda lead to sincere voting.

6.4 Certiorari at the U.S. Supreme Court (Caldeira et al., 1999)

Most cases decided by the U.S.Supreme Court arise from petitions to review decisions of lower courts.[17] The decision process is successive: first, a decision to grant or deny the cert is made, where granting requires at least 4 out of the 9 judges to be in favor (thus less than a simple majority!).[18] This decision need not be explained to outsiders (the process is rather secretive), and a denial does not constitute per-se an acknowledgment of the earlier decision. If a cert is granted, a decision "on merits", roughly speaking to affirm, or to reverse the opinion of the lower court, follows by simple majority (5 out of 9). Thus, the basic decision procedure is binary and successive and does not use a constant adoption threshold (see Figure 8).

The main source of uncertainty about the preferences of others is the presence of

[16] One senator voted only for the 40% level, which explains the small increase, from 46 to 47.

[17] Certiorari (or certs) is the Latin name for requests of more information, used because the Supreme Court requests the relevant documents from the lower court.

[18] See **?** for an analysis of this rule.

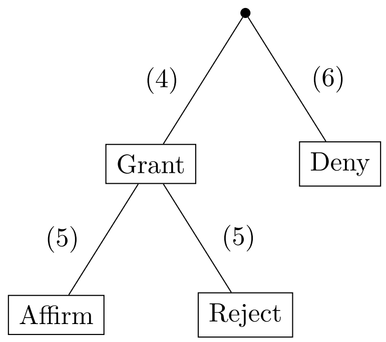


Figure 8: Structure of the decision process at the U.S. Supreme Court

"moderate" judges who do not have strong ideological convictions that could be used to predict their opinions with high probability. Unless the earlier "status quo" decision is clearly on one side of the possible decisions on merit, i.e., more "liberal" or more "conservative" or obviously wrong on pure judicial terms, the procedure is not convex: a convex successive procedure must start with the decision on one of the most extreme alternatives, which is simply infeasible here. In particular, the monotonicity of the adoption thresholds is only apparent.

Most certs are denied because they are deemed frivolous, but it has been argued, and frequently documented, that some cert decisions are "defensively" denied, for example, by liberal (conservative) judges that are afraid of a reversal (affirmation) of an earlier liberal decision. In other words, the vote at the first stage is not sincere, on the potential merits of the cert. Such a strategic vote is indeed optimal in this non-convex procedure for beliefs that attach a sufficiently high probability to a subsequent unfavorable outcome, and is more likely to occur for courts where the number of moderates is high enough, so that pivotality at both stages matters. **?** identify 18 such potential cases in 1982.

7. Conclusion

This paper offers a tractable, robust theoretical framework for analyzing ubiquitous voting procedures where voters are privately informed about their preferences. If the outcome of a voting procedure should not be sensitive to beliefs about others, nor to the deployment of strategic skills, the agenda needs to be built "from the extremes to the middle" so that more extreme alternatives are both more difficult to adopt, and are put to vote before other, more moderate options.

Our results illuminate the empirical discussion about the relative frequency of strategic voting in real life situations. Firstly, they imply that sincere voting is often a very robust, ex-post perfect equilibrium, and hence that sincere and sophisticated voting are observationally equivalent in many relevant cases. Secondly, for all reviewed case studies where strategic deviations from sincere voting are found, our theoretical results predict that sincere voting is not an ex-post perfect equilibrium, and hence that profitable deviations matching the observed actions do exist for some types of agents and some of their beliefs.

Appendix

Arguments for Example 2.

Sincere voting by all types in the second stage is a weakly dominant strategy. Moreover, sincere voting at the first stage is optimal for voters of type t_2 and t_3: Whenever they are pivotal, they can be sure to obtain their most preferred alternative given the sincere strategies used by the others.

It is optimal for voters of type t_1 to vote left if, conditional on being pivotal, they prefer the left branch:

$$\begin{aligned}\mathbb{E}U(\ell|t_1,\ \text{pivotal}) &= \frac{1}{(q_1+q_4)(q_2+q_3)}\Big[q_1(q_2+q_3)+q_4q_2\Big]\\ &= \frac{q_1}{q_1+q_4}+\frac{q_2q_4}{(q_1+q_4)(q_2+q_3)}\\ \mathbb{E}U(r|t_1,\ \text{pivotal}) &= v\end{aligned}$$

Similarly, it is optimal for type t_4 to vote left if, conditional on being pivotal, they prefer the left branch:

$$\begin{aligned}\mathbb{E}U(\ell|t_4,\ \text{pivotal}) &= \frac{1}{(q_1+q_4)(q_2+q_3)}\Big[q_4(q_2+q_3)+q_1q_3\Big]\\ &= \frac{q_4}{q_1+q_4}+\frac{q_1q_3}{(q_1+q_4)(q_2+q_3)}\\ \mathbb{E}U(r|t_4,\ \text{pivotal}) &= \frac{1}{(q_1+q_4)(q_2+q_3)}\Big[q_4(q_2+q_3)+q_1(q_2+q_3)v\Big]\\ &= \frac{q_4}{q_1+q_4}+\frac{q_1}{q_1+q_4}v.\end{aligned}$$

It can be easily verified that, given the strategies of the others, if $v \leq \frac{\min\{q_2,q_3\}}{q_2+q_3}$, then voting left in the first stage is a best response for voters of type t_1 and t_4. □

Lemma 1. *Consider an arbitrary voting procedure satisfying CONV and MON. Given any $u \in V$ such that $\max_{SP} L(u \oplus \ell) = k$ and $\min_{SP} L(u \oplus r) = k+1$, let $\tau(k) = \tau^\ell(u)$.*[19] *Then the following statements hold:*

(i) If $k \in L(v \oplus \ell)$ but $k+1 \notin L(v \oplus \ell)$, then $\tau^\ell(v) \geq \tau(k)$.

(ii) If $k-1 \notin L(v \oplus r)$ but $k \in L(v \oplus r)$, then $\tau^\ell(v) \leq \tau(k-1)$.

(iii) If $k \in L(v \oplus \ell)$ but $k \notin L(v \oplus r)$, then $\tau^\ell(v) \leq \tau(k)$.

(iv) If $k \notin L(v \oplus \ell)$ but $k \in L(v \oplus r)$, then $\tau^\ell(v) \geq \tau(k)$.

[19] Since the voting procedure satisfies MON, this definition is independent of the exact choice of u.

Proof.

(i) If $L(v \oplus r) = \{k+1, ...\}$, then we are done. If $L(v \oplus r) = \{..., k+1, ...\}$, then a node v' will be reached s.t. $L(v' \oplus \ell) = \{..., k\}$ and $L(v' \oplus r) = \{k+1, ...\}$. Hence, $\tau^\ell(v') = \tau(k)$. MON implies $\tau^\ell(v) \geq \tau^\ell(v')$.

(ii) Analogously.

(iii) If $L(v \oplus \ell) = \{..., k\}$, then we are done. If $L(v \oplus \ell) = \{..., k, ...\}$, then there is a node v' following $v \oplus \ell$ such that $L(v' \oplus \ell) = \{..., k\}$ and $L(v' \oplus r) = \{k+1, ...\}$. Hence, $\tau^\ell(v') = \tau(k)$. Because of MON, $\tau^\ell(v) \leq \tau^\ell(v') = \tau(k)$.

(iv) Analogously. □

Proof of Theorem 3.

Fix a preference profile $\succ$, a coalition $C \subseteq N$, and a node v. We show that there is no profitable deviation for C starting at v. This is done by proving that, for any possible deviation, the members of coalition C are weakly better off by voting sincerely at v and following specific strategies afterwards. Since this holds for any v, this implies that there is no coalition that has a profitable deviation, and hence that sincere voting is a strong ex-post perfect equilibrium.

To obtain a contradiction, suppose that C has a profitable deviation starting at v. If the decision at v was the same as under sincere voting, then the members of coalition C could, without incurring a loss, vote sincerely at v and then follow the actions prescribed by the deviation from the next node on. This holds because all other voters are assumed to vote sincerely, which is a Markovian strategy, i.e., a strategy that does not condition on the past history of play.

Therefore, there exists a subset $\tilde{C} \subseteq C$ such that sincere voting prescribes a left vote at v for $i \in \tilde{C}$, but the deviation prescribes a right vote (or vice-versa, which yields an analogous argument). Moreover, the left branch must be selected if all members of coalition $\tilde{C}$ vote left, but the right branch is selected if all members follow the deviation and vote right. Let k be the alternative chosen if coalition C plays its deviation strategy profile, while all other voters vote sincerely. Hence, k must be contained in the right branch at v. We show first that k is also contained in the left branch.

Claim 1: $k \in L(v \oplus \ell)$

We show that $k \notin L(v \oplus \ell)$ implies that the outcome after the deviation is strictly worse for some members of C, contradicting the assumption that the deviation was profitable. Let l be chosen under sincere voting. By CONV, $l < k$. Because the deviation is profitable for C, $k \succ_i l$ for all $i \in C$. Since preferences are single-peaked, this implies $l + 1 \succ_i l$ for all $i \in C$.

We now show that $\max_{SP} L(v \oplus \ell) = l$. Suppose instead that $l + 1 \in L(v \oplus \ell)$, and consider any node v' following $v \oplus \ell$ such that $l + 1 \notin L(v' \oplus \ell)$. At this node, the right branch is chosen under sincere voting because, by MON, the threshold for a consecutive left decision is weakly higher than at v, and because coalition C votes for the right branch at v'. This implies that $l + 1$ is never rejected, which contradicts the assumption that l is selected. Thus, we can conclude that $l + 1 \notin L(v \oplus \ell)$.

Because l is the largest alternative in the left branch, and because a sincere vote at v for $i \in \tilde{C}$ is to vote left, single-peaked preferences imply that $l \succ_i k$ for all $i \in \tilde{C}$.

This implies that the deviation is not profitable for $i \in \tilde{C}$, which contradicts our initial assumption. Thus, we can conclude that $k \in L(v \oplus \ell)$.

Claim 2: If coalition C votes sincerely at v, and afterwards always votes for the branch containing alternative k, then alternative k will be selected.

Because alternative k is contained in the both branches at v, the right branch must contain a strictly larger alternative. Condition CONV implies then that $k+1 \in L(v \oplus r)$. Since alternative k is selected if the right branch is chosen at v and since coalition C plays its profitable deviation, a node v' following $v \oplus r$ is reached such that the left branch is chosen at v' and $k+1 \notin L(v' \oplus \ell)$. By Lemma 1 (i), $\tau^{\ell}(v') \geq \tau(k)$. Since the left branch is chosen at node v', there are at least $\tau^{\ell}(k) - |C|$ voters in $N \setminus C$ having a peak weakly to the left of alternative k.

Since alternative k is selected if the right branch is chosen at v and since coalition C plays its profitable deviation, a node v'' following v is reached such that the right branch is chosen and $k-1 \notin L(v'' \oplus r)$. By Lemma 1 (ii), $\tau^{\ell}(v'') \leq \tau(k-1)$. Since the right branch is chosen, at most $\tau(k-1)-1$ of the voters in $N \setminus C$ have a peak weakly lower than alternative $k-1$.

We show now that, if coalition C always votes for a division containing alternative k, then k is chosen following the left branch at v as well. Consider any node v' following $v \oplus \ell$ such that k is contained in the left branch, but not in the right branch: that is, $k \in L(v' \oplus \ell)$ and $k \notin L(v' \oplus r)$. By Lemma 1 (iii), $\tau^{\ell}(v') \leq \tau(k)$. Since there are at least $\tau^{\ell}(k) - |C|$ voters in $N \setminus C$ having a peak weakly lower than alternative k, the left division is chosen if coalition C votes left.

Analogous arguments imply that the right branch is chosen whenever alternative k is only in the right branch and when C votes for the right branch. As a consequence, the branch containing alternative k is chosen at each node, and k is finally selected even if the left branch is chosen at v. To conclude, we have shown that, for any deviation of coalition C at v, the same outcome can be obtained by voting sincerely at v and following specific strategies thereafter. □

Proof of Theorem 4.

Note that, starting at an arbitrary node, sincere voting by all voters results in the election of the alternative preferred by the median type among all remaining alternatives. This holds because the procedure satisfies CONV, because simple majority is used at each stage of the voting process and because, at each stage, if the median voter votes left, all voters with smaller types will vote left as well (an analogous argument holds if the median voter votes right). Sincere voting is an ex-post perfect equilibrium if there are no profitable one-shot deviations, .

Fix now an arbitrary profile of preferences, an arbitrary node, and an arbitrary voter i. We show that voter i cannot gain by voting insincerely at this node and voting sincerely afterwards. If i's most preferred alternative is in the left branch, but the median type's most preferred alternative is in the right branch, then by the previous arguments the right branch will be chosen no matter how i votes. Analogous arguments apply if i's most preferred alternative is in the right branch, but the median type's most preferred alternative is in the left branch.

Hence, suppose the median votes left and i's sincere vote would be to vote left as well (analogous arguments apply if both vote right under sincere voting). If the median type's most preferred remaining alternative, say alternative m, is also in the right branch, a deviation is not profitable as m will be elected no matter how i votes at this node. If m is only in the left branch and i is pivotal, then CONV implies that if i deviates then a larger alternative will be elected, say alternative k.

Suppose i's type is smaller than the median type. Because the median prefers m to any larger alternative still available, and because the median's type is larger than i's type, $m \succ_i k$ follows by single-crossing. Hence, the deviation is not profitable.

Suppose instead that i's type is weakly larger than the median type. Then i is pivotal only if his type equals the median type, but then sincere voting yields his most preferred remaining alternative. Hence, there is no profitable deviation, and sincere voting is consequently an ex-post perfect equilibrium.

□

Chapter 4

Costly Verification in Collective Decisions

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.

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 **?**, 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 (**?**, **?**).[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.

? introduces costly verification in a principal-agent model. Our model differs from his, and the literature building on it (see e. g. **?**, **?**) 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 **?**. 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

[1]See also **?** for a recent extension to settings with more than two alternatives.

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. **?** 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]

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 (**?**, **?**). 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 **?** that have been introduced to the mechanism design literature by **?**.

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 mechansims and discuss their optimality, while Section 5 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 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 7 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

[2]For papers on mechanism design with evidence, see also **?**, **?**, **?**, **?**.

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 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 $t_i^- < t_i^+$ for all $t_i^- \in T_i^-$, $t_i^+ \in T_i^+$. 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 $t_i, t_i' \in T_i$

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

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

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

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 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 it 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:

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

[3]With a 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 $a_i(m)$.

s.t. (d, a) 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 $t_i \in T_i$,*

$$\inf_{t'_i \in T_i^+} \mathbb{E}_{t_{-i},s}[d(t'_i, t_{-i}, s)] \geq \mathbb{E}_{t_{-i},s}[d(t_i, t_{-i}, s)[1 - a_i(t_i, t_{-i}, s)]] \quad \text{and}$$
$$\sup_{t'_i \in T_i^-} \mathbb{E}_{t_{-i},s}[d(t'_i, t_{-i}, s)] \leq \mathbb{E}_{t_{-i},s}[d(t_i, t_{-i}, s)[1 - a_i(t_i, t_{-i}, s)] + a_i(t_i, t_{-i}, s)].$$

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 $t_i \in T_i$,

$$\mathbb{E}_{t_{-i},s}[d(t'_i, t_{-i}, s)] \geq \mathbb{E}_{t_{-i},s}[d(t_i, t_{-i}, s)[1 - a_i(t_i, t_{-i}, s)]]. \tag{1}$$

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 $t_i \in T_i$,

$$\mathbb{E}_{t_{-i},s}[d(t'_i, t_{-i}, s)] \leq \mathbb{E}_{t_{-i},s}[d(t_i, t_{-i}, s)[1 - a_i(t_i, t_{-i}, s)] + a_i(t_i, t_{-i}, s)]. \tag{2}$$

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 introduce and illustrate the class of voting-with-evidence mechanisms. To describe any voting-with-evidence 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 voting-with-evidence 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 voting-with-evidence 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 *voting-with-evidence mechanism* with cutoffs $\{\alpha_i^+, \alpha_i^-\}_{i \in \mathcal{I}}$, where $\alpha_i^- + c_i \leq \alpha_i^+ - c_i$ we first formalize the concept of altered reports. For each report t_i, the *altered report* is

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

Figure 1 illustrates how the altered reports are determined.

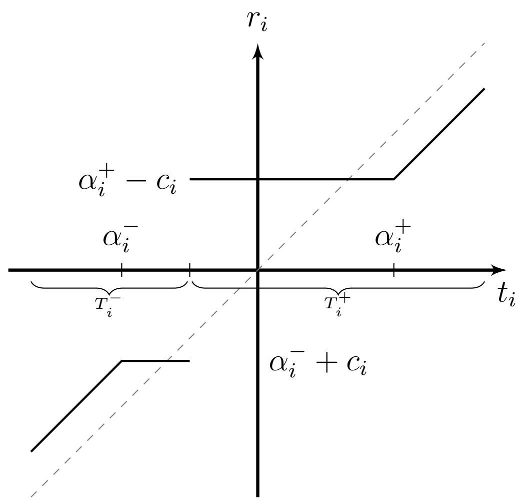


Figure 1: Example illustrating how altered reports are determined.

Given the altered reports, a voting-with-evidence 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 voting-with-evidence mechanism verifies an agent if and only if he is decisive.[4]

Remark 1 (Incentive compatibility of voting-with-evidence mechanisms)**.** We will now show that a voting-with-evidence 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 by reporting some $t_i' > t_i$ and $t_i' > \alpha_i^+$. But

[4]Our definition of a voting-with-evidence 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.

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 voting-with-evidence mechanism is *ex-post incentive compatible.*

A voting-with-evidence 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 voting-with-evidence 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.

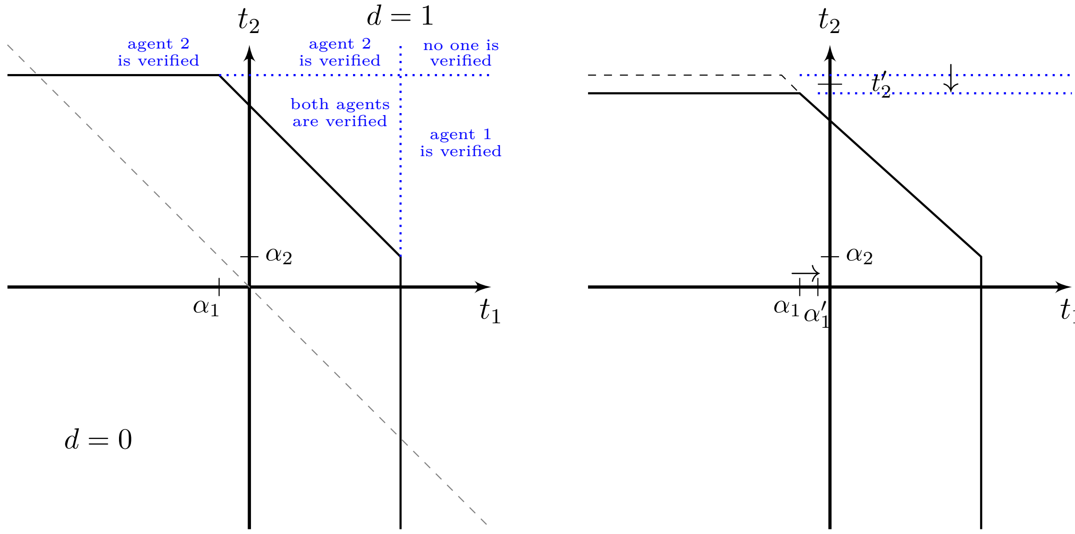
We are now ready to state our main result.

Theorem 1. *Voting-with-evidence 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 7. Before illustrating a voting-with-evidence mechanism in a two-agent example we will give an intuition for why voting-with-evidence 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 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 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. More precisely any report $t_i \in T_i^+$ above α_i^+ is replaced by the net type $t_i - c_i$. Similarly are types $t_i \in T_i^-$ 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.



(a) A voting-with-evidence mechanism.

(b) Optimal cutoffs.

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

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 voting-with-evidence mechanism.

We end this section by explaining a voting-with-evidence 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 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, 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 (illustrated in Figure 2b). 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 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 voting-with-evidence mechanism is not only Bayesian incentive compatible, it satisfies the stronger notion of ex-post incentive compatibility (see Remark 1). This robustness

[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.

of the voting-with-evidence 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 (**?** and **?**).

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 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,

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

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

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

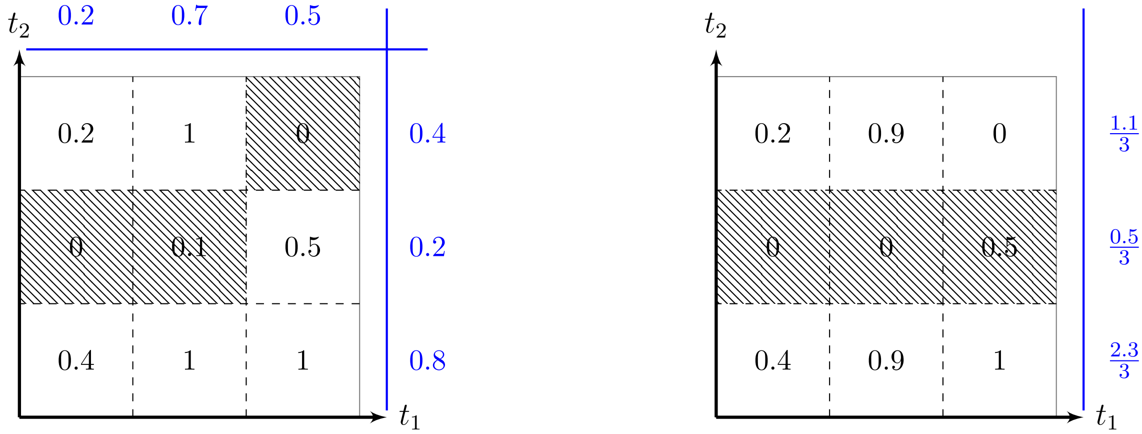
$$\inf_{t_i' \in T_i^+} \mathbb{E}_{s}[d(t_i', t_{-i}, s)] \geq \mathbb{E}_{s}[d(t_i, t_{-i}, s)[1 - a_i(t_i, t_{-i}, s)]] \quad \text{and} \tag{5}$$

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

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 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).



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

(b) Verification probabilities that are necessary for EPIC.

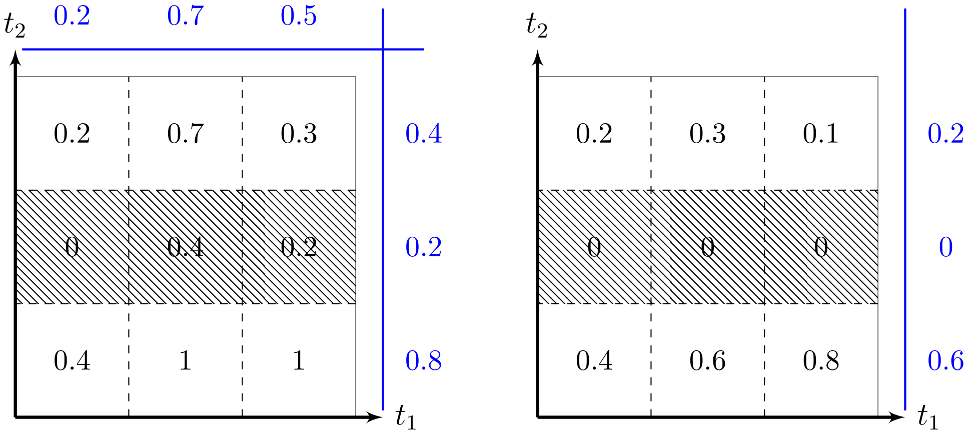
Figure 3: Failure of a naive BIC-EPIC equivalence.

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 $\mathbb{E}_{t_{-i}}\Big[\inf_{t'_i \in T_i^+} \mathbb{E}_s[d(t'_i, t_{-i}, s)]\Big]$. 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 which 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 **?**. 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



(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.

in **?** 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., $\mathbb{E}_{t_{-i}}[\inf_{t_i \in B} \hat{g}(t_i, t_{-i})] = \inf_{t_i \in B} \mathbb{E}_{t_{-i}}[\hat{g}(t_i, t_{-i})]$. The basic idea sketched above is extended via an approximation argument to a complete proof in Appendix 7.

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 $(\hat{d}, \hat{a})$; verification probabilities under this rule are therefore weakly lower than under the original rule. If they are striclty lower, one can always add additional verifications to 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 4a 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.

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 $T_i^+ = \{t_i | t_i \geq 0\}$ and $T_i^- = \{t_i | t_i < 0\}$. Consider the voting-with-evidence mechanism with cutoffs $\alpha_i^+ = 1$ and $\alpha_i^- = -1$ for all i. Let $t = (-5, 2, 2)$. Given truthful reporting the voting-with-evidence 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 voting-with-evidence mechanism specifies $d(t_1, t_2', t_3') = 1$. The voting-with-evidence mechanism is therefore not dominant-strategy incentive compatible, no matter how we specify the mechanism off-equilibrium.*

5. Proof of Theorem 1

In this section we show that a voting-with-evidence 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):

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

where $\mathbf{1}_A$ denotes the indicator function for the set A, 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)*,* $V_P(d, a) \leq V_R(d)$.

Proof.

$$V_P(d,a) = \mathbb{E}_t\left[\sum_i (d(t)[t_i - c_i(t_i)] + c_i \mathbf{1}_{T_i^+}[d(t) - a_i(t)] - c_i \mathbf{1}_{T_i^-}[d(t) + a_i(t)]\right]$$

$$\leq \mathbb{E}_t\left[\sum_i (d(t)[t_i - c_i(t_i)] + c_i \mathbf{1}_{T_i^+}[d(t)(1 - a_i(t))] - c_i \mathbf{1}_{T_i^-}[d(t)(1 - a_i(t)) + a_i(t)]\right] \tag{7}$$

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

$$= V_R(d).$$

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 follows. 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 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 voting-with-evidence mechanism.*

Proof. We assume here that T is finite. We extend this proof in Appendix 7 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:

$$\max_{0 \leq d \leq 1} \mathbb{E}_t\Big[\textstyle\sum_i d(t)[t_i - c_i(t_i)]\Big] \tag{Aux}$$

$$\text{s.t. for all } i \in \mathcal{I}:$$
$$\mathbb{E}_{t_{-i}} d(t) \geq \varphi_i^+ \text{ for all } t_i \in T_i^+, \text{ and}$$
$$\mathbb{E}_{t_{-i}} d(t) \leq \varphi_i^- \text{ for all } t_i \in T_i^-,$$

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

[7] If this condition is not satisfied, consider the sequence of problems where $\{\phi_i^+, \phi_i^-\}_i$ is replaced by

$\alpha_i^+ = \inf_{\alpha \in T_i^+}\{\alpha | \mathbb{E}_{t_{-i}}[d^*(\alpha, t_{-i})] > \varphi_i^+\}$ 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 **?**) implies that there exist Lagrange multipliers $\lambda_i^*(t_i)$, such that d maximizes

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

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.

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

Feasibility implies that

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

for all $t_i \in T_i^+$ and

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

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

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

$\{\phi_i^+ - \frac{1}{m}, \phi_i^- + \frac{1}{m}\}_i$ 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.

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

Since d^* maximizes the Lagrangian by assumption, we conclude that it is a voting-with-evidence mechanism. □

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

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

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

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

6. Relation to BDL

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. In a short note we provide an alternative proof for the optimality of threshold mechanisms. Our proof makes a

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

connection to the literature on reduced form auctions.[9] In the private good environment—the case considered by BDL—the set of feasible reduced form auctions has an explicit description (?) and a nice combinatorial structure (see e. g. ?). Using the fact that the relevant incentive constraints can be formulated in terms of reduced form auctions. We can write the optimization only in 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 (?), and we can show that threshold mechanisms are optimal. The approach to optimize directly over reduced forms, instead of the more complicated 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 (?). We had to use other tools and methods for showing that a voting-with-evidence mechanism is optimal in the collective choice environment.

Let us now look closer at the voting-with-evidence 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 voting-with-evidence 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.*

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

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 voting-with-evidence mechanism is optimal for the principal. The voting-with-evidence 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 and analyze imperfect verification in this setting, and it would also be interesting to look closer at a model with limited commitment.

Appendix

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.[10] Fix a Bayes-Nash equilibrium σ of the game induced by the indirect mechanism.[11]

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.

[10] 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.

[11] 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.

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

Let

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

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.

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 $i \in \mathcal{I}$, $n \geq 1$, $l_i = 1, \ldots, 2^{n+1}$, let

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

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

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

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

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

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

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

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

Omitted proofs from Section 4

Proof of Theorem 2.
The proof applies Theorem 6 in (**?**) to a discrete approximation of A and by taking limits we establish Theorem 2.

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

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

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 (**?**) 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

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

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

Moreover,

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

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 $0 \le \hat{g} \le 1$ and its marginals are equal almost everywhere to 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,
$\operatorname{ess\,inf}_{t_i \in B} \int_{A_{-i}} g'_n(t_i,t_{-i}) dF_{-i} \to \operatorname{ess\,inf}_{t_i \in B} \int_{A_{-i}} \hat{g}(t_i,t_{-i}) dF_{-i}$. Note further that,
$\mathbb{E}_{t_{-i}}[\inf_{t_i \in T_i^+} \hat{g}(t_i,t_{-i})] \le \inf_{t_i \in T_i^+} \mathbb{E}_{t_{-i}}[\hat{g}(t_i,t_{-i})]$ always holds. By way of contradiction suppose now that for some i,

$$\int \operatorname*{ess\,inf}_{t_i \in B} \hat{g}(t_i,t_{-i}) dF_{-i} < \operatorname*{ess\,inf}_{t_i \in B} \int \hat{g}(t_i,t_{-i}) dF_{-i}.$$ [12]

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 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

[12] 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.

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

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

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

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

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

Moreover,

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

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. □

Chapter 5

Optimal Private Good Allocation: The Case for a Balanced Budget

In an independent private value auction environment, we are interested in strategy-proof mechanisms that maximize the agents' residual surplus, that is, the utility derived from the physical allocation minus transfers accruing to an external entity. We find that, under the assumption of an increasing hazard rate of type distributions, an optimal deterministic mechanism never extracts any net payments from the agents, that is, it will be budget-balanced. Specifically, optimal mechanisms have a simple "posted price" or "option" form. In the bilateral trade environment, we obtain optimality of posted price mechanisms without any assumption on type distributions.

1. Introduction

Most parts of the mechanism design literature studying welfare maximization problems focus on mechanisms implementing the efficient allocation. However, in general it is not possible to implement the efficient allocation in dominant strategies using budget-balanced mechanisms (**?**). Given this result, we study how to choose among different mechanisms that cannot attain both, allocative efficiency and budget-balancedness. Since we are concerned with welfare maximization, the social planner's objective function should consist of the agents' aggregate utility and therefore include aggregate transfers. In other words, one seeks to find mechanisms that maximize what we call residual surplus. This is the surplus, or utility, the agents derive from the chosen physical allocation, reduced by the amount of transfers that are lost to an external agency (this is often called "money burning").

A common approach is to implement the efficient allocation via Groves mechanisms and to redistribute as much money to the agents as possible without distorting incentives (**?**, **?**, **?**, **?**). This approach aims at characterizing the optimal mechanism for allocating private goods that implements the *efficient allocation* in *dominant strategies*, is *individually rational* and *never creates a budget deficit* (ex-post).[1] However, if mechanisms that allocate inefficiently yield higher residual surplus (**?**) it is not clear why

[1] In our setting, the best Groves mechanism is implemented by a second-price auction with two bidders.

one should use a mechanism that allocates efficiently.

Consequently, we drop the requirement that mechanisms allocate efficiently. Instead, we take an optimal mechanism design approach and consider mechanisms that are comparable to the ones considered before in that they are *strategy-proof*, *deterministic*, *never run a deficit* and satisfy *ex-post participation constraints*. We analyze which mechanism maximizes residual surplus when an indivisible good is auctioned among two agents with independent private values that are distributed according to prior type distributions. We show that under an increasing hazard rate assumption on type distributions, the optimal mechanism will never waste any payments, thereby deviating distinctly from the efficient allocation (Theorem 1). In fact, our proof method reveals that all mechanisms that allocate efficiently are worse than the simple mechanism where the object is always given to the same agent (the one with the higher expected valuation; Corollary 1), showing that our general mechanism design approach has clear advantages over the previous approach to search for the optimal Groves mechanism. We show that the optimal mechanism is either a "posted price" or an "option" mechanism: The object is assigned to one of the agents unless both agents agree to trade at a prespecified price (posted price mechanism) or unless the second agent uses his option to buy the object at a fixed price from the first agent (option mechanism). Therefore, the optimal mechanisms do not invoke money burning and are of a particularly simple form. Moreover, numerical simulations indicate that these simple mechanisms obtain a large share of first-best welfare (92 per cent on average in our simulations). In the bilateral trade setting, we establish optimality of posted price mechanisms without any restrictions on type distributions (Theorem 2). This provides an argument for the focus on budget-balanced mechanisms (see **?**, **?**).

The requirement that a mechanism does not produce a budget deficit ex-post is considerably stronger than the requirement that this holds in expectation. However, in many situations it is reasonable that a budget breaker is infeasible and therefore ex-post constraints need to be obeyed. This includes situations where there is no insurance or where agents have restricted access to capital markets. Also, hidden information issues towards a third party cannot always be resolved, and autarkic mechanisms that can be implemented without explicit intervention by a third party might be preferable (e.g., when mechanisms are used to model bargaining situations (**?**, **?**)). If all these considerations do not apply and mechanisms that create no deficit in expectation can be implemented (for example, because the designer has unlimited liability), then one can achieve the first-best solution (see Section 5). Similarly, we show that one can achieve the first-best if mechanisms are only required to be Bayesian incentive compatible (Proposition 1). In contrast to these two constraints, which are the main driving forces behind our results, we argue that the participation constraint and the restriction to deterministic mechanisms are not essential to the spirit of our results (Section 5).

Our work is part of a small literature that searches for mechanisms maximizing residual surplus when the first-best is not achievable. **?** studies a model of firms colluding in a Bertrand oligopoly. A mechanism used by a cartel to allocate market shares should maximize residual surplus. Miller shows that under general conditions it is never optimal to allocate market shares efficiently and gives numerical evidence that for some type distributions it is optimal to give up efficiency in order to obtain a balanced budget.

However, other examples indicate that this observation does not hold for all distributions. ? study residual surplus maximization in a repeated bilateral trade setting and obtain numerical results suggesting that for many type distributions the optimal mechanism is a posted price mechanism. Closely related to our paper is independent work by ?, who obtain the characterization of our Theorem 1 when restricting to symmetric distributions of types and allowing mechanisms to violate individual rationality.

The result that the efficient allocation is never optimal contrasts with the literature cited above that restricts attention to efficient rules (?, ?, ?, ?). Recently, ? and ? relax this requirement. Similarly to our work, they focus on mechanisms that are deterministic, strategy-proof, ex-post individually rational and create no deficit ex-post. However, they require mechanisms in addition to be anonymous, which immediately implies that whenever the object is allocated, it is allocated to the agent that values it the most (weak assignment efficiency). This restricts the set of mechanisms severely and excludes the mechanisms that turn out to be optimal in our analysis.

The restriction to efficient allocation rules has also been relaxed in a series of papers that study specific mechanisms in a multi-unit setting. ? and ? propose simple mechanisms where one agent is designated as a residual claimant and is allocated one unit (or no unit, respectively) independent of his type. The remaining units are auctioned among the other agents and the residual claimant receives all payments accruing in the auction. Faltings uses numerical examples to argue that his mechanism often outperforms the VCG mechanism. Moreover, Moulin shows that his mechanism provides a higher worst-case welfare guarantee than any VCG mechanism given that there are sufficiently many objects and agents. In our setting with two agents and one object, these mechanisms always allocate the object to a fixed agent and therefore correspond to a degenerate option mechanism. Our Corollary 1 supports Faltings' numerical results in the two agent setting by showing that under regular prior distributions his mechanism indeed outperforms the VCG mechanism. Building on the ideas of Faltings and Moulin, ? provide worst-case welfare guarantees for two specific classes of mechanisms that allocate inefficiently: *Burning allocation mechanisms* burn a (random) number of units and assign the remaining units efficiently. *Partitioning mechanisms* partition units and agents randomly into two groups, allocate the objects in each partition efficiently to the agents in the corresponding partition and distribute the payments to agents in the other partition. Similarly, ? propose a deterministic mechanism where the burning of items is contingent on the reports of the agents; they provide worst-case welfare guarantees that converge to 0.88 asymptotically as the number of agents grows. Our work differs from these papers by evaluating mechanisms according to a Bayesian prior, restricting ourselves to the two agent setting and using a general optimal mechanism design perspective.

Another related strand of the literature studies the expected residual surplus of Bayesian incentive compatible mechanisms when it is not possible to redistribute any payments among the agents (?, ?, ?). This implies that methods similar to those in ? can be applied. It is shown that for a large class of type distributions (those which exhibit an increasing hazard rate) it is optimal to always assign the object to the same agent. Maximization of residual surplus also plays a role in the analysis of optimal mechanisms used by bidding rings (?). It is worth noting that the equivalence between

Bayes-Nash and dominant strategy implementation (**?**, **?**) does not apply to our model.[2]

We present the basic model for the auction environment in Section 2 and characterize incentive compatible mechanisms in Section 3. The optimization problem is solved in Section 4, the role of the assumptions is discussed in Section 5. We study this mechanism design problem in the bilateral trade context in Section 6, and conclude in Section 7.

2. Model

An indivisible object is auctioned among two agents. Each agent $i = 1, 2$ has a valuation x_i for the object, which is his private information. Valuations are drawn independently from $X_i = [0, \bar{x}_i]$ according to distribution functions F_i with corresponding densities f_i, which we assume to be bounded.[3] We denote by $X = X_1 \times X_2$ the product type space and by F the joint distribution on X. For notational convenience, when concentrating on agent i, we will write (x_i, x_{-i}) for $x = (x_1, x_2) \in X$.

If agent i is given a payment of p_i (usually negative), his utility is $x_i + p_i$ for winning the object, and p_i if the other agent gets the object.

Mechanisms

Due to the Revelation Principle we focus on truthfully implementable direct revelation mechanisms for selling the object.

Definition 1. *A* mechanism *M is a tuple (d, p), where $d : X \to \{0, 1\}^2$ and $p : X \to \mathbb{R}^2$ are measurable functions, such that $d_1(x) + d_2(x) = 1$.*[4]

The interpretation is that $d_i(x) = 1$ if and only if agent i gets the object. If the agents report x, then agent i receives as payment the component $p_i(x)$ of $p(x)$.

Equilibrium Concept

We consider strategy-proof mechanisms where truthful reporting is a dominant strategy for both agents. Thereby, we ensure that the mechanisms can robustly be implemented without specific assumptions on the beliefs of the agents. Hence, we define the following notion of incentive compatibility:

Definition 2. *A mechanism M is* incentive compatible (IC) *if for every agent i and for each $x_i \in X_i$, $r_i \in X_i$,*

$$d_i(x_i, r_{-i}) \cdot x_i + p_i(x_i, r_{-i}) \geq d_i(r_i, r_{-i}) \cdot x_i + p_i(r_i, r_{-i})$$

[2] See Section 5 for more details.

[3] Assuming that the lower bound of the type space is 0 simplifies the analysis. The details are explained in footnote 7.

[4] For a discussion of stochastic mechanisms, see Section 5. We follow **?** and **?** and assume that the good is always allocated. This is reasonable, for example, when considering how a cartel allocates market shares, or how the government sells licenses to firms. While there can be welfare gains from not allocating the good when one focuses on anonymous mechanisms (**?**), these gains seem to be minor in our model. Moreover, the assumption that the good is always allocated is without loss of generality in the trade setting (Section 6).

holds for each $r_{-i} \in X_{-i}$.

This definition is independent of the distribution of valuations, which reflects the robustness of strategy-proof mechanisms as compared to mechanisms that are Bayes-Nash incentive compatible. Although the set of mechanisms we consider does therefore not depend on F, the next section shows that the distributions determine which mechanism is optimal.

Objective and Further Constraints

We aim at finding the mechanism that maximizes the sum of agents' ex-ante (expected) residual surplus, that is, utility derived from the physical allocation minus aggregate payments. We impose the constraint that the mechanism has to be *ex-post no-deficit (ND)*, that is, for every type profile x, we require $p_1(x) + p_2(x) \leq 0$.[5] Also, the mechanism has to be *ex-post individually rational (IR)*, that is, for all type profiles x, we require $d_i(x)x_i + p_i(x) \geq 0, i = 1, 2$. Summarizing, we want to solve the following optimization problem:

$$\max_{M=(d,p)} \int_X \Big[d_1(x)x_1 + d_2(x)x_2 + p_1(x) + p_2(x) \Big] dF(x) \tag{1}$$
$$\text{s. t. } M \text{ satisfies IC, ND and IR.}$$

We say that a mechanism is optimal if it solves problem (1).

3. Characterization of Incentive Compatibility

The aim of this section is to give a characterization of incentive compatibility in order to simplify problem (1). The conditions characterizing incentive compatible mechanisms involve a monotonicity and an integrability condition. We first define monotonicity.

Definition 3. *The allocation function d is* monotone *if d_i is non-decreasing in x_i for $i = 1, 2$.*

Now given a monotone allocation function d, define the following functions for $i = 1, 2$:

$$g_i(x_{-i}) := \inf\{x_i : d_i(x_i, x_{-i}) = 1\}.$$

If there is no x_i such that $d(x_i, x_{-i}) = 1$, then we set $g_i(x_{-i}) = \bar{x}_i$. Note that if d is monotone, these functions define d almost everywhere. The following lemma, which is a corollary of **?**, gives a characterization of incentive compatibility.

Lemma 1. *A deterministic mechanism $M = (d, p)$ is incentive compatible if and only if the following two conditions are satisfied:*

[5] Ex-post budget constraints are commonly imposed on mechanism design problems: see, for example, the literature on optimal redistribution (**?**, **?**, **?**) and bilateral trade (**?**, **?**), or **?**. The role of this assumption is discussed in Section 5.

1. *The allocation rule d is a monotone step function.*

2. *For all $x_i \in X_i$ and $x_{-i} \in X_{-i}$,*

$$p_i(x_i, x_{-i}) = q_i(x_{-i}) - g_i(x_{-i})d_i(x_i, x_{-i}) \tag{2}$$

for some function $q_i: \ X_{-i} \to \mathbb{R}$.

The interpretation of condition (2) is that an agent who receives the object is punished by receiving a lower payment: she receives $q_i(x_{-i})$ if she does not receive the object, and this payment is reduced in case she gets the object to make the agent's marginal type $g_i(x_{-i})$ indifferent between receiving and not receiving the object.

This can be interpreted as a payoff-equivalence result: Payments are completely determined by the allocation as soon as one fixes the payment for some type x_i. Hence, the only freedom that is left regarding the payment scheme, is to give the agent an additional payment that is independent of his type. These additional payments can serve as a possibility to redistribute certain amounts of payments to another agent. Given an allocation rule d and a payment rule p, we say that the redistribution payment q implicitly defined by the above equality is associated with p.

The simplified formulation of problem (1) is the following:

$$\max_{M=(d,p)} \int_X \Big[d_1(x)[x_1 - g_1(x_2)] + d_2(x)[x_2 - g_2(x_1)] + q_1(x_2) + q_2(x_1)\Big] dF(x)$$
$$\text{s. t. } M \text{ satisfies IR and ND, } q \text{ is associated with } p \text{ and } d \text{ is monotone.}$$

We will write $U(M)$ for the above integral and from now on only consider mechanisms that are IC, IR and ND.

4. The Optimal Auction

In this section, we present the first main result of this paper: if we impose an increasing hazard rate condition on the type distributions, then the optimal mechanism is always budget-balanced. Specifically, it turns out that the optimal mechanism takes one of two simple forms:

Either it is a posted price mechanism which by default allocates the object to one of the agents (agent 1, say) and changes the allocation if and only if both agents agree to trade at a prespecified price a, i.e., agent 1 reports a valuation below a fixed price a and agent 2 reports a valuation above a. If agent 2 is allocated the object, he makes a payment a to agent 1, otherwise no transfers accrue.

Or it is an option mechanism where the good is allocated by default to agent 1, but agent 2 has the option to buy the object at price a. Hence, if agent 2's valuation is above the strike price a, he buys the object and pays a to agent 1 (see also **?**).

Formally, these two mechanisms are defined as follows:

Definition 4. *A mechanism $M = (d, p)$ is a* posted price mechanism with default agent

1 and price a, *if*

$$\begin{array}{lll} d_2(x) = 1, & p(x) = (a, -a) & \text{if } x_1 \leq a \text{ and } x_2 \geq a, \\ d_2(x) = 0, & p(x) = (0, \quad 0) & \text{otherwise.} \end{array}$$

M is an option mechanism with default agent 1 and price a, *if*

$$\begin{array}{lll} d_2(x) = 1, & p(x) = (a, -a) & \text{if } x_2 \geq a, \\ d_2(x) = 0, & p(x) = (0, \quad 0) & \text{otherwise.} \end{array}$$

Similarly, one can define posted price and option mechanisms with default agent 2. If we do not specify the agent or price we just say that M is option or posted price.

Both classes of mechanisms are parameterized by the price a and it is easy to check that all these mechanisms are budget-balanced as well as incentive compatible and individually rational.

Our assumption on type distributions is the following:

Condition 1 (HR)**.** *The hazard rates of the type distributions are monotone. That is, the functions $h_i(x_i) = \frac{f_i(x_i)}{1-F_i(x_i)}$ are non-decreasing in $x_i \in [0, \bar{x}_i)$ for $i = 1, 2$.*

Theorem 1. *Suppose that the hazard rates of the type distributions are monotone. Then the optimal mechanism is either a posted price or an option mechanism.*

It is known that if payments are wasteful by assumption, then for regular distributions it is optimal to make the allocation independent of reports (**?**): a more efficient allocation is more than offset by the waste of payments that are required for incentive-compatibility. Given that money can be redistributed in our model, there are better budget-balanced mechanisms (essentially posted price and option mechanisms). One might argue naively that, if wasting money is suboptimal in the setting of (**?**), it must also be suboptimal in this setting, and hence a budget-balanced mechanism must be optimal. However, redistribution payments allow for additional flexibility, which makes the argument more subtle and requires that we optimize jointly allocations and redistribution payments.[6]

The proof can be sketched as follows: We first show the important auxiliary result that either an option mechanism or a posted price mechanism is optimal in $\mathcal{M}_0$, the class of mechanisms such that g_i is monotone and piecewise constant for each agent (Lemma 2). We then argue that the residual surplus $U(M)$ of a given mechanism M can be approximated arbitrarily well by a mechanism in $\mathcal{M}_0$ (Lemma 3). The Theorem then follows by the following observation: Suppose there is a mechanism $\bar{M}$ being strictly better than the best option or posted price mechanism, and denote the difference in residual surplus by ε. It follows from Lemma 3 that there is a mechanism in the class $\mathcal{M}_0$ whose residual surplus is within $\frac{\varepsilon}{2}$ of $U(\bar{M})$, thus being better than

[6]Indeed, if the arguments from a model without redistribution could simply be extended, our conclusion would also hold for stochastic mechanisms. However, numerical results in Section 5 show that this is not the case.

the best option or posted price mechanism. But this contradicts Lemma 2, hence there cannot be a mechanism being better than the best option or posted price mechanism.

While the approximation part of the proof can be found in the Appendix, we state and prove Lemma 2, which contains the essence of why Theorem 1 holds.

Lemma 2. *Suppose that the hazard rates of the type distributions are monotone and let $M = (d, p)$ be any mechanism in $\mathcal{M}_0$. Then there exists a mechanism M' that is posted price or option such that $U(M') \geq U(M)$.*

Proof. The proof consists of three steps: *Step 1* determines for an arbitrary allocation rule the maximal possible redistribution payments q_i. Hence, the allocation rule from this point on completely determines the optimal payments and we can constructively manipulate the allocation rule in *Steps 2* and *3* until we end up with an option or posted price mechanism.

Step 1: We denote the jump points of $g_2(x_1)$ and $g_1(x_2)$ by α_j and β_j, respectively (see Figure 1). Note that, without loss of generality, we can assume that for the first segment of g_1 we have $g_1(x_2) = 0$ since otherwise we could switch the roles of the agents.

We now claim that $q_2(x_1) = 0$, $\forall x_1 \in X_1$; that is, no money is redistributed to agent 2. To see this, pick arbitrary x_1 and observe that $g_1(0) = 0$[7] and $d_2(x_1, 0) = 0$; therefore $g_1(0)d_1(x_1, 0) = g_2(x_1)d_2(x_1, 0) = 0$. From (ND) it follows that $q_1(0)+q_2(x_1) = p_1(x_1, 0) + p_2(x_1, 0) \leq 0$. Also, (IR) for agent 2 at $(x_1, 0)$ implies $q_2(x_1) \geq 0$, and (IR) for agent 1 at $(0, 0)$ implies $q_1(0) \geq 0$, and therefore $q_2(x_1) = 0$.

Next, we can assume that

$$q_1(x_2) = \min_{x_1} \Big\{ g_1(x_2)d_1(x_1, x_2) + g_2(x_1)d_2(x_1, x_2) \Big\} \tag{3}$$

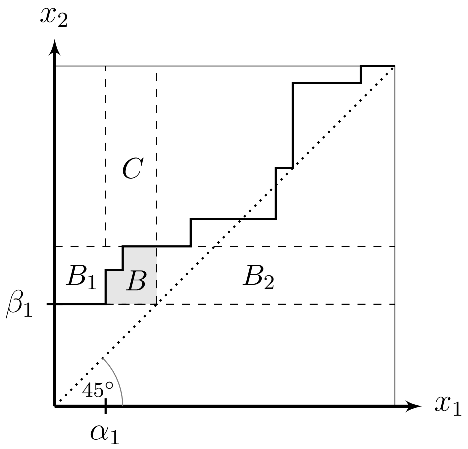
always holds, since by (ND) this relation always holds with $\leq$ and changing it to equality does not reduce $U(M)$. In this way, the complete payment-scheme is determined through the allocation rule d. Note that setting the function q this way implies that (ND) and (IR) are always satisfied.

Step 2: In this step we argue that changing the allocation to the one shown in Figure 1b does not increase money burning, but increases allocative efficiency and hence aggregate welfare.

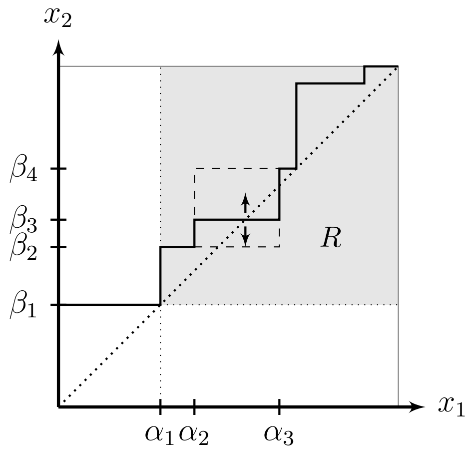
Define the set $B = \{x \mid x_1 \leq \beta_1 \leq x_2, d_2(x) = 0\}$ and consider the sets B_1, B_2 and C as shown in Figure 1a. We change the allocation rule and allocate the object to agent 2 for types in B. Since $x_2 \geq x_1$ for $x \in B$, this improves the physical allocation and we can concentrate on payments. Note that q_1, as defined in (3), increases to the same extent as g_1, hence any additional payments in the set B_2 can be redistributed. Also, transfers are weakly increased for types in B_1 and C. As the change in allocation has no effect outside these sets, the claim follows.

Step 3: This step studies the effects of shifting steps in the set R, shown as the shaded area in Figure 1b, while fixing redistribution payments. Our condition on the

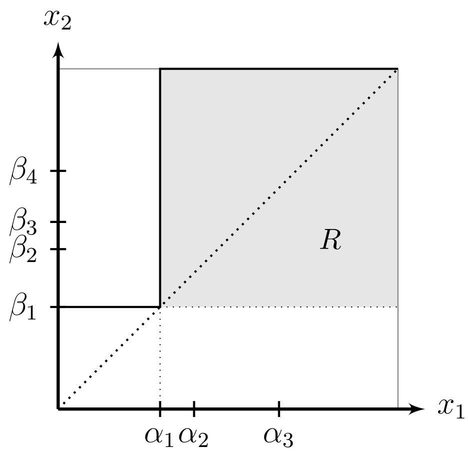
[7] At this step we use that 0 is the lower bound of the type space. This assumption implies that participation constraints pin down the maximal redistribution payments. Without this assumption, one would have to optimize over redistribution payments. If, for example, $f_1 = f_2$ and f_i is log-concave then exactly the same results can be obtained.



(a) Illustration of the sets used in Step 2 of the proof.



(b) Situation after the object is assigned to agent 2 in the set B. The dashed lines indicate the interval in which the jump point in Step 3 is varied.



(c) A posted price mechanism with default agent 1 and price β_1 improves over M.

Figure 1: Illustration of the proof of Lemma 2.

hazard rate ensures that each step should optimally be moved to either the lowest or the highest possible position. Hence, proceeding iteratively, we obtain either an option mechanism or a posted price mechanism. This will complete the proof.

Changing the allocation in R does not change q_1 as defined in (3) and we ignore the functions q_i from now on.

The following is a procedure to remove one step contained in R without decreasing $U(M)$. We do this exemplarily with the jump point at β_3 (see Figure 1b). We vary β_3 on the interval $[\beta_2, \beta_4]$ and show that welfare is quasi-convex in β_3. This implies that setting $\beta_3^* = \beta_2$ or β_4 increases $U(M)$. The part of $U(M)$ that depends on β_3 is the following:

$$\int_{\alpha_2}^{\alpha_3} \left[\int_{\beta_2}^{\beta_3} (x_1 - \alpha_2) dF_2(x_2) + \int_{\beta_3}^{\bar{x}_2} (x_2 - \beta_3) dF_2(x_2) \right] dF_1(x_1) \\ - \int_{\alpha_3}^{\bar{x}_1} \left[\int_{\beta_2}^{\beta_3} \alpha_2 dF_1(x_1) + \int_{\beta_3}^{\beta_4} \alpha_3 dF_2(x_2) \right] dF_1(x_1)$$

Differentiating with respect to β_3 using Leibniz' rule yields

$$\int_{\alpha_2}^{\alpha_3} \Big[f_2(\beta_3)(x_1 - \alpha_2) - \big[1 - F_2(\beta_3)\big] \Big] dF_1(x_1) + \int_{\alpha_3}^{\bar{x}_1} f_2(\beta_3)[\alpha_3 - \alpha_2] dF_1(x_1).$$

Writing constants C_1, C_2 and C_3 for the terms that do not depend on β_3, we get

$$C_1 f_2(\beta_3) - C_2 \big[1 - F_2(\beta_3)\big] + C_3 f_2(\beta_3).$$

Assuming $C_2[1 - F_2(\beta_3)] > 0$ (if either $C_2 = 0$ or $1 - F_2(\beta_3) = 0$, we set $\beta_3^* = \beta_4$ without reducing U), we can divide by $C_2[1 - F_2(\beta_3)]$ and get that the derivative is non-negative if and only if

$$C \cdot h_2(\beta_3) - 1 \geq 0,$$

where $C = (C_1 + C_3)/C_2 > 0$. Because $h_2(\beta_3)$ is non-decreasing by condition (HR), quasi-convexity follows and $U(M)$ is increased by either setting $\beta_3^* = \beta_2$ or $\beta_3^* = \beta_4$. In either case, we have decreased the number of steps by one and the procedure ends.

Iteratively applying this procedure establishes the lemma. □

A consequence of the theorem is that, given the increasing hazard rates of the agents' type distributions, finding the best mechanism reduces to finding the best posted price and option mechanisms and comparing these two. For example, if the agents have the same distribution function, all option and posted price mechanisms with the same strike price yield the same welfare and therefore the best mechanism is characterized by the strike price a^* satisfying

$$a^* = \mathbb{E}[x_1] = \mathbb{E}[x_2].$$

Our intermediate results (see the proof of Lemma 2) also allow for a refined judgment of the welfare implied by the efficient allocation, which is employed by the literature on optimal redistribution (**?**, **?**, **?**, **?**). We provide a mechanism that improves upon

Distribution	Average share	Minimum share
Weibull (IHR)	0.930 %	0.788 %
Weibull (DHR)	0.916 %	0.753 %
Exponential	0.905 %	0.752 %

Table 1: Simulation results showing the share of first-best welfare that is obtained by the optimal posted price or option mechanism. We solved a discretized model and ran 500 trials for each distribution using randomly drawn parameters.

all efficient mechanisms.[8] Surprisingly, this improvement can be achieved using an extremely simple mechanism:

Corollary 1. *If the hazard rates of the type distributions are monotone, then every mechanism that allocates efficiently is dominated by a mechanism that always allocates the good to the same agent.*

More precisely, a mechanism that is better than every efficiently allocating mechanism can be found simply by comparing the agents' type distributions, giving the good to the agent with the higher expected valuation and completely ignoring any reported types.

Despite their simplicity, the optimal mechanisms obtain a surprisingly large share of first-best welfare, as the following example suggests (see Table 1 for further numerical estimates of the share of first-best welfare that the optimal mechanism obtains). Note that randomly allocating the object to one of the agents provides a worst-case welfare guarantuee of $\frac{1}{2}$; in all our numerical examples the optimal posted price or option mechanism improves significantly over this lower bound.

Example 1. *Suppose that $\theta_i \sim U[0,1]$ for $i = 1,2$. First-best welfare is given by $U_{FB} = \frac{2}{3}$, whereas the optimal mechanism M is an option mechanism with price $\frac{1}{2}$, yielding $U(M) = \frac{5}{8}$. Hence, the optimal mechanism yields a 93.8 per cent share of first-best welfare. In contrast, a random allocation yields only a 75 per cent share of first-best welfare.*

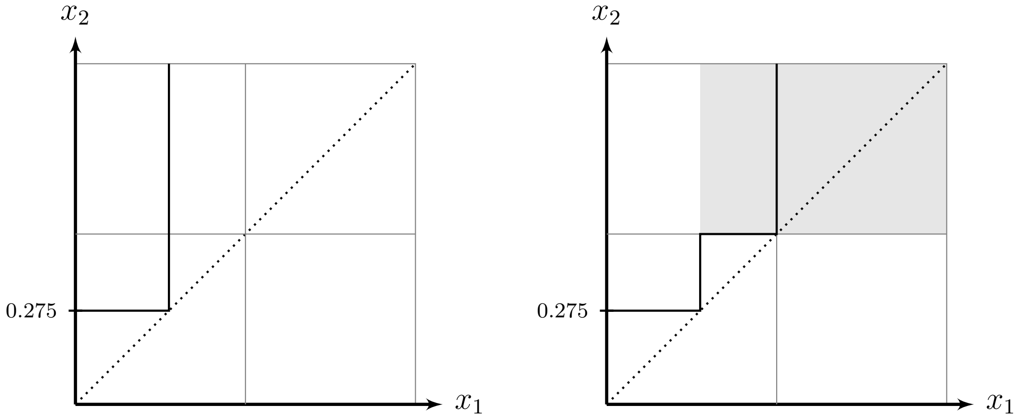
The following example shows that if Condition (HR) is not satisfied the optimal mechanism need not be of the form stated in Theorem 1. The example also illustrates the role of (HR) in establishing the result.

Example 2. *Let the distribution function of two symmetric agents be given as*

$$f(x_i) = \begin{cases} 0.9 & \text{if } x_i \leq 0.5 \\ 0.1 & \text{otherwise.} \end{cases}$$

Due to the downwards jump at 0.5, f does not satisfy condition (HR). The optimal posted price mechanism (which is as good as the optimal option mechanism) has a strike price of $a^* = 0.275$, attaining a social welfare of 0.0718. However, the following

[8] Indeed, this mechanism improves upon any mechanism that treats agents symmetrically in a neighborhood of 0. This observation extends to settings with more than two agents.



(a) Optimal posted price mechanism.

(b) Mechanism M.

Figure 2: Mechanisms presented in Example 2

mechanism M attains a higher social welfare of 0.0741*: Set*

$$d_2(x) = 1 \quad \Leftrightarrow \quad (x_2 \geq a^* \text{ and } x_1 \leq a^*) \text{ or } (x_2 \geq 0.5 \text{ and } x_1 \leq 0.5),$$

and set $q_2(x_1) \equiv 0$, as well as

$$q_1(x_2) = \begin{cases} 0 & \text{if } x_2 \leq a^* \\ a^* & \text{otherwise.} \end{cases}$$

This mechanism and the best option mechanism are depicted in Figure 2. One can see that the allocation of mechanism M is more efficient. Because the induced higher payments cannot be redistributed, payments of $(0.5 - 0.275) = 0.225$ *are lost for type profiles in the shaded area in Figure 2b. But still, since type profiles x with $x_1, x_2 \geq 0.5$ appear so rarely (with density* 0.01*), this does not counter the positive effect due to the better allocation. In this sense, an increasing hazard rate ensures that lost payments can never be weighed out by an improved efficiency of the allocation.*

5. Relaxing the constraints

In this section, we in turn relax the no-deficit, incentive compatibility and participation constraints as well as the restriction to deterministic mechanisms, and analyze how sensitive our characterization in Theorem 1 is to these relaxations.

Ex-ante Budget Constraints

While ex-post budget constraints are imposed commonly in the literature and seem appropriate for many settings, they would effectively be turned into ex-ante constraints if insurance against budget deficits was available.[9] Relaxing the no-deficit constraint to an ex-ante constraint, thus requiring the mechanism to run at no deficit on average, simplifies the problem and allows the planner to implement the first-best (i. e., the efficient allocation and a balanced budget). This can be achieved by running the VCG mechanism. This mechanism is ex-post individually rational and creates no deficit ex-post. By redistributing the expected surplus in an arbitrary fixed way to the agents, the mechanism becomes ex-ante budget-balanced and therefore achieves the first-best.

Bayesian Incentive Compatible Mechanisms

If stronger assumptions can be made on the information structure (namely, if the agents' beliefs equal a common prior that is known to the designer), we can relax the constraints on the mechanisms to Bayesian incentive compatibility and interim individual rationality. This allows the implementation of mechanisms that achieve higher expected welfare. Notably, if the distribution of types is symmetric across agents, then the expected externality mechanism (**?**, **?**) achieves the first-best. To see this, observe that this mechanism allocates efficiently, has a balanced budget, and has payments given by

$$t_i^{EEM}(x) = \int_{x_i}^{\overline{x}_{-i}} x_{-i} dF_{-i}(x_{-i}) - \int_{x_{-i}}^{\overline{x}_i} x_i dF_i(x_i). \tag{4}$$

Therefore, an agent reporting a type of 0 receives a weakly positive transfer and hence a weakly positive utility. The following Proposition shows that with ex-ante symmetric agents, the expected externality mechanism is even ex-post individually rational. More generally, it shows that the first-best can be achieved whenever virtual values are increasing (in particular, under condition (HR)).

Proposition 1. *Consider the problem of finding the optimal mechanism that is Bayesian incentive compatible, interim individually rational and satisfies ex-post no-deficit.*

1. *If virtual valuations are increasing, (i. e., $x_i - \frac{1-F_i(x_i)}{f_i(x_i)}$ is increasing for $i = 1, 2$), then the optimal mechanism allocates efficiently and is ex-post budget-balanced.*

2. *If agents are ex-ante symmetric (i. e., $F_1 \equiv F_2$), then the expected externality mechanism is optimal. It allocates efficiently, is ex-post budget-balanced, and ex-post individually rational.*

This implies that the equivalence of Bayesian and dominant strategy incentive compatible mechanisms established by **?** does not apply. They show that in a large class of mechanism design problems, for any Bayesian incentive compatible and interim individually rational mechanism, there exists an equivalent dominant strategy incentive compatible mechanism that is ex-post individually rational. However, this equivalence is established in the absence of budget constraints, and the above arguments imply that it cannot be extended to our setting.

[9] Note also, that the exact form of the budget constraints can be irrelevant when considering Bayesian incentive compatible mechanisms (**?**).

Distribution	Average loss	Maximum loss	Instances without loss
Random	0.018 %	0.874 %	92.500 %
IHR	0.003 %	0.420 %	97.850 %
Weibull	0.000 %	0.000 %	100.000 %
All	0.007 %	0.874 %	96.785 %

Table 2: Simulation results comparing the welfare loss due to the restriction to deterministic mechanisms.

Participation Constraints

While our general characterization of the optimal mechanism does not hold with relaxed participation constraints, these constraints are not the main driving force behind our results and the inefficiency of the optimal allocation. Indeed, our characterization can be obtained without participation constraints if one restricts attention to settings where agents are symmetric ex-ante (**?**).

Stochastic Mechanisms

In the previous section we restricted attention to deterministic mechanisms in order to be able to analytically characterize the optimal mechanism. Deterministic mechanisms have additional benefits: They are simpler to implement, and more plausible in some settings (e.g., when modeling bargaining between agents).

While there are instances where the focus on deterministic mechanisms is not without loss, numerical simulations suggest that the induced loss in welfare is small. We generated $n = 2\,000$ random instances for three classes of distributions of types: Random distributions, random distributions with an increasing hazard rate, and distributions from the Weibull class with different shape and scale parameters such that the distribution has an increasing hazard rate. We then computed the optimal deterministic and stochastic mechanism for every instance. The results are summarized in Table 2 which shows, for each distribution class, the average and maximum welfare loss of the optimal deterministic mechanism, as a percentage of the welfare of the best stochastic mechanism. The fourth column shows the percentage of instances where there is no loss due to the restriction to deterministic mechanisms. As can be seen, instances where the deterministic constraint is binding appear only rarely. Further, even if this is the case, the percentage loss in expected welfare is very small. Note that whenever a stochastic mechanism is strictly better in our simulations, the optimal mechanism is not budget-balanced.

6. Bilateral Trade

? showed that one cannot implement the efficient allocation in the bilateral trade setting in an ex-post budget-balanced and interim individually rational way, and characterized the optimal mechanism satisfying these constraints. In the same environment, **?** study the set of dominant-strategy implementable mechanisms that are ex-post budget-balanced and individually rational, showing that essentially only posted price mechanisms fulfill these conditions. However, a priori it is not clear why one should restrict the search for the optimal mechanism to mechanisms with a balanced budget. After all, it is conceivable that deviating from a balanced budget could improve incentives and therefore lead to higher welfare. In fact, **?** show by example that relaxing budget-balancedness to a no-deficit constraint can improve upon posted price mechanisms. The result in this section shows that this holds only for stochastic mechanisms; when looking at deterministic mechanisms, the restriction to budget-balanced mechanisms does not reduce aggregate welfare.

Let the model and notation be as in Section 2, but assume now that agent 1 (called the "seller" from now on and indexed by S) is the owner of the good before participating in the mechanism (whereas agent 2 is called the "buyer" and indexed by B). By a buyer posted price mechanism we denote a posted price mechanism in which the buyer gets the object if and only if he announces a type high enough, and the seller a type that is low enough. Again, we are looking for a mechanism that maximizes the sum of the expected utilities of the agents, taking monetary transfers into account. The fact that in the bilateral trade setting the seller initially owns the good requires a stronger condition for a mechanism to be individually rational: now the outside option for a seller is to not participate in the mechanism and to keep the object. Hence, for a mechanism to be individually rational,

$$d_S(x)x_S + p_S(x) \geq x_S \qquad \text{and} \qquad d_B(x)x_B + p_B(x) \geq 0 \tag{IR'}$$

must hold for all $x \in X$.

Thus, a mechanism is optimal if it solves

$$\max_{M=(d,p)} \int_X \Big[d_S(x)x_S + d_B(x)x_B + p_S(x) + p_B(x) \Big] dF(x) \tag{5}$$
$$\text{s. t. } M \text{ satisfies IC, ND and IR'.}$$

Theorem 2. *There is a buyer posted price mechanism that solves problem* (5)*.*

Proof. We first show that (IR') implies that the seller keeps the object whenever his valuation is higher. Assume to the contrary that trade takes place at $x_S > x_B$. Then (IR') for the seller implies that the seller receives at least x_S and (IR') for the buyer implies that he pays at most x_B, violating (ND).

Recall that $g_B(0)$ denotes the smallest buyer type such that trade takes place when $x_S = 0$, and $g_S(\overline{x}_B)$ denotes the highest seller type such that trade takes place when $x_B = \overline{x}_B$. We claim that $g_S(\overline{x}_B) \leq g_B(0)$. Constraints (IC) and (IR') for the seller imply that he receives at least a payment of $g_S(\overline{x}_B)$ whenever the buyer reports $\overline{x}_B$ and trade takes place, in particular at $(0, \overline{x}_B)$ (if no trade takes place at $(0, \overline{x}_B)$, trade will never happen, corresponding to a posted price mechanism with a price above the highest possible valuation). Similarly, (IC) and (IR') for the buyer imply that he pays at most $g_B(0)$ whenever the seller reports 0, in particular at $(0, \overline{x}_B)$. Therefore, $g_S(\overline{x}_B) > g_B(0)$ would violate (ND) at $(0, \overline{x}_B)$.

Finally, we claim that the buyer posted price mechanism with strike price $g_B(0)$

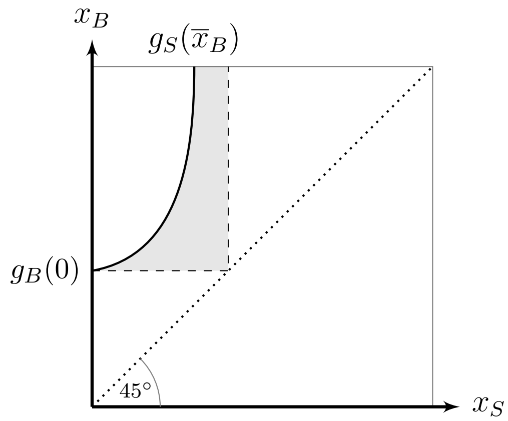


Figure 3: Illustration of the proof of Theorem 2. The shaded area indicates the type profiles where the initial mechanism differs from the posted price mechanism with strike price $g_B(0)$ (dashed line).

weakly dominates the given mechanism. To see this, note that $p_S(x) + p_B(x) \leq 0$ by (ND) and a posted price mechanism is budget-balanced. Hence, the posted price mechanism dominates the old mechanism with respect to payments. Since the allocation only differs for x such that $x_B \geq g_B(0) \geq x_S$ and the posted price allocation rule prescribes $d_B(x) = 1$ for such x (see also Figure 3), the posted price mechanism also dominates the old mechanism with respect to the allocation rule. □

In contrast to Theorem 1, this result shows that a posted price mechanism is optimal for any type distribution. The difference is due to the stronger individual rationality constraint. While any allocation rule is compatible with (IR), the stronger constraint (IR') in the trade setting restricts the set of allocation rules that can be implemented without a budget deficit. Within this smaller class of feasible allocation rules, for any distribution of types a posted price mechanism is optimal.

The stronger individual rationality constraint also implies that mechanisms which do not allocate the object are infeasible. This is because if the buyer does not get the object, no money can be collected to compensate the seller for losing the object. Therefore, assuming that the object is always allocated is without loss of generality in this setting.

7. Discussion

We have studied the trade-off between efficiency and budget-balancedness in an independent private values auction model. We incorporated this into the model by letting the social welfare objective function include all payments, that is, by maximizing residual surplus.[10] We showed that, if one focuses on robust implementation in dominant strategies, an increasing hazard rate condition on agents' type distributions guarantees a resolution of the trade-off completely in favor of a balanced budget. In addition, budget-balanced mechanisms have a very simple form and can easily be implemented

[10] For other ways to analyze the frontier that describes possible ways to resolve the trade-off between efficiency and budget-balancedness, see, for example, **?** or **?**.

as posted price or option mechanisms. Further, we showed without any assumption on the prior distribution of types that a posted price mechanism is optimal in the bilateral trade setting. Our results imply that our approach of optimal mechanism design yields higher welfare than approaches concentrating on the efficient allocation.

In the section on robustness we have seen that the restriction to deterministic and ex-post individually rational mechanisms is not crucial for our main result. Instead, it is primarily driven by the focus on strategy-proof mechanisms that satisfy the ex-post no-deficit constraint: Without these constraints the first-best can be achieved, implying that these two restrictions are relatively costly in terms of welfare.

An interesting open question is how the result generalizes to a model including more than two agents. We strongly believe that the optimal mechanism will still be budget-balanced. An important argument for this is that, as the number of agents gets large, the efficient allocation can be approximated in a budget-balanced way: in the spirit of **?**, allocate efficiently while ignoring one agent who then receives all payments from the other agents. This can be implemented by tentatively giving the object to one of the agents and then simulating a second price auction with reserve price where this agent sells the object to the remaining agents.

Appendix: Proof of Theorem 1

The following lemma enables us to approximate any mechanism with mechanisms from the class $\mathcal{M}_0$.

Lemma 3. *For every mechanism $M = (d, p)$ and for every $\varepsilon > 0$ there exists a mechanism $\tilde{M} = (\tilde{d}, \tilde{p})$ in $\mathcal{M}_0$ such that $U(M) - U(\tilde{M}) < \varepsilon$.*

Proof. Let the mechanism $M = (d, p)$ and $\varepsilon > 0$ be given and let $g_1(x_2)$ and $g_2(x_1)$ be defined as above. Define $D_i := \{x \in X : d_i(x) = 1\}$ as the set of type profiles where agent i gets the object and define $\tilde{D}_i$ similarly. Since g_2 is a monotone function it can be approximated uniformly by a monotone and piece-wise constant function $\tilde{g}_2$. Denote the associated allocation rule by $\tilde{d}$. By choosing the step width small enough the approximation can be done such that for given $\delta > 0$,

$$||g_1 - \tilde{g}_1||_\infty < \delta \quad \text{and} \quad ||g_2 - \tilde{g}_2||_\infty < \delta$$

holds. The approximation can be chosen such that $g_i(x_{-i}) = \bar{x}_i$ implies $\tilde{g}_i(x_{-i}) = \bar{x}_i$ and $\tilde{g}$ can be chosen such that $\tilde{g}_2 \leq g_2$, implying that $\tilde{D}_1 \subset D_1$.

Without loss of generality, we can assume that $q_2(x_1) \equiv 0$ (see Step 1 in the proof of Lemma 2). By construction of $\tilde{g}_2$ and since M satisfies (ND), we can define functions $\tilde{q}_i(x_{-i})$ such that $\tilde{q}_2(x_1) \equiv 0$, $0 \leq \tilde{q}_1(x_2) \leq \inf_{x_1}\{\tilde{g}_1(x_2)\tilde{d}_1(x_1, x_2) + \tilde{g}_2(x_1)\tilde{d}_2(x_1, x_2)\}$ $\forall x_2 \in X_2$ and $||\tilde{q}_1 - q_1||_\infty < \delta$. We then have:

$$\begin{aligned} U(d,p) - U(\tilde{d},\tilde{p}) \leq \int_X q_1(x_2) - \tilde{q}_1(x_2)\ dF(x) \\ + \int_{D_1} x_1 - g_1(x_2)\ dF(x) - \int_{\tilde{D}_1} x_1 - \tilde{g}_1(x_2)\ dF(x) \\ + \int_{D_2} x_2 - g_2(x_1)\ dF(x) - \int_{\tilde{D}_2} x_2 - \tilde{g}_2(x_1)\ dF(x) \end{aligned}$$

$$\leq \delta + \int_{D_1 \setminus \tilde{D}_1} x_1 - g_1(x_2)\ dF(x) + \int_{\tilde{D}_1} \delta\ dF(x)$$
$$+ \int_{\tilde{D}_2 \setminus D_2} x_2 - g_2(x_1)\ dF(x) + \int_{D_2} \delta\ dF(x)$$
$$\leq 3\delta + B_1 \overline{x}_1 \delta + B_2 \overline{x}_2 \delta,$$

where B_i is an upper bound for $f_i(x_i)$. Hence, by choosing $\delta < \frac{\varepsilon}{3+B_1\overline{x}_1+B_2\overline{x}_2}$, it follows that $U(d,p) - U(\tilde{d},\tilde{p}) < \varepsilon$. □

We combine the approximation lemma with Lemma 2 in order to prove the theorem.

Proof of Theorem 1. Without loss of generality, we restrict ourselves to posted price mechanisms for agent 2. We first establish that U maps the set of all posted price mechanisms to a compact subset of $\mathbb{R}$. Let $\bar{a} = \min\{\bar{x}_1, \bar{x}_2\}$ and let $a \in [0, \bar{a}]$ be some price for a posted price mechanism M_a. Then $U(M_a)$ can be written as

$$U(M_a) = \int_0^a \int_a^{\bar{x}_2} x_2 dF(x) + \int_0^{\bar{x}_1} \int_0^a x_1 dF(x) + \int_a^{\bar{x}_1} \int_a^{\bar{x}_2} x_1 dF(x).$$

Due to the continuity of F, this function is continuous with respect to a. Since $[0, \bar{a}]$ is compact, so is $\{U(M_a) \mid a \in [0, \bar{a}]\}$ and therefore there exists an a^* such that $U(M_{a^*})$ is maximal among all posted prices.

Next, assume that the theorem is false, i.e., there exists a mechanism M and $\varepsilon > 0$ such that $U(M) > U(M_{a^*}) + \varepsilon$. Then apply Lemma 3 to M and ε to get a mechanism $\tilde{M} \in \mathcal{M}_0$ with $U(\tilde{M}) > U(M_{a^*})$. This contradicts Lemma 2, establishing the theorem. □

Proof of Corollary 1. The arguments in Step 1 in the proof of Lemma 2 imply that agent 1 receives no redistribution payments; symmetric arguments imply that agent 2 also gets no redistribution payments. Hence, all payments that are collected must be wasted. The result then follows from **?**. □

Proof of Proposition 1.

(1) Let d^* denote an efficient allocation rule. Given a mechanism (d^*, t), let

$$U_i(x_i) := \int_{\underline{x}_{-i}}^{\overline{x}_{-i}} x_i \cdot d_i^*(x_i, x_{-i}) + t_i(x_i, x_{-i}) dF_{-i}(x_{-i})$$

denote the interim expected utility, and $S := -\mathbb{E}[t_1(x_1, x_2) + t_2(x_1, x_2)]$ the expected budget surplus. Observe that

$$\int_X x_1 \cdot d_1^*(x_1, x_2) + x_2 \cdot d_2^*(x_1, x_2) dF(x_1, x_2) - S$$
$$= \int_X U_1(x_1) + U_2(x_2) dF(x_1, x_2)$$
$$= U_1(0) + U_2(0) + \int_X \int_0^{x_1} d_1^*(s, x_2) ds + \int_0^{x_2} d_2^*(x_1, s) ds\ dF(x_1, x_2)$$
$$= U_1(0) + U_2(0) + \int_X \frac{1 - F_1(x_1)}{f_1(x_1)} d_1^*(x_1, x_2) + \frac{1 - F_2(x_2)}{f_2(x_2)} d_2^*(x_1, x_2)\ dF(x_1, x_2).$$

Hence,

$$U_1(0) + U_2(0) + S = \int_X \left[x_1 - \frac{1 - F_1(x_1)}{f_1(x_1)}\right] d_1^*(x_1, x_2) \\ + \left[x_2 - \frac{1 - F_2(x_2)}{f_2(x_2)}\right] d_2^*(x_1, x_2) dF(x_1, x_2).$$

We claim that

$$U_1(0) + U_2(0) + S \geq 0. \tag{6}$$

Indeed, if $U_1(0) + U_2(0) + S < 0$ were true, then

$$\int_X \left[x_i - \frac{1 - F_i(x_i)}{f_i(x_i)}\right] d_i^*(x_1, x_2) dF(x) < 0$$

would hold for some i, as $S \geq 0$ follows from no-deficit. Together with the assumption that virtual valuations are increasing, this would imply that

$$\int_X \left[x_i - \frac{1 - F_i(x_i)}{f_i(x_i)}\right] [1 - d_i^*(x_1, x_2)] dF(x) < 0.$$

Hence

$$\int_X \left[x_i - \frac{1 - F_i(x_i)}{f_i(x_i)}\right] dF(x) < 0,$$

contradicting the fact that $\int_{X_i} x_i - \frac{1-F_i(x_i)}{f_i(x_i)} dF_i(x_i) = 0$.

Define $t_1(x) := t_1^{EEM}(x) - \int_{X_2} t_1^{EEM}(0, s) dF_2(s)$ and $t_2(x) := -t_1(x)$. The mechanism (d^*, t) is ex-post budget balanced by construction (hence, $S = 0$); it is Bayesian incentive compatible since payments differ from the payments in the expected externality mechanism only by a constant. Moreover, $U_1(0) = 0$ by construction; therefore, $U_2(0) \geq 0$ follows from (6), showing individual rationality.

(2) Optimality of the expected externality mechanism follows from the observations before Proposition 1. Ex-post individual rationality follows from the following two facts: $x_i \leq x_j$ implies $t_i^{EEM}(x) \geq 0$ by (4), and $x_i > x_j$ similarly implies $t_i^{EEM}(x) = -\int_{x_j}^{x_i} s dF_i(s) \geq -x_i$. This implies that $x_i \cdot d_i^*(x) + t_i^{EEM} \geq 0$ for all x. □