Essays on Mechanism Design without Transfers

Inauguraldissertation

zur Erlangung des Grades eines Doktors
der Wirtschaftswissenschaften

durch

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

vorgelegt von

Axel Niemeyer

aus Stendal

2023

Dekan:	Prof. Dr. Jürgen von Hagen
Erstreferent:	Prof. Dr. Benny Moldovanu
Zweitreferent:	Prof. Dr. Daniel Krähmer
Tag der mündlichen Prüfung:	27. Juni 2023

Acknowledgements

I would like to express my sincere gratitude to the many people who helped me in the process of completing this dissertation.

First, I would like to thank my advisors, Benny Moldovanu and Daniel Krähmer, for their countless insightful comments on my work, from feedback on early ideas to writing advice on papers to be submitted. They have given me the freedom to work independently on my own topics, at my own pace, while providing me with the necessary guidance that steered my research in a fruitful direction. Their input was crucial to the success of my work, and I consider myself fortunate to have been able to count on their support throughout my doctoral studies.

I would also like to thank my fantastic co-authors Tangren Feng, Qinggong Wu, Deniz Kattwinkel, Alexander Winter, and especially Justus Preusser. Not only did they contribute many great ideas, hard work, and tremendous expertise to our joint projects, which now form a significant part of this dissertation, but they also taught me through many lively discussions what it means to do research in economic theory, from asking the right questions to effectively communicating results. I am proud of what we have accomplished.

I am grateful to the excellent theory group in Bonn and to all my fellow PhD students for the many stimulating discussions and productive coffee breaks that helped me throughout my PhD journey. I have benefited from the infrastructure and financial support of my graduate institution, the BGSE, and the ECONtribute Cluster of Excellence, which have enabled me to pursue my doctoral studies. I would also like to thank Mirjam Beyer and Farzad Saidi for helping me organize and navigate the job market.

Finally, I am grateful to my family and friends for their unwavering support in whatever I choose to do. They encourage and motivate me when I need it most, and they are always there to lend a listening ear or offer helpful advice. Throughout my PhD journey, they have been a source of joy and laughter and helped me keep my balance and sanity, especially in times of Covid.

Contents

List of Figures

Introduction

Mechanism design is a subfield of economic theory that is concerned with the design of rules—mechanisms—for the interactions of self-interested agents. The goal is to achieve optimal outcomes, such as maximizing social welfare, in situations where agents have private information (e.g. their preferences) and act strategically. The theory of mechanism design has been successfully applied in numerous domains, including resource allocation and public good provision, auction and contract design, or the organization of matching markets and elections.

In many classical applications, it is well-understood how monetary transfers between agents can help align individual interests towards a targeted outcome. However, the theory is less developed when such transfers are unavailable. What are the outcomes that can be achieved by mechanisms without transfers? This thesis gives answers in two important domains: collective decision problems (Chapter 1 and Chapter 2) and simple allocation problems (Chapter 3 and Chapter 4). In the remainder of this introduction, these domains are discussed in greater detail, along with an overview of the main findings of each chapter.

Collective decision problems involve a group of agents having to choose one of several social alternatives. Consider, for instance, a parliament deliberating on a bill, a hiring committee evaluating a job candidate, or a regulatory board deciding whether to approve a new drug. In such situations, decision-relevant information is often dispersed among agents (preference interdependence), and agents sometimes disagree on which decision to make even if all private information were publicly known (preference heterogeneity). A thorough understanding of these collective decision problems can help explain the use of certain voting or deliberation procedures in practice and reveal their potential for improvement through mechanism design.

In Chapter 1, based on Feng, Niemeyer, and Wu (2023), we show that deterministic ex post implementation without transfers is impossible if the underlying environment is neither almost an environment with private values nor almost one with common values. This finding suggests that the equilibrium outcomes of collective decision problems with preference interdependence and heterogeneity are likely sensitive to what agents know and believe about each others' information. It also suggests the use of weaker solution concepts for these problems, such as posterior implementation, which requires that each agent's strategy is optimal against the strategies of other agents for every possible profile of equilibrium messages (Green and Laffont, 1987).

In Chapter 2, based on Niemeyer (2022), I characterize posterior implementable social choice functions in a setting with two social alternatives in terms of score voting mechanisms. In such a mechanism, each agent submits a number from a set of consecutive integers, and the collective decision is determined by whether or not the sum exceeds a given quota. This characterization generalizes an earlier, geometric characterization by Green and Laffont (1987) for the two-agent case. It also yields the key insight that posterior implementation is not significantly more demanding than Bayesian implementation in the two-agent setting of Green and Laffont but becomes very stringent as one moves beyond this special case. The practical relevance of posterior implementation stems from it being the exact solution concept that ensures the robustness of equilibrium outcomes against the extensive form of the mechanism. I discuss applications to sequential voting games (Dekel and Piccione, 2000) and jury decision-making (Li, Rosen, and Suen, 2001).

The simple allocation problem is to distribute a single desirable good among a group of agents, where each agent has private information about the social benefit or value to the principal that would result if they were to be allocated the good. Prominent examples include the allocation of tasks, resources, or money within organizations such as firms, governments, or clubs. With few exceptions, the literature on the problem maintains the classical mechanism design assumption of independent types. However, types are correlated in many real-world applications in that each agent's private information is informative about the value of allocating to others. We investigate how such correlation can be leveraged to design effective allocation mechanisms without transfers, even if no further instruments such as verification, future allocations, or ex-post punishments are available.

In Chapter 3, based on Niemeyer and Preusser (2022), we investigate dominant-strategy incentive compatible (DIC) mechanisms. First, we show that theoretically optimal DIC mechanisms require a degree of complexity that is undesirable in practice. On the one hand, we give a full characterization of when the set of DIC mechanisms is fully described by deterministic mechanisms—such a description holds only in special cases. On the other hand, we show that it is impossible for a DIC mechanism to process the agents' reports anonymously. Second, we make a positive and normative case for a simple class of mechanisms that is actually observed in practice—we call them jury mechanisms. In a jury mechanism, the set of agents is split into jurors and candidates; the allocation only depends on the reports of the jurors, but jurors themselves never win the object. We show that jury mechanisms are optimal when there are three agents, are approximately optimal in symmetric environments with many agents, and are the only deterministic DIC mechanisms that satisfy a reasonably relaxed notion of anonymity.

In Chapter 4, based on Kattwinkel et al. (2022), we investigate Bayesian incentive compatible (BIC) mechanisms. Our analysis begins with the case of two agents, where the simple allocation problem has a natural interpretation: a decision-maker has two options; which option she prefers depends on the private information of two agents who are fully biased: one agent always prefers the first option, the other always prefers the second. First, we fully characterize the set of all BIC mechanisms. Second, we show that in stark contrast to settings with transfers, the set of BIC mechanisms shrinks as the correlation structure becomes richer. Third, we show that the principal can actually benefit from consulting fully biased agents. The key insight is that profitable mechanisms can be constructed when the principal has interdependent payoffs; that is, whenever the information of one agent is relevant to the principal's return of allocating to the other agent. This insight generalizes to the n-agent case.

References

Dekel, Eddie, and Michele Piccione. 2000. "Sequential voting procedures in symmetric binary elections." *Journal of Political Economy* 108 (1): 34–55. [2]

Feng, Tangren, Axel Niemeyer, and Qinggong Wu. 2023. "The limits of ex post implementation without transfers." *Theoretical Economics* 18 (2): 463–79. [1]

Green, Jerry R, and Jean-Jacques Laffont. 1987. "Posterior implementability in a two-person decision problem." *Econometrica: Journal of the Econometric Society,* 69–94. [2]

Kattwinkel, Deniz, Axel Niemeyer, Justus Preusser, and Alexander Winter. 2022. "Mechanisms without transfers for fully biased agents." https://arxiv.org/abs/2205.10910. [3]

Li, Hao, Sherwin Rosen, and Wing Suen. 2001. "Conflicts and common interests in committees." *American Economic Review* 91 (5): 1478–97. [2]

Niemeyer, Axel. 2022. "Posterior Implementability in an N-Person Decision Problem." Available at SSRN 4270283. [2]

Niemeyer, Axel, and Justus Preusser. 2022. "Simple Allocation with Correlated Types." Working Paper. [2]

Chapter 1

The Limits of Ex Post Implementation without Transfers*

1.1 Introduction

Collective decision-making takes place everywhere, from a committee choosing which job candidates to hire, a congress deciding whether to pass a bill, to a country electing its next president. When designing a decision mechanism for such situations, an important consideration is informational robustness: The mechanism should function effectively for a wide range of information structures, i.e., what agents know and believe about each other's information. Robustness is important because decision mechanisms are often institutionalized for repeated use, each time tackling a new problem with a different information structure. Thus, robust, all-purpose mechanisms are best suited for institutions such as committees, legislatures or elections. Moreover, even in a single decision problem, there is usually uncertainty about the underlying information structure. Thus, narrowly tailored mechanisms may misfire if the actual information structure turns out to be different from what was expected.

One might then ask: Are robust decision mechanisms viable? If monetary transfers are allowed, then the answer can be positive — even if one requires robustness against *all* possible information structures, which, by Bergemann and Morris (2005), amounts to the mechanism in question admitting an *ex post equilibrium*. More specifically, it is known that in interdependent value environments, non-trivial, even efficient, social choice functions can be *ex post incentive compatible (EPIC)*, i.e., implementable in an ex post equilibrium of some mechanism, if private information is one-dimensional.[1] There are limits to ex post implementation with transfers, though, as Jehiel et al. (2006) show: If private information is continuous and multi-

* This chapter is based on Feng, Niemeyer, and Wu (2023).

1. For public goods provision, see Section 5 in Chung and Ely (2003). In auction settings, efficient social choice functions are ex post implementable when preferences satisfy appropriate single-

dimensional, then deterministic EPIC social choice functions must be constant in generic environments.

In many collective decision problems, including the examples mentioned above, monetary transfers cannot be used. One would expect ex post implementation to be further constrained by the absence of transfers, but to what extent? — This is the central question we address in this paper. Our main result is as follows: for collective decision problems with a continuous state space, if transfers are not allowed, then deterministic EPIC social choice functions must be constant as long as there is a "small amount" of preference interdependence and preference heterogeneity in the environment, regardless of whether types are one- or multidimensional. If there are only two alternatives, then the conclusion even extends to stochastic social choice functions. Thus, we sharpen the findings of Jehiel et al. (2006) for settings without transfers — we will compare the two papers in more detail after taking a closer look at our result first.

Let us elaborate on the setting. A group of n agents must collectively choose one of finitely many alternatives. Each agent i's private information — her type — is a number or vector θ_i, whereas the collection of everyone's types, $\theta = (\theta_1, ..., \theta_n)$, constitutes the payoff-relevant *state*. An agent's preferences over the alternatives depend on the state, which includes others' as well as her own information.

The sufficient condition for our impossibility result can be more precisely stated as follows: if in state θ some agents are indifferent between two alternatives (a, b), then among the indifferent agents there exists a certain agent i whose indifference between (a, b) is broken by a slight change in the information of another agent j, and moreover, the preferences of i and j regarding (a, b) do not agree entirely in any arbitrary neighborhood around θ. Thus, locally around θ, there is preference interdependence because the preference of i depends non-trivially on j's information, and there is preference heterogeneity because the preferences of i and j differ.

There are three reasons why we suggest that this sufficient condition requires only a "small amount" of preference interdependence and heterogeneity. First, the condition only imposes restrictions on those "indifference" states where agents are actually indifferent between alternatives. Second, the "magnitude" of preference interdependence and heterogeneity, locally at a state, need not be large. Indeed, the condition is satisfied at θ even if i's preference is barely sensitive to j's information, and their preferences are almost but not entirely identical. Third, for an indifference state and a corresponding pair of alternatives, we merely need two agents, i and j, whose preferences are interdependent and heterogeneous. In other words, two agents are enough to disrupt ex post implementation.

The range of environments where our impossibility theorem applies is not only broad in theory, but also relevant in practice: decision-relevant information is of-

crossing conditions; see Crémer and McLean (1985), Maskin (1992), Dasgupta and Maskin (2000), Jehiel and Moldovanu (2001), Bergemann and Välimäki (2002), and Perry and Reny (2002).

ten dispersed across individuals with diverse intentions and tastes, which formally translates into interdependence and heterogeneity of preferences. In terms of how mechanisms such as voting or deliberation procedures operate in the real world, our result therefore predicts that equilibrium outcomes are likely sensitive to what agents believe about each others' information.

Although, as we have argued above, the sufficient condition for our impossibility result is satisfied in a broad range of environments, there are two prominent types of environments in which it is violated: environments with *private values*, where agents' preferences never depend on the information of others, and environments with *common values*, where agents share the same preferences in every state. It is therefore not surprising that these environments admit non-constant EPIC social choice functions. In the case of private values, EPIC is known to be equivalent to strategy-proofness. There, dictatorships are strategy-proof, and further non-constant social choice functions become strategy-proof when the famous Gibbard-Satterthwaite Theorem (Gibbard, 1973; Satterthwaite, 1975) is circumvented through restrictions on the underlying preference domain.[2] In the case of common values, the social choice function that chooses the common first-best alternative in each state is clearly EPIC. Yet, as we have seen, the possibility of ex post implementation quickly fades as we move away from these two extremes, when both preference interdependence and heterogeneity come into play. In particular, not even dictatorship is EPIC when values are interdependent,[3] and various exceptions to Gibbard-Satterthwaite are killed by even a small amount of preference interdependence.

We already mentioned that our result strengthens the finding of Jehiel et al. (2006) for settings without transfers: while Jehiel et al. (2006) show that deterministic ex post implementation with multi-dimensional types is generically impossible, even when transfers are available, we show that shutting down transfers further limits the scope of ex post implementation, especially in environments with one-dimensional types and in important "non-generic" environments that survive Jehiel et al. (2006), such as those with additively separable preferences.[4]

Both results arise from roughly the same conceptual barrier to ex post implementation, namely that agents who can change the social choice around a given state must have aligned preferences. However, compared to Jehiel et al. (2006), the absence of transfers allows us to expose this barrier more explicitly and translate it into an easily interpretable condition on the underlying preferences. The nature of our result as an impossibility result is then established by the argument that this condition

2. See, for example, Moulin (1980) and Saporiti (2009).

3. The reason is that a dictator who decides based on her own information would revise her choice in some states after learning about other agents' information. Also see Jehiel et al. (2006) for disambiguation of the term *dictatorship* in interdependent value environments.

4. See Section 5.4.2 in Jehiel et al. (2006) for the formal definition.

is satisfied in many economically relevant and practically prevalent environments, rather than via a mathematical genericity argument as in Jehiel et al. (2006).

Another difference between the two papers is that Jehiel et al. (2006) only consider a two-agent, two-alternative model. This simple model is sufficient for their goal of establishing generic impossibility, and in principle, the two-by-two setting is also enough to illustrate the key insights of our paper (see Section 1.2). However, our general analysis with many agents and alternatives covers economically relevant (but in the sense of Jehiel et al. (2006), non-generic) cases where it is a priori unclear whether some form of ex post implementation becomes possible, e.g., when subsets of agents have aligned preferences over subsets of alternatives.

There are only a few other papers on ex post implementation without transfers. Che, Kim, and Kojima (2015) and Fujinaka and Miyakawa (2020) as well as Pourpouneh, Ramezanian, and Sen (2020) study specific settings, namely object assignment and matching problems, respectively. In these settings, non-trivial ex post implementation is typically possible: our preference interdependence condition entails allocative externalities, which are typically assumed away in the assignment and matching literature; see Section 1.5 for a more detailed discussion. The impossibility of ex post implementation can be overcome in the same way when transfers are available: genericity in the sense of Jehiel et al. (2006) also entails allocative externalities. In fact, Bikhchandani (2006) shows by construction that non-trivial ex post implementation is possible in environments with private objects and multidimensional types.

For more general settings, Barberà, Berga, and Moreno (2018, 2019) and Feng and Wu (2020, Section 4.3) also discuss necessary and sufficient conditions for the impossibility of ex post implementation. Unlike us, these papers impose no topological structure on the state space, making their conditions more general yet also more abstract and harder to interpret and verify than our conditions. Indeed, it is precisely because we are working with a continuous state space that we are able to obtain a much sharper result about ex post implementability.

This paper is organized as follows. Section 1.2 illustrates the main insight in a simple example. Section 1.3 sets up the general model. Section 1.4 presents the main result. Section 1.5 discusses ex post implementation in situations where our result is silent: (1) allowing transfers; (2) matching and assignment problems; (3) discrete state spaces; (4) stochastic social choice with three or more alternatives. All proofs are in Section 1.A.

1.2 Example

Two agents, 1 and 2, need to make a collective choice from two alternatives, S(afe) and R(isky), e.g., whether or not to pass a law, implement a project, or convict a defendant. The value of S is always 0 to both agents, whereas the value of R depends on an unknown state $\theta = (\theta_1, \theta_2)$, which can take any value from $\Theta = [-1,1]^2$. Specifically, the value of R to agent $i = 1, 2$ is given by

$$v_i^R(\theta) \;=\; \theta_i + \beta\theta_{-i}$$

where $\beta \in [0,1]$.

Agent i can observe θ_i but not θ_{-i}. Thus, each agent only has partial information about the true payoff-relevant state, and β is a parameter that captures the degree to which agent i's valuation depends on the information of the other agent $-i$. Note that when $\beta = 0$, this is a *private value* environment where an agent's preference depends only on her own information, whereas when $\beta = 1$, this is a *common value* environment where the agents preferences are identical. We will return to these special cases in a moment.

We first focus on an intermediate case $\beta = 1/2$. Since each agent's valuation for R is twice as sensitive to her own information as to the other agent's information, the two agents do not always agree on which alternative is better. Indeed, in Figure 1.1a, which graphically represents the setting, the two agents' *indifference curves* $IC_i = \{\theta \mid v_i^R(\theta) = 0\}$, i.e., the respective sets of states where 1 and 2 are indifferent between S and R, partition the state space into four regions, $\{RR, RS, SR, SS\}$, where region XY has the interpretation that within it, agent 1 strictly prefers alternative X and agent 2 strictly prefers alternative Y.

Which deterministic social choice functions $\phi : [-1,1]^2 \to \{S, R\}$ are EPIC when $\beta = 1/2$? ϕ is EPIC if and only if it is optimal for each agent i to truthfully report her type θ_i to the direct mechanism induced by ϕ *in every state*, given that the other agent also reports truthfully. Obviously, any constant ϕ is EPIC. It turns out that the converse is also true: Any EPIC ϕ must be constant.

Let us briefly sketch the gist of the formal argument. Note that if an agent has the same preference across two states that differ only in her own information, then an EPIC social choice function must choose the same alternative in both states. As an example, consider the two states θ and θ' in Figure 1.1a. These states are aligned vertically (thus differ only in agent 2's information) and are respectively located in RR and SR (thus agent 2 strictly prefers R in both states). If some ϕ chose different alternatives in θ and θ', then agent 2 would be decisive in either state: she could induce the choice of one alternative by reporting her information truthfully, or the choice of the other alternative by misreporting her information to be dimension 2 of the other state. But since she strictly prefers R in both states, she would induce the choice of R in one of the states by misreporting her private information, contradicting EPIC.

Now, any EPIC ϕ must be constant within each of the four regions where both agents' preferences are strict and constant because we could otherwise find two states in the same region that differ only in one agent's information but where different alternatives are chosen, contradicting our previous observation about EPIC.

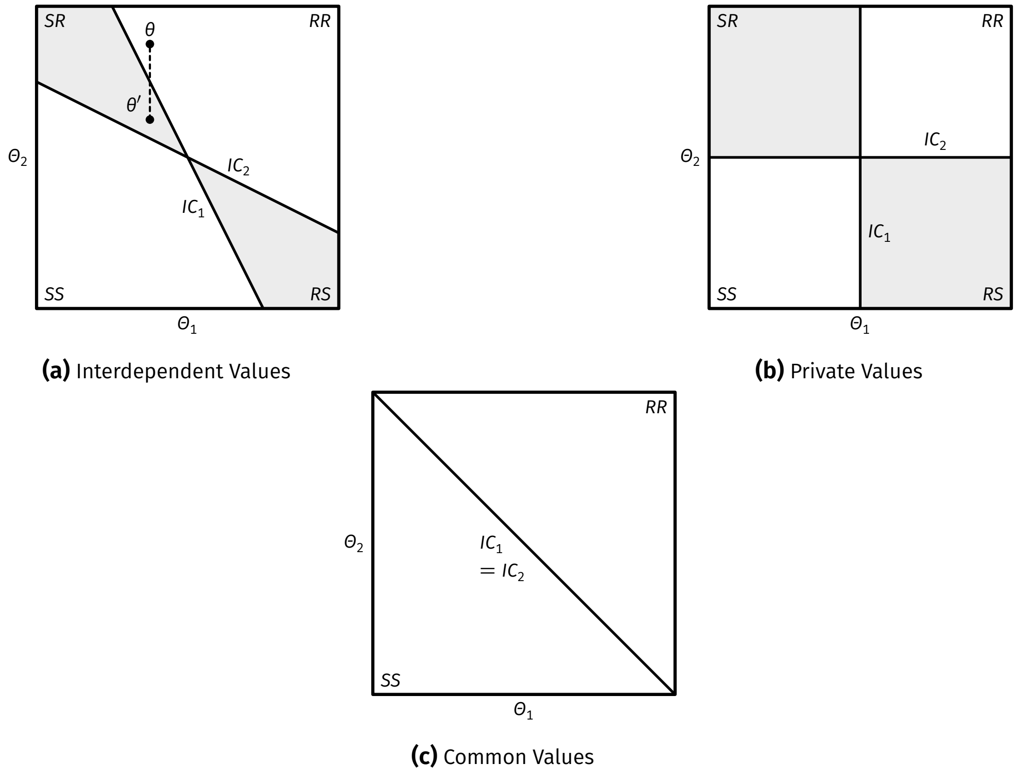


Figure 1.1. An illustration of the example.

In addition, ϕ must choose the same alternative across any two adjacent regions because we can always find states such as θ and θ' that link two regions through an agent whose preference is the same. It follows that any EPIC ϕ must choose the same alternative across all four regions.[5]

It is worth noting that the linking argument across regions relies on the existence of the two states (θ, θ') that (1) differ only in one dimension $j \in \{1, 2\}$, and in which (2) agent $i \neq j$ has different ordinal preferences but (3) agent j has the same ordinal preference. Conditions (1) and (2) jointly entail *preference interdependence* between the agents: the change of agent j's information leads to a change in agent i's ordinal preference. Conditions (2) and (3) jointly entail *preference heterogeneity*: the agents' ordinal preferences do not always agree, so that a change in the state may cause a change in one agent's preference but not in the other's. In short, that ϕ is constant relies on the presence of preference interdependence and heterogeneity.

Not surprisingly, there exist non-constant EPIC ϕ if preference interdependence is absent as in the private value case $\beta = 0$ (Figure 1.1b) or if preference heterogeneity is absent as in the common value case $\beta = 1$ (Figure 1.1c) because we cannot

5. In this example, it is easy to show that ϕ must then also choose the same alternative on the indifference curves IC_1 and IC_2. In general, one can only show this for the interior of the state space; see the Appendix.

find the desired linking states (θ, θ') in either case. For example, the function ϕ that chooses R only in RR is EPIC in both cases.

On the other hand, the argument goes through for any $\beta \in (0,1)$, i.e., when there is at least some preference interdependence and heterogeneity, regardless of how close β is to one of the two exceptional cases. In this sense, if we think of the environments with interdependent values as a spectrum parametrized by $\beta \in [0,1]$ with private values at one end and common values at the other, then even a slight departure from the two extremes leads to an impossibility of ex post implementation. This insight, as formalized and generalized in Theorem 1.1, is the main contribution of the paper.

1.3 Model

A group of agents $N = \{1, \ldots, n\}$ must collectively choose an alternative from a finite set A without using monetary transfers. The valuation of agent $i \in N$ for alternative $a \in A$ depends on an underlying *state* $\theta \in \Theta$, where Θ is the set of all possible states. We represent i's valuation for alternative a by a *valuation function* $v_i^a : \Theta \to \mathbb{R}$. In addition, we let $v_i^{ab}(\theta) := v_i^a(\theta) - v_i^b(\theta)$ denote i's *relative valuation function* for a versus another alternative b. Thus, i weakly prefers a over b in state θ if and only if $v_i^{ab}(\theta)$ is non-negative.

Preference interdependence among the agents is typically modeled by assuming that each agent is only partially informed about the payoff-relevant state θ. Specifically, θ consists of n components, $\theta = (\theta_1, ..., \theta_n)$, and each agent i only observes θ_i — her *type*. The state space Θ is therefore $\prod_{i \in N} \Theta_i$. We assume $\Theta_i = [-1,1]^{d_i}$ where $d_i \in \mathbb{N}$ is the dimension of agent i's type and thus allow for multidimensional types.[6]

Valuation functions are continuously differentiable. Given a relative valuation function v_i^{ab}, let ∇v_i^{ab} denote its gradient, and let $\nabla_{\theta_j} v_i^{ab}$ denote the d_j—dimensional vector of components of ∇v_i^{ab} with respect to the type of agent j. We follow Jehiel et al. (2006) in assuming that an agent's indifference between two alternatives is broken by a slight change in her own information. More precisely,

$$\forall i \in N, \forall \theta \in \Theta, \forall a, b \in A : a \neq b, \quad \left(v_i^{ab}(\theta) = 0 \implies \nabla_{\theta_i} v_i^{ab}(\theta) \neq \mathbf{0} \right).^{7} \tag{RESP}$$

As motivated in the introduction, we are interested in the ex post implementability of social choice functions. By the Revelation Principle, we can focus on those that

6. Our result still obtains if each Θ_i is a subset of a Euclidean space with connected interior. Moreover, Θ need not be a product state space, provided its interior is connected. Our proof explicitly assumes only these properties of the state space.

7. This assumption is not necessary for the gist of our result but simplifies statement and proof: without (RESP), the result's conclusion must be slightly weakened, making the result harder to communicate. See Feng and Wu (2020) for an earlier version of the result without (RESP).

are truthfully ex post implementable in direct mechanisms, or in other words, ex post incentive compatible. Specifically, a (deterministic) *social choice function* $\phi : \Theta \to A$ is *ex post incentive compatible (EPIC)* if truth-telling is an ex post equilibrium of the direct mechanism induced by ϕ, i.e.

$$\forall i \in N, \forall \theta \in \Theta, \forall \tilde{\theta}_i \in \Theta_i, \quad v_i^{\phi(\theta_i,\theta_{-i})}(\theta) \geq v_i^{\phi(\tilde{\theta}_i,\theta_{-i})}(\theta). \tag{EPIC}$$

Following Jehiel et al. (2006), we say that social choice function ϕ is *trivial* if it is constant on the interior of Θ.

Although we have set up the model in terms of cardinal valuation functions, our findings can be easily transferred to a model where preferences are ordinal. After all, only ordinal preferences matter for ex post incentives when there are no transfers and mechanisms are deterministic. In Section 1.5, we discuss this alternative specification in more detail.

1.4 Impossibility of Ex Post Implementation

Let us first formally present the main result and explain it in more detail right after. For a pair of distinct alternatives (a, b), let

$$I^{ab}(\theta) = \{i \in N \mid v_i^{ab}(\theta) = 0\}$$

denote the set of agents who are indifferent between this pair in state θ. If $I^{ab}(\theta)$ is nonempty, we say that (a, b) is an *indifference pair* of θ. Moreover, we say that θ is an *indifference state* if it has at least one indifference pair.

Theorem 1.1. *Suppose for any indifference state θ and any of its indifference pairs (a, b), there exists an agent $i \in I^{ab}(\theta)$ and another agent $j \in N$ such that:*
(1) (local interdependence) $\nabla_{\theta_j} v_i^{ab}(\theta) \neq \mathbf{0}$*;*
(2) (local heterogeneity) $j \notin I^{ab}(\theta)$ *or* $\nabla v_i^{ab}(\theta) \neq \lambda \nabla v_j^{ab}(\theta)$ *for any* $\lambda \geq 0$.
Then, all EPIC social choice functions are trivial.

To better understand the result, let us parse the statement. Note first that the sufficient condition only constrains indifference states regarding their indifference pairs. That is, only for the indifference states θ and their indifference pairs (a, b) do we need to find two agents i and j whose preferences regarding (a, b) are interdependent but nonetheless heterogeneous locally around θ. More precisely, local interdependence means that i, who is indifferent between (a, b) in θ, is no longer indifferent following some small change in j's type, i.e., the *ordinal preference* of i depends on j's information around θ. Local interdependence is satisfied in Figure 1.1a but not in Figure 1.1b because it requires agent 1's indifference curve IC_1 to not be entirely vertical and agent 2's indifference curve IC_2 to not be entirely horizontal. Local heterogeneity means that i and j disagree on whether a or b is better in or near state θ. Specifically, if $j \notin I^{ab}(\theta)$, then heterogeneity in θ is immediate: i is indifferent, but j is not. On the other hand, if j is also indifferent in θ, then the condition that $\nabla v_i^{ab}(\theta) \neq \lambda \nabla v_j^{ab}(\theta)$ for any $\lambda \geq 0$, i.e., that the two gradients are not co-directional at θ, implies that there is an arbitrarily close state in which i and j rank (a, b) differently.[8] Local heterogeneity is satisfied in Figure 1.1a but not in Figure 1.1c because it requires that IC_1 and IC_2 cross each other when they intersect,[9] as only then would the gradients, which are respectively normal to the indifference curves, be misaligned at the intersection.

We view Theorem 1.1 as a strong negative result — an "impossibility" theorem — for the following reasons. First, its sufficient condition only puts restrictions on indifference states, which typically compose a very small subset of all states.[10] Second, local interdependence only rules out the knife-edge case that $\nabla_{\theta_j} v_i^{ab}(\theta)$ is exactly equal to $\mathbf{0}$, and likewise, in case $j \in I^{ab}(\theta)$, local heterogeneity only rules out the knife-edge case that $\nabla v_i^{ab}(\theta)$ and $\nabla v_j^{ab}(\theta)$ are exactly co-directional. In other words, the sufficient condition is satisfied even if, locally around θ, there is only a minimal amount of preference interdependence and heterogeneity. Third, for there to be local interdependence and heterogeneity, we only need two agents whose preferences jointly satisfy the respective requirements, and these agents need not be the same across indifference states or even pairs. In particular, our result still holds if subsets of agents, say, parties in a parliament, have identical preferences as long as there is preference interdependence and heterogeneity between parties.

In fact, the result can be further strengthened. First, what we prove in the Appendix is actually stronger (Theorem 1.2): non-trivial social choice functions do not exist even under the weaker notion of *local ex post incentive compatibility*, which requires that no agent i has an incentive to slightly misrepresent her true type θ_i as some $\tilde{\theta}_i$ that is close to θ_i. Moreover, the presence of local interdependence and heterogeneity in *every* indifference state is an overkill for deriving the impossibility result. All that is needed is a specific discrete set of indifference states satisfying the conditions; see Remark 1.1 in the Appendix.

Why is the existence of a minimal amount of preference interdependence and heterogeneity in some indifference states already enough to disrupt even local ex post implementation? With transfers absent and mechanisms deterministic, incentives are determined by preference rankings only. Thus, it is local incentives around indifference states that matter most to implementation because indifference states are precisely those where preference rankings change. Moreover, since minimal movements around an indifference state are enough to change an agent's prefer-

8. Another way to think of this condition is that there are two pieces of information in state θ between which i and j have different marginal rates of substitution.

9. For local heterogeneity to hold at the intersection, the indifference curves of i and j may be tangent only when their preferences regarding (a, b) are diametrically opposed in a neighborhood of the intersection, which is not possible in the example for any $\beta \in [0, 1]$.

10. Clearly, for any continuous distribution on Θ, the set of indifference states has measure zero.

ence ranking, EPIC admits no "margin of error" there when it comes to the magnitude of preference interdependence or heterogeneity. The implied discontinuity in implementability between pure private/common value environments and interdependent value environments reflects how chokingly stringent EPIC is as a constraint on mechanism design.

1.5 Discussion

1.5.1 An ordinal framework

As we have mentioned earlier, when transfers are absent and mechanisms are deterministic, ex post incentives are determined by preference rankings only, whereas cardinal valuations per se are irrelevant. Although the key conditions for our analysis — those about local interdependence and local heterogeneity — are formulated in terms of cardinal valuations, they are essentially about how ordinal preferences change from an indifference state to nearby states. In principle, these conditions can be alternatively defined in terms of ordinal preferences, but one can imagine that such definitions would be more tedious to formulate and use for our analysis. Since our result hinges on preferences in and around indifference states, we have imposed mild regularity conditions on valuation functions to ensure that the set of indifference states is well-behaved. In the ordinal model, if we were to impose analogous conditions on the boundaries that separate the regions where a given agent's preferences are constant, then our analysis would go through analogously with the appropriately modified notions of local interdependence and heterogeneity.[11]

1.5.2 Transfers

In the introduction, we mentioned that transfers facilitate ex post implementation. If transfers are allowed and an agent only cares about her own transfer, as is typically assumed, then she is indifferent between any two outcomes where the chosen non-monetary alternative and her own transfer are the same, despite differences in the other agents' transfers. These indifferences persist across states and thus violate both local interdependence and (RESP), rendering our result silent. Transfers can be used to overcome preference interdependence or heterogeneity — the two roadblocks to ex post implementation suggested by our result — by either making values effectively private or by aligning the agents' interests.[12] In the following, we illustrate these two possibilities in the context of our leading example.

11. With continuous preferences, these boundaries are nothing but the agent's indifference curves. Whether or not the agent is actually indifferent in states where her preferences change is not relevant for the result.

12. See Section 5 in Chung and Ely (2003) for further discussion on how transfers can be used to align individual interests in collective choice problems.

Example (continued from Section 1.2). Suppose monetary transfers are now allowed and agents have quasi-linear utilities: $u_i(\theta) = v_i^X(\theta) + t_i(\theta)$, where X is the chosen alternative and t_i is the transfer agent i receives.

First, consider the transfer scheme $(t_i)_{i=1,2}$ where $t_i(\theta) = -\beta\theta_{-i}$ if R is chosen and $t_i = 0$ if S is chosen. Agent i's "post-transfer" utility is then θ_i if R is chosen and 0 if S is chosen. Thus, transfers eliminate preference interdependence and transform the environment into one of private values as in Figure 1.1b. Consequently, mechanisms such as dictatorship or unanimity voting are EPIC.

Second, consider the transfer scheme $(t_i')_{i=1,2}$ where $t_i'(\theta) = (1-\beta)\theta_{-i}$ if R is chosen and $t_i = 0$ if S is chosen. Both agents have the same "post-transfer" utility, namely $\theta_1 + \theta_2$ if R is chosen and 0 if S is chosen. Thus, transfers eliminate preference heterogeneity and transform the environment into one of common values as in Figure 1.1c. Consequently, the mechanism that chooses R if and only if $\theta_1 + \theta_2 > 0$ is EPIC.

1.5.3 Assignment and matching problems

A common assumption in matching is that each agent only cares about her own assigned object or match. Thus, similar to the case of transfers, local interdependence and (RESP) generally fail to hold in such problems.[13] It is therefore not surprising that non-trivial EPIC social choice functions exist even when preferences are interdependent.[14] However, as Che, Kim, and Kojima (2015) show, such EPIC social choice functions cannot be efficient, at least in the housing allocation problem where each agent is assigned exactly one object. Moreover, our negative result can still apply to assignment or matching problems with allocative externalities, e.g., when students not only care about which dorm room they get but also which rooms their friends get.

1.5.4 Discrete state spaces

We have assumed that the state space is a connected subset of a Euclidean space. If instead the state space is discrete, then counterexamples to our result are easy to find. For instance, see Feng and Wu (2020). One way to understand the discrepancy between discrete and continuous state spaces is to think of a discrete state space as a low-resolution discretization of a continuous space. For example, suppose each agent's underlying type can be any number between -1 and 1, yet each agent is only aware of whether her type is above or below 0, making her effective type space

13. The housing allocation problem with two objects and two agents is an exception since the assignment of one object to one agent implies that the remaining object must be assigned to the other agent. See also the illustrative example in Che, Kim, and Kojima (2015).

14. This observation echoes how Jehiel et al. (2006) relies on allocative externalities; also see Bikhchandani (2006).

binary. Since the agents' indifference curves are then being squeezed into a discrete grid, they tend to become more aligned, and this alignment gives leeway to non-trivial ex post implementation.

1.5.5 Stochastic social choice functions

What if we allow for randomization so that the collective choice can be a *lottery* over alternatives? It turns out that Theorem 1.1 still holds as long as there are only two alternatives. The reason is simple: an agent is indifferent between lotteries if and only if she is indifferent between the two underlying alternatives, and she otherwise prefers lotteries in which her preferred alternative is chosen with a higher rather than lower probability. Thus, our arguments immediately extend to stochastic implementation with two alternatives. However, if there are three or more alternatives, then an agent can get the same expected utility from different lotteries despite having strict preferences over the underlying alternatives. In the following example, these indifferences can indeed be used to construct a non-trivial stochastic EPIC social choice function.

Example (continued from Section 1.2)**.** Agents 1 and 2 now decide between three alternatives, R, S, and P. For $i = 1, 2$, still assume $\Theta_i = [-1, 1]$, $v_i^R = \theta_i + \theta_{-i}/2$, and $v_i^S = 0$. Additionally, assume $v_i^P = -1$. Theorem 1.1 applies here, so any deterministic EPIC social choice function must be trivial. However, consider the stochastic social choice function $\phi = (\phi^R, \phi^P, \phi^S)$ given by

$$\begin{aligned} \phi^R(\theta) &= \frac{4 + 2\theta_1 + 2\theta_2}{11}, \\ \phi^P(\theta) &= \frac{\theta_1^2 + \theta_1\theta_2 + \theta_2^2}{11}, \\ \phi^S(\theta) &= 1 - \phi^R(\theta) - \phi^P(\theta), \end{aligned}$$

where $\phi^X(\theta)$ denotes the probability that alternative X will be chosen in state θ. It is readily verified that ϕ is EPIC.

Appendix 1.A Proof of Theorem 1.1

Endow Θ with the norm topology. Let $B_\varepsilon(\theta)$ denote the open ball with radius $\varepsilon > 0$ centered at θ. A social choice function ϕ is said to be *locally EPIC* if there exists some $\varepsilon > 0$ such that for any $\theta \in \Theta$, ϕ restricted to $B_\varepsilon(\theta)$ is EPIC, i.e.

$$\begin{aligned} &\forall i \in N,\ \forall \theta \in \Theta,\ \forall \tilde{\theta}_i \in \Theta_i : \\ &\qquad (\tilde{\theta}_i, \theta_{-i}) \in B_\varepsilon(\theta) \implies v_i^{\phi(\theta_i,\theta_{-i})}(\theta) \geq v_i^{\phi(\tilde{\theta}_i,\theta_{-i})}(\theta). \qquad \text{(LEPIC)} \end{aligned}$$

Let $\overline{\Theta} := \{\theta \in \operatorname{int}\Theta \mid \forall i \in N,\ \forall a,b \in A : a \neq b,\ v_i^{ab}(\theta) \neq 0\}$ denote the set of interior states where all agents have strict preferences, and let $\mathscr{C}$ denote the set of all connected components of $\overline{\Theta}$. $\overline{\Theta}$ is open because valuation functions are continuous. Similarly, each connected component $C \in \mathscr{C}$ is open. Note that the ordinal preferences of all agents are strict and constant on each $C \in \mathscr{C}$.

Lemma 1.1. *If ϕ is locally EPIC, then ϕ is constant on each $C \in \mathscr{C}$.*

Proof. Suppose ϕ satisfies (LEPIC) for some $\varepsilon > 0$. Pick any $C \in \mathscr{C}$. Suppose for the sake of contradiction that ϕ is not constant on C, then there exists some $\theta \in C$ and $\tilde{\varepsilon} \in (0,\varepsilon)$ such that $B_{\tilde{\varepsilon}}(\theta) \subset C$ and $\phi(\theta) \neq \phi(\theta')$ for some $\theta' \in B_{\tilde{\varepsilon}}(\theta)$. Clearly we can find a sequence of states $(\theta^0, \dots, \theta^n)$ in $B_{\tilde{\varepsilon}}(\theta)$ where $\theta^0 = \theta$, $\theta^n = \theta'$, and for every $k = 0, \dots, n-1$, θ^k and θ^{k+1} differ at most in the $k+1$th entry. Thus $\phi(\theta) \neq \phi(\theta')$ implies that $\phi(\theta^k) \neq \phi(\theta^{k+1})$ for some k. By construction, θ^k and θ^{k+1} differ only in the type of agent $k+1$ who has the same strict ordinal preferences in both states. Therefore, she either could profit from misreporting her type as θ^k_{k+1} in state θ^{k+1} or from misreporting her type as θ^{k+1}_{k+1} in state θ^k, contradicting (LEPIC). □

Given Lemma 1.1, it causes no confusion to write $\phi(C)$ for the choice by ϕ on $C \in \mathscr{C}$.

Distinct $C, C' \in \mathscr{C}$ are said to be *adjacent at* $\theta \in \operatorname{int}\Theta$ if (1) $\theta \in [\operatorname{cl} C \cap \operatorname{cl} C']$, and moreover (2) $B_\varepsilon(\theta) \subset [\operatorname{cl} C \cup \operatorname{cl} C']$ for some $\varepsilon > 0$. In addition, we consider every $C \in \mathscr{C}$ as being adjacent to itself (at every $\theta \in \operatorname{cl} C$).

A collection $\boldsymbol{X}$ of vectors are said to be *collinear* if for any $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{X}$, $\boldsymbol{x} = \lambda \boldsymbol{y}$ for some $\lambda \in \mathbb{R}$, i.e., these vectors lie on a common line passing through the origin. If, in addition, for any $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{X}$, $\boldsymbol{x} = \lambda \boldsymbol{y}$ for some $\lambda \geq 0$, i.e., these vectors lie on a common ray emanating from the origin, then they are said to be *co-directional.*

Lemma 1.2. *Suppose ϕ is locally EPIC. If $C, C' \in \mathscr{C}$ are adjacent at $\theta \in \operatorname{int}\Theta$, and $\phi(C) := a \neq b =: \phi(C')$, then*
(1) $\nabla_{\theta_j} v_i^{ab}(\theta) = \mathbf{0}$ for any $i \in I^{ab}(\theta)$ and $j \in N \setminus I^{ab}(\theta)$, and
(2) $\left(\nabla v_i^{ab}(\theta)\right)_{i \in I^{ab}(\theta)}$ are co-directional.

Proof. The lemma's premises imply that we can find $\varepsilon > 0$ such that (1) (LEPIC) holds for $B_\varepsilon(\theta)$, (2) $B_\varepsilon(\theta) \subset [\operatorname{cl} C \cup \operatorname{cl} C']$, and (3) for any agent i and any distinct pair of alternatives (x,y), if i strictly prefers x to y in θ, then she strictly prefers x to y in every state in $B_\varepsilon(\theta)$.

Arbitrarily pick alternatives $x, y, w, z \in A$ where $x \neq y$ and $w \neq z$ and agents $i \in I^{xy}(\theta)$ and $j \in I^{wz}(\theta)$. Claim that $\nabla v_i^{xy}(\theta)$ and $\nabla v_j^{wz}(\theta)$ are collinear. Indeed, if not, then we can find $\theta', \theta'' \in B_\varepsilon(\theta)$ such that $v_i^{xy}(\theta') = 0$ but $v_j^{wz}(\theta') \neq 0$, and $v_i^{xy}(\theta'') \neq 0$ but $v_j^{wz}(\theta'') = 0$.

By (RESP), we can find two states arbitrarily close to θ' (hence within $B_\varepsilon(\theta)$) in which j has the same strict preference regarding (w,z) but i has different strict preferences regarding (x,y). Similarly, we can find two states arbitrarily close to θ''

(hence also within $B_\varepsilon(\theta)$) in which i has the same strict preference regarding (x,y) but j has different strict preferences regarding (w,z). Thus $B_\varepsilon(\theta)$ must intersect at least three distinct connected components of $\overline{\Theta}$ as it contains at least three profiles of strict preferences of the agents. This contradicts that $B_\varepsilon(\theta)$ only intersects two such components, namely C and C'.

Towards proving part (1), suppose for the sake of contradiction that there exists $i \in I^{ab}(\theta)$ and $j \in N \setminus I^{ab}(\theta)$ such that $\nabla_{\theta_j} v_i^{ab} \neq \mathbf{0}$. Thus, we can find $\rho > 0$ sufficiently small such that $\theta' := (\theta_j + \rho \nabla_{\theta_j} v_i^{ab}(\theta),\ \theta_{-j}) \in B_\varepsilon(\theta)$, $\theta'' := (\theta_j - \rho \nabla_{\theta_j} v_i^{ab}(\theta),\ \theta_{-j}) \in B_\varepsilon(\theta)$, and

$$v_i^{ab}(\theta')v_i^{ab}(\theta'') < 0, \quad v_j^{ab}(\theta')v_j^{ab}(\theta'') > 0.$$

In other words, agent i has different strict preferences regarding (a,b) in θ' and θ'', whereas agent j has the same strict preference. By the collinearity observation above, we can further conclude that, for ρ small enough, any agent k who is indifferent between any pair (x,y) in θ has different strict preferences regarding this pair in θ' and θ''. Together with $\theta', \theta'' \in B_\varepsilon(\theta)$ we thus establish $\theta', \theta'' \in \overline{\Theta}$, i.e., all agents have strict preferences in both states. Moreover, θ' and θ'' must be in distinct connected components of $\overline{\Theta}$ — one in C, the other in C' — because i's preferences differ across the two states. Since agent j has the same strict preference regarding (a,b) in θ' and θ'' and since the two states differ only in j's type, (LEPIC) implies $\phi(\theta') = \phi(\theta'')$, a contradiction.

Now we show part (2). From the collinearity observation we conclude that $\nabla v_i^{ab}(\theta)$ and $\nabla v_j^{ab}(\theta)$ are collinear for any $i,j \in I^{ab}(\theta)$. If, for the sake of contradiction, for some $i,j \in I^{ab}(\theta)$ the two gradients are not also co-directional, then they must be diametrically opposed. By (RESP), we can find $\rho > 0$ sufficiently close to 0 such that the following three statements are true. First, i strictly prefers a to b and j strictly prefers b to a in both of the following two states:

$$\hat{\theta} := (\theta_i + \rho \nabla_{\theta_i} v_i^{ab}(\theta),\ \theta_{-i}) \quad \text{and} \quad \tilde{\theta} := (\theta_j - \rho \nabla_{\theta_j} v_j^{ab}(\theta),\ \theta_{-j}).$$

Second, i strictly prefers b to a and j strictly prefers a to b in both of the following two states:

$$\hat{\theta}' := (\theta_i - \rho \nabla_{\theta_i} v_i^{ab}(\theta),\ \theta_{-i}) \quad \text{and} \quad \tilde{\theta}' := (\theta_j + \rho \nabla_{\theta_j} v_j^{ab}(\theta),\ \theta_{-j}).$$

Third, the above two pairs of states are in $B_\varepsilon(\theta)$. Following the argument in the previous paragraph, the four states are also in $\overline{\Theta}$ and thus either in C or in C'. In addition, one pair must fall in C and the other pair must fall in C' because the preferences of agent i (equivalently, j) regarding (a,b) are the same within each pair but differ across pairs. Therefore, $\phi(\hat{\theta}) = \phi(\tilde{\theta})$ but $\phi(\hat{\theta}) \neq \phi(\hat{\theta}')$. (LEPIC) implies that $\phi(\hat{\theta}) = a$ for otherwise i would misreport her type as $\hat{\theta}_i'$ in state $\hat{\theta}$. Similarly, $\phi(\tilde{\theta}) = b$ for j not to misreport, but then $\phi(\hat{\theta}) \neq \phi(\tilde{\theta})$, a contradiction. □

Lemma 1.3. *For any $C, C' \in \mathscr{C}$ there exists a finite sequence of connected components $C^0, \dots, C^K \in \mathscr{C}$ and a finite sequence of indifference states $\theta^1, \dots, \theta^K \in \operatorname{int}\Theta$ such that $C^0 = C, C^K = C'$, and C^k and C^{k+1} are adjacent at θ^{k+1} for every $k = 0, \dots, K-1$.*

Proof. Pick any $\overline{C} \in \mathscr{C}$. Let $\mathscr{C}'$ denote the set of all $C' \in \mathscr{C}$ that can be linked to $\overline{C}$ through a finite sequence of connected components with the same properties as in the statement of the lemma. Clearly the lemma is established if $\mathscr{C} \setminus \mathscr{C}'$ is empty, thus, for the sake of contradiction, suppose $\mathscr{C} \setminus \mathscr{C}'$ is non-empty.

Define

$$S := \operatorname{int}\Theta \cap \left[\operatorname{cl} \bigcup_{\tilde{C} \in \mathscr{C} \setminus \mathscr{C}'} \tilde{C}\right] \cap \left[\operatorname{cl} \bigcup_{\tilde{C} \in \mathscr{C}'} \tilde{C}\right].$$

Geometrically speaking, S is the frontier separating the components in $\mathscr{C}'$ from those in $\mathscr{C} \setminus \mathscr{C}'$. Note that $S \subset \left[\operatorname{int}\Theta \setminus \overline{\Theta}\right]$. Moreover, S is non-empty, for otherwise, the two (relatively) closed sets $\operatorname{int}\Theta \cap \left[\operatorname{cl}\bigcup_{\tilde{C} \in \mathscr{C} \setminus \mathscr{C}'} \tilde{C}\right]$ and $\operatorname{int}\Theta \cap \left[\operatorname{cl}\bigcup_{\tilde{C} \in \mathscr{C}'} \tilde{C}\right]$ would partition $\operatorname{int}\Theta$, which contradicts that $\operatorname{int}\Theta$ is connected.

For any agent $i \in N$ and distinct alternatives (a, b), let $IC_i^{ab} := \{\tilde{\theta} \in \Theta \mid v_i^{ab}(\tilde{\theta}) = 0\}$ denote the set of states where $i \in N$ is indifferent between (a, b). As an intermediate step, we will show that there exists a state $\theta \in S$ such that if an open ball B centered at θ is sufficiently small, then for any agent $i \in N$ and pair of distinct alternatives (a, b), $[B \cap S] \subset [B \cap IC_i^{ab}]$ if $B \cap IC_i^{ab}$ is non-empty.

The desired state θ can be obtained constructively as follows. Fix an arbitrary state $\theta' \in S$. Since valuation functions are continuous, if an open ball B' centered at θ' is sufficiently small, then $\theta' \in IC_i^{ab}$ for any $i \in N$ and alternatives (a, b) such that $B' \cap IC_i^{ab}$ is non-empty. Now we look for a state $\theta'' \in B' \cap S$ such that for some $i \in N$ and alternatives (a, b), $\theta' \in IC_i^{ab}$ whereas $\theta'' \notin IC_i^{ab}$. If such θ'' does not exist, then θ' is the desired state θ. If such θ'' exists, then we proceed analogously with θ'' in place of θ'. The procedure terminates after finitely many iterations because there are only finitely many agents and pairs of distinct alternatives, thus eventually yielding the desired state θ.

Observe that there is a sufficiently small open ball B centered at θ such that each $B \cap IC_i^{ab}$, if non-empty, not only satisfies $[B \cap S] \subset [B \cap IC_i^{ab}]$ (established above) but also is diffeomorphic to a hyperplane (by (RESP) and the inverse function theorem). Hence, $B \setminus IC_i^{ab}$ consists of two open connected components, $U = \{\tilde{\theta} \in B \mid v_i^{ab}(\tilde{\theta}) < 0\}$ and $U' = \{\tilde{\theta} \in B \mid v_i^{ab}(\tilde{\theta}) > 0\}$, with common boundary $B \cap IC_i^{ab}$.[15]

Now we show that $[B \cap S] = [B \cap IC_i^{ab}]$ if $B \cap IC_i^{ab}$ is non-empty. Since $\theta \in S$, both $\operatorname{int}\Theta \cap \left[\operatorname{cl}\bigcup_{\tilde{C} \in \mathscr{C} \setminus \mathscr{C}'} \tilde{C}\right]$ and $\operatorname{int}\Theta \cap \left[\operatorname{cl}\bigcup_{\tilde{C} \in \mathscr{C}'} \tilde{C}\right]$ must intersect $B \setminus S$, which implies

15. Specifically, using (RESP), suppose without loss of generality that $\partial v_i^{ab}(\theta)/\partial \theta_{is} \neq 0$, where θ_{is} is the s-th entry of θ_i. Then the Jacobian of $h(\theta) = (\theta_{-is}, v_i^{ab}(\theta))$ is invertible, hence h is the desired local diffeomorphism: h^{-1} maps the hyperplane defined by the equation $\theta_{is} = 0$ to IC_i^{ab} and maps the halfspaces separated by that hyperplane to U and U', respectively. Finally, recall that connectivity is preserved under the continuous map h^{-1}.

that $B \setminus S$ is disconnected because the two sets are relatively closed and disjoint in $B \setminus S$. If there exists $\theta' \in B \cap IC_i^{ab}$ such that $\theta' \notin B \cap S$ for some $i \in N$ and alternatives (a,b), then S would be diffeomorphic to a hyperplane missing some points, hence $B \setminus S$ would have to be connected, a contradiction.

It follows that $B \cap \overline{\Theta}$ intersects exactly two connected components in $\mathscr{C}$ because all non-empty $B \cap IC_i^{ab}$ coincide by the previous paragraph. One of these connected components is some $C \in \mathscr{C} \setminus \mathscr{C}'$ and the other is some $C' \in \mathscr{C}'$ because $\theta \in S$ by construction. Moreover, $\theta \in [\mathrm{cl}\, C] \cap [\mathrm{cl}\, C']$. Thus, $C \in \mathscr{C} \setminus \mathscr{C}'$ and $C' \in \mathscr{C}'$ are adjacent at θ, contradicting the initial assumption that no component in $\mathscr{C}'$ is adjacent to a component in $\mathscr{C} \setminus \mathscr{C}'$. □

We will now state and prove a stronger impossibility theorem that immediately implies Theorem 1.1 as a corollary.

Theorem 1.2. *Suppose the premises of Theorem 1.1 hold. Then, all* ***locally*** *EPIC social choice functions are trivial.*

Proof. Fix any ϕ that is locally EPIC for radius $\varepsilon > 0$. Let $\Theta^k \subset \mathrm{int}\,\Theta$ denote the set of interior states where exactly k agents have indifferences in their preferences. Thus $\mathrm{int}\,\Theta = \bigcup_{k=0}^{n} \Theta^k$. It suffices to show that ϕ is constant on Θ^k for every $k = 0, \ldots, n$ and, moreover, that $\phi(\Theta^0) = \ldots = \phi(\Theta^n)$. We proceed by induction on k.

For $k = 0$, note that $\Theta^k = \overline{\Theta}$. Suppose, for the sake of contradiction, that ϕ is not constant on $\overline{\Theta}$. By Lemma 1.3, there exist two connected components C and C' of $\overline{\Theta}$ adjacent at some indifference state θ such that $\phi(C) \neq \phi(C')$. For any indifference pair (a,b) of θ, one of the following two cases must hold by assumption: (1) There is $i \in I^{ab}(\theta)$ and $j \notin I^{ab}(\theta)$ such that $\nabla_{\theta_j} v_i^{ab}(\theta) \neq \mathbf{0}$. (2) There are $i, j \in I^{ab}(\theta)$ where $\nabla v_i^{ab}(\theta)$ and $\nabla v_j^{ab}(\theta)$ are not co-directional. Hence we have $\phi(C) = \phi(C')$ by the contrapositive of Lemma 1.2, a contradiction. Thus ϕ must be constant on $\overline{\Theta} = \Theta^0$.

Now suppose ϕ is constant on Θ^ℓ for every $\ell < k$ and, moreover, $\phi(\Theta^0) = \ldots = \phi(\Theta^{k-1})$. Pick any $\theta \in \Theta^k$. By iteratively using (RESP), we can find states $\theta', \theta'' \in B_\varepsilon(\theta)$ arbitrarily close to θ such that (1) θ, θ' and θ'' differ from each other only in some agent i's type, (2) agent i is indifferent between one or more pairs of distinct alternatives in θ, and, in addition, for any such pair she has strict and opposite preferences in θ' and θ'', (3) for any agent whose preference regarding any given pair of distinct alternatives is strict in θ, her preference regarding this pair remains the same in θ' and θ''. Thus $\theta' \in \Theta^\ell$ and $\theta'' \in \Theta^{\ell'}$ for $\ell, \ell' < k$. Consequently, the inductive hypothesis implies that $\phi(\theta') = \phi(\theta'') = \phi(\Theta^0)$. Suppose for the sake of contradiction that $\phi(\theta) = a$ but $\phi(\Theta^0) = b \neq a$. On the one hand, if i has a strict preference regarding (a,b) in θ, then she has the same strict preference in θ and θ', and hence by (LEPIC), we must have $a = \phi(\theta) = \phi(\theta') = b$ for there to be no incentive for i to misreport, a contradiction. On the other hand, if i is indifferent between (a,b) in θ, then, by construction, i strictly prefers a over b in one of θ' and θ'', and in that state, she has an incentive to misreport her type as θ_i, also a contradiction. Thus $\phi(\theta) = \phi(\Theta^0)$. Since θ was arbitrarily chosen from Θ^k, we conclude that ϕ must be constant on Θ^k and, moreover, $\phi(\Theta^k) = \phi(\Theta^0)$. □

Remark 1.1. The sufficient condition for Theorem 1.1 and Theorem 1.2 can be weakened. Indeed, Lemma 1.3 guarantees the existence of a *discrete* set of indifference states Θ^* such that for any $C, C' \in \mathscr{C}$ there is a finite sequence of connected components $C^0, \ldots, C^K \in \mathscr{C}$ where $C^0 = C$, $C^K = C'$, and C^k and C^{k+1} are adjacent at some $\theta \in \Theta^*$ for every $k = 0, \ldots, K-1$. The proof of Theorem 1.2 goes through as long as local interdependence and heterogeneity are present in such a set of indifference states. Importantly, if $\mathscr{C}$ is finite, then Θ^* can be chosen as a finite set.[16] Thus, the set of indifference states where local interdependence and heterogeneity need actually be present is much smaller than the set of all indifference states.

16. One can imagine that preferences must be rather special for $\mathscr{C}$ to be infinite, but such preferences do exist. We are grateful to an anonymous referee for suggesting the following illustrative example: Suppose there are two agents, two alternatives, and one-dimensional types $\theta = (\theta_1, \theta_2)$ as in our example from Section 1.2. If relative valuations are given by

$$v_1(\theta) = \theta_1^3 \sin(1/\theta_1) - \theta_2 \quad \text{and} \quad v_2(\theta) = \theta_2,$$

where $v_1(0, \theta_2) = -\theta_2$, then any neighborhood of $\theta = (0,0)$ contains infinitely many components in $\mathscr{C}$. The example can be modified to satisfy all of our assumptions, including (RESP), as well as the premises of Theorem 1.1 by rotating the state space.

References

Barberà, Salvador, Dolors Berga, and Bernardo Moreno. 2018. “Restricted environments and incentive compatibility in interdependent values models.” Barcelona GSE Working Paper Series. Working Paper no. 1024. [8]

Barberà, Salvador, Dolors Berga, and Bernardo Moreno. 2019. “Domains Admitting Ex Post Incentive Compatible and Respectful Mechanisms: A Characterization for the Two Alternatives Case.” *Social Design: Essays in Memory of Leonid Hurwicz,* 295–306. [8]

Bergemann, Dirk, and Stephen Morris. 2005. “Robust mechanism design.” *Econometrica* 73 (6): 1771–813. [5]

Bergemann, Dirk, and Juuso Välimäki. 2002. “Information acquisition and efficient mechanism design.” *Econometrica* 70 (3): 1007–33. [6]

Bikhchandani, Sushil. 2006. “Ex post implementation in environments with private goods.” *Theoretical Economics* 1 (3): 369–93. [8, 15]

Che, Yeon-Koo, Jinwoo Kim, and Fuhito Kojima. 2015. “Efficient assignment with interdependent values.” *Journal of Economic Theory* 158: 54–86. [8, 15]

Chung, Kim-Sau, and Jeffrey C Ely. 2003. “Ex-post incentive compatible mechanism design.” Working Paper, Department of Economics, Northwestern University. [5, 14]

Crémer, Jacques, and Richard P. McLean. 1985. “Optimal Selling Strategies under Uncertainty for a Discriminating Monopolist when Demands are Interdependent.” *Econometrica* 53 (2): 345–61. [6]

Dasgupta, Partha, and Eric Maskin. 2000. “Efficient auctions.” *Quarterly Journal of Economics* 115 (2): 341–88. [6]

Feng, Tangren, Axel Niemeyer, and Qinggong Wu. 2023. “The limits of ex post implementation without transfers.” *Theoretical Economics* 18 (2): 463–79. [5]

Feng, Tangren, and Qinggong Wu. 2020. “Robust Binary Voting.” Mimeo. [8, 11, 15]

Fujinaka, Yuji, and Toshiji Miyakawa. 2020. “Ex-post incentive compatible and individually rational assignments in housing markets with interdependent values.” *Journal of Mathematical Economics* 91: 157–64. [8]

Gibbard, Allan. 1973. “Manipulation of Voting Schemes: A General Result.” *Econometrica* 41 (4): 587–601. [7]

Jehiel, Philippe, Moritz Meyer-ter-Vehn, Benny Moldovanu, and William R Zame. 2006. “The limits of ex post implementation.” *Econometrica* 74 (3): 585–610. [5–8, 11, 12, 15]

Jehiel, Philippe, and Benny Moldovanu. 2001. “Efficient design with interdependent valuations.” *Econometrica* 69 (5): 1237–59. [6]

Maskin, Eric. 1992. “Auctions and privatization.” *Privatization, ed. by H. Siebert (Institut für Weltwirtschaften der Universität Kiel),* 115–36. [6]

Moulin, H. 1980. “On Strategy-Proofness and Single Peakedness.” *Public Choice* 35 (4): 437–55. [7]

Perry, Motty, and Philip J Reny. 2002. “An efficient auction.” *Econometrica* 70 (3): 1199–212. [6]

Pourpouneh, Mohsen, Rasoul Ramezanian, and Arunava Sen. 2020. “The Marriage Problem with Interdependent Preferences.” *International Game Theory Review (IGTR)* 22 (02): 1–16. [8]

Saporiti, Alejandro. 2009. “Strategy-proofness and single-crossing.” *Theoretical Economics* 4 (2): 127–63. [7]

Satterthwaite, Mark Allen. 1975. “Strategy-proofness and Arrow’s conditions: Existence and correspondence theorems for voting procedures and social welfare functions.” *Journal of Economic Theory* 10 (2): 187–217. [7]

Chapter 2

Posterior Implementability in an N-Person Decision Problem*

2.1 Introduction

This paper studies a model of collective decision-making where a group of agents must decide whether to accept or reject an alternative without using monetary transfers. The range of applications is broad: imagine a parliament deciding whether to pass a law, a regulatory board deciding whether to approve a new drug, or a group of firms deciding whether to proceed with a merger. In such situations, decision-relevant information is often dispersed among agents (preference interdependence), and agents sometimes disagree on which decision to make even if all private information were publicly known (preference heterogeneity). What are the equilibrium outcomes of such collective decision problems, and by what mechanisms can they be implemented? A thorough understanding of implementability can help explain why certain voting or deliberation procedures are used in practice and whether they can be improved through mechanism design.

The state of knowledge about implementability in collective decision problems with preference interdependence and heterogeneity is very different for the two most commonly used solution concepts: Bayesian implementation and ex post implementation. Bayesian implementation requires mutual optimality of agents' strategies given their interim beliefs about the information of others, whereas ex post implementation requires mutual optimality for every realized information profile. In general, there is no known characterization of Bayesian implementation. By contrast, it is known that non-trivial ex post implementation is impossible (Feng, Niemeyer, and Wu, 2023).[1]

* This chapter is based on Niemeyer (2022).

1. The difficulties associated with ex post implementation are a common theme in the literature, even in settings where monetary transfers are available; see Jehiel et al. (2006). Recall that the

This paper settles the question of implementability for the notion of posterior implementation (Green and Laffont, 1987), which is stronger than Bayesian but weaker than ex post implementation. Posterior implementation requires that each agent's strategy is optimal against the strategies of other agents for every possible message profile. This weakens ex post implementation in that optimality is only required with respect to the information that can be inferred from observed equilibrium behavior in the mechanism—but not necessarily with respect to the revelation of *all* private information.[2] Before delving into the results, it will be illuminating to discuss the practical and theoretical relevance of the concept in more detail.

Posterior implementation is interesting from a practical perspective because it is the exact solution concept that ensures the robustness of agents' equilibrium behavior against the extensive form of the mechanism. Such robustness is important because it is often beyond the designer's control whether a mechanism will be played simultaneously or in some extensive form. For example, voting mechanisms in practice are frequently implemented by show of hands, roll call, or division of the assembly rather than secret ballots. Consequently, a mechanism designed for simultaneous play might misfire if it is played sequentially and if agents adjust their behavior to account for information revelation along the equilibrium path of play.[3] The relevance of this issue is illustrated by the FDA's decision to eliminate sequential voting in favor of a simultaneous electronic voting system for its advisory committees in 2007, citing the risk of momentum in sequential voting as its leading concern (see Urfalino and Costa (2015) and Newham and Midjord (2020) for more details).[4]

From a theoretical perspective, posterior implementation is interesting because it assumes no *fixed* informational position between Bayesian and ex post implementation. Indeed, the meaning of posterior implementation changes with the message space of the underlying mechanism: the fewer messages an agent uses in equilibrium, the less information can be inferred about her private information and the closer posterior implementation moves to Bayesian implementation on its informational basis. By contrast, in a direct revelation mechanism where agents truthfully report their information, posterior implementation exactly coincides with ex post implementation. Thus, a better understanding of posterior implementation can shed

---

impossibility of ex post implementation immediately implies the impossibility of implementation in dominant strategies.

2. If there is no preference interdependence, i.e., if values are private, then posterior implementation, ex post implementation, and implementation in dominant strategies are all equivalent.

3. Indeed, even if the planner intends the use of sequential voting or deliberation procedures, she often cannot specify in advance in which order the mechanism is going to be played—who votes or talks first, second, and so forth. The voting order may be random, e.g., due to a committee's particular seating arrangement, or it may be subject to strategic manipulation, e.g., by the committee chair.

4. This characterization and interpretation of posterior implementation are new. Green and Laffont (1987) give a different motivation for posterior implementation by arguing that it is the appropriate solution concept for certain situations where agents lack commitment to their actions; see their introduction for the details.

light on the restrictions that Bayesian implementability imposes on social choice functions (when posterior implementation is close to Bayesian implementation) and on the maximum degree of information revelation that still allows for non-trivial implementation (when posterior implementation is close to ex post implementation).

This paper provides three main insights into posterior implementation. The first insight is how an earlier characterization of posterior implementation can be generalized from $n = 2$ to $n \geq 3$ agents. Green and Laffont (1987) show that every posterior implementable social choice function for $n = 2$ agents is characterized by a decreasing step function that partitions the type space into two components on each of which the social choice is constant. Instead of generalizing this *geometric* characterization to higher-dimensional spaces, this paper gives a simple *economic* characterization of posterior implementation: every (responsive) posterior implementable social choice function is posterior implementable by score voting. In a score voting mechanism, each agent submits a number from a set of consecutive integers; the alternative is accepted if and only if the sum of these numbers exceeds a pre-specified quota.[5] Thus, score voting requires that agents transmit their private information in coarse categories, but it allows finer information transmission than majority voting. The characterization is obtained under the appropriate generalizations of Green and Laffont's assumptions to the many-agent case (monotone and heterogeneous preferences; continuous and affiliated types).

The second insight is that the possibility of posterior implementation depends crucially on the number of agents. In generic environments with $n \geq 3$ agents, a (responsive) social choice function is posterior implementable if and only if it is Bayesian implementable by unanimity voting.[6] By contrast, with $n = 2$ agents, every monotone and deterministic social choice function that is Bayesian implementable is also posterior implementable; the set of these social choice functions is generally much richer than those that are implementable by unanimity voting. Thus, posterior implementation becomes very stringent as one moves beyond the two-agent setting of Green and Laffont (1987). Some further results on the existence of Bayesian and posterior implementable social choice functions are presented in the paper—the proof technique via the Poincaré-Miranda theorem might be of independent interest to voting theory.

The third insight is that two well-known results from the literature are intimately related to posterior implementation:

(1) Dekel and Piccione (2000) show that with informationally symmetric agents, any symmetric Bayes-Nash equilibrium of a simultaneous majority voting game is also a perfect Bayesian equilibrium for any voting order in the associated sequential voting game. Their result can be recovered from the present paper by

5. Score voting is perhaps the most common name for the mechanism described above; see the corresponding Wikipedia page for a brief overview and examples of practical use.

6. The result is shown to hold in both a topological and measure-theoretic sense of genericity.

verifying the following two observations: their notion of "sequential robustness" is equivalent to posterior implementation; given their symmetry assumptions, posterior implementation and Bayesian implementation via majority voting are equivalent. The genericity analysis in the present paper now implies that symmetric environments are a knife-edge case: in general, posterior implementation and Bayesian implementation are not equivalent—the voting order typically matters, where unanimity voting is the *unique* exception.

(2) Li, Rosen, and Suen (2001) characterize monotone and deterministic social choice functions that are Bayesian implementable in the "Condorcet jury model" with two agents. They show that any such function must be a "partition outcome" (in the sense of being characterized by a decreasing step function). Their result is an immediate corollary from the present paper, which embeds the Condorcet jury model as a special case: partition outcomes are exactly those outcomes that are Bayesian implementable via score voting. The result remains valid even beyond the Condorcet jury model due to the underlying equivalence of monotone Bayesian implementation and posterior implementation for $n = 2$ agents. However, this equivalence breaks down for $n \geq 3$ agents, signifying the difficulty of characterizing Bayesian implementation in collective decision problems with an arbitrary number of agents.

This paper proceeds as follows. Section 2.2 introduces the model and the concept of posterior implementation. Section 2.3 presents the characterization of posterior implementation in terms of score voting mechanisms. Section 2.4 discusses the (non-)existence of posterior implementable social choice functions and the striking difference between the two-agent and many-agent cases. Section 2.5 shows that robustness against the extensive form of the mechanism is the characterizing property of posterior implementation. Section 2.6 relates these findings to work by Dekel and Piccione (2000) and Li, Rosen, and Suen (2001). Section 2.7 concludes.

2.1.1 Related Literature

The literature on posterior implementation is sparse. Green and Laffont (1987) introduce the concept and provide a geometric characterization for two-agent decision problems. Lopomo (2001) shows that the symmetric equilibrium outcome of the English auction (with a suitable reserve price) is revenue-optimal among all posterior implementable outcomes of auctions. Jehiel et al. (2007) give an example of an auction with multi-dimensional types showing that posterior implementation can be possible when ex post implementation is impossible. Kawakami (2016) studies posterior implementation along with a certain type of renegotiation-proofness in the two-agent decision problem of Green and Laffont.[7]

7. The informational considerations underlying renegotiation-proofness share some similarities with posterior implementation. Suppose that agents are negotiating the use of an alternative mecha-

The literature on implementability in collective decision problems with preference interdependence and heterogeneity is equally sparse. Feng, Niemeyer, and Wu (2023) show that non-trivial ex post implementation is impossible. Feng and Wu (2020) study implementation in interim dominant strategies, which neither implies nor is implied by ex post or posterior implementation. They provide a characterization of the concept in terms of certain (yes/no)-voting mechanisms. Finally, there do not seem to be any follow-up studies on the Bayesian implementability result of Li, Rosen, and Suen (2001).

Other work following Dekel and Piccione (2000) discusses limitations of their equivalence result *within* the class of symmetric environments. Battaglini (2005) shows that the result no longer holds with costly voting and abstention. Ali and Kartik (2012) construct an equilibrium in the sequential voting games of Dekel and Piccione that exhibits herding. Callander (2007) studies herding when voters have a preference to vote for the winning alternative.

Variations of score voting have been studied axiomatically as aggregation rules for ordinal preferences over multiple alternatives; see Myerson (1995), Gaertner and Xu (2012), and Macé (2018). In the present paper, score voting mechanisms are studied for an entirely different reason. They emerge here as a solution to a particular implementation problem *with interdependent values*, as a way for agents to communicate continuous information in coarse categories. The only other work that discusses something similar to score voting in the context of implementation is Li, Rosen, and Suen (2001).

2.2 Preliminaries

2.2.1 Model

A group of $n \geq 2$ agents must decide whether to accept or reject an alternative. The valuation $v_i(\theta)$ of agent $i \in N = \{1, \ldots, n\}$ for the alternative depends on an unknown state of the world $\theta \in \Theta$. Agent i weakly prefers to accept the alternative in state θ if and only if her valuation $v_i(\theta)$ is non-negative.

Each agent only has partial information about the payoff-relevant state θ. Specifically, θ consists of n components, $\theta = (\theta_1, \ldots, \theta_n)$, and agent i only observes θ_i—her type. Assume that θ_i is a real number from a compact interval Θ_i. Without loss of generality, normalize $\Theta_i = [0, 1]$. Thus, the type (state) space is given by

nism versus a status quo mechanism. In this case, they draw inferences about the information of others from the (prospective) outcome of the negotiation, and these inferences affect equilibrium behavior in the mechanism that is played following the negotiations; see Holmström and Myerson (1983), Crawford (1985), Forges (1994), or Cramton and Palfrey (1995). Similarly, with posterior implementation, agents draw inferences about the information of others from equilibrium behavior in the given mechanism, and equilibrium behavior is then required to remain unaffected by these inferences. Kawakami (2016) discusses the relationship in more detail.

$\Theta = \prod_{i \in N} \Theta_i = [0,1]^n$. It is common knowledge that states are distributed according to a probability measure $\mu \in \Delta(\Theta)$ with a continuously differentiable and strictly positive density f.[8]

The agents use a mechanism without transfers to make a collective decision. A *mechanism* (M, ψ) is a collection of measurable message spaces M_i, one for each agent i, and a measurable outcome function $\psi : M \to [0,1]$ that assigns an acceptance probability to every message profile $m \in M = \prod_{i \in N} M_i$. A mechanism induces a game of incomplete information in which agents maximize expected utility.

2.2.2 Assumptions

Assume that each valuation function $v_i : \Theta \to \mathbb{R}$ is continuously differentiable. Let $\nabla v_i : \Theta \to \mathbb{R}^n$ denote the gradient. The following assumptions generalize the ones by Green and Laffont (1987) to settings with $n \geq 2$ agents and will be maintained throughout:

- *monotonicity*: for each agent $i \in N$, ∇v_i is strictly positive in every component;
- *heterogeneity*: if $v_i(\theta) = 0$ for all $i \in N$, then $\left(\nabla v_i(\theta)\right)_{i \in N}$ are not collinear;
- *affiliation*: agents' types $(\theta_1, \dots, \theta_n)$ are affiliated, i.e., for all $\theta, \theta' \in \Theta$,

$$f(\theta) f(\theta') \leq f(\theta \wedge \theta') f(\theta \vee \theta').$$[9]

Monotonicity is a sorting condition; it ensures that agents are more inclined to accept the alternative in higher states. Heterogeneity captures the idea that agents may attach importance to different dimensions of a state and that some agents may require stronger evidence than others to accept the alternative. Specifically, for each state in which all agents are indifferent towards the alternative, there exists a nearby state and two agents, one of whom strictly prefers the alternative and the other the status quo. Thus, heterogeneity is a mild assumption because it merely rules out pure common values (for which the exercise of implementability is pointless because first-best can always be achieved in ex post equilibrium). Affiliation, a form of positive correlation, ensures that monotonicity is preserved by taking conditional expectations, i.e., that each agent's expected valuation is also monotone with respect to her partial information about the state.

2.2.3 The concept of posterior implementation

In order to define posterior implementation, one must first specify the posterior beliefs that an agent holds after observing the equilibrium behavior of others.

8. Any topological space in this paper is endowed with its Borel σ-algebra. For a measurable space X, let $\Delta(X)$ denote the space of probability measures on X. For $\mu \in \Delta(X)$ and a measurable set $B \subset X$, let $\mu[B]$ denote the associated probability.

9. $\theta \wedge \theta'$ ($\theta \vee \theta'$) denotes the component-wise minimum (maximum) of θ and θ'.

A *strategy* for agent i in (the game induced by) mechanism (M, ψ) is a function (formally, Markov kernel) $\sigma_i : \Theta_i \to \Delta(M_i)$. As usual, for some n-tuple $\theta = (\theta_1, \ldots, \theta_n)$, write $\theta_{-i} = (\theta_1, \ldots \theta_{i-1}, \theta_{i+1}, \ldots, \theta_n)$; more generally, for a subset $J \subset N$ of agents, write $\theta_J = (\theta_i)_{i \in J}$ and $\theta_{-J} = (\theta_i)_{i \in N \setminus J}$. Let $\mu(\cdot|\theta_i) \in \Delta(\Theta_{-i})$ denote the belief that agent i holds about the types of other agents when her own type is θ_i. When i knows that other agents play strategies σ_{-i} and observes messages $m_{-i} \in M_{-i}$, she forms a posterior belief $\mu(\cdot|\theta_i, m_{-i}) \in \Delta(\Theta_{-i})$. Assume that posterior beliefs are derived via Bayes' rule whenever possible.[10]

Let

$$V_i(\theta_i|m_{-i}) \;=\; \int_{\Theta_{-i}} v_i(\theta_i, \theta_{-i})\, \mu(d\theta_{-i}|\theta_i, m_{-i}) \tag{2.1}$$

denote the *posterior expected valuation* of agent i given m_{-i} when she is of type θ_i.

Definition 2.1. A strategy profile $\sigma = (\sigma_i)_{i \in N}$ in (M, ψ) is a *posterior equilibrium* if for all $i \in N$, $\theta_i \in \Theta_i$, $m_{-i} \in M_{-i}$, and $\tilde{m}_i \in M_i$

$$V_i(\theta_i|m_{-i})\psi(\sigma_i(\theta_i), m_{-i}) \;\geq\; V_i(\theta_i|m_{-i})\psi(\tilde{m}_i, m_{-i}).^{11} \tag{2.2}$$

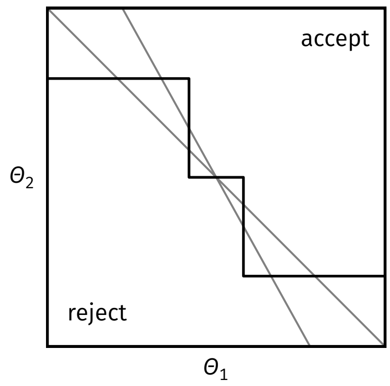
In a posterior equilibrium, the strategy σ_i of each agent i is optimal against the strategies σ_{-i} of other agents for every possible message profile m_{-i}.

A *social choice function* is a measurable function $\phi : \Theta \to [0, 1]$ that assigns an acceptance probability in $[0, 1]$ to every state $\theta \in \Theta$. A tuple (M, ψ, σ) is an *implementation* of ϕ if $\psi \circ \sigma = \phi$ μ-almost everywhere. It is a *posterior implementation* if, in addition, σ is a posterior equilibrium in (M, ψ). A social choice function ϕ is *posterior implementable* if there exists a posterior implementation of ϕ. Posterior implementation is stronger than Bayesian implementation but weaker than ex post implementation; the definitions of these two standard concepts are relegated to Section 2.B.

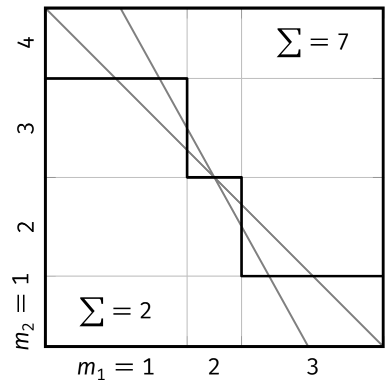
Remark 2.1 (Revelation principle)**.** The standard revelation principle does not apply to posterior implementation. Indeed, suppose that some posterior implementable social choice function is posterior implementable by a direct revelation mechanism in which agents truthfully report their types. Since agents perfectly learn

10. Bayes' rule might not apply if agents play pathological strategies. Nevertheless, posterior beliefs still exist and are uniquely determined for almost all $\theta_i \in \Theta_i$ and $m_{-i} \in M_{-i}$; see Section 2.A for the formal details. In this case, assume that the undetermined beliefs are "consistent" with affiliation in that they preserve the monotonicity of valuation functions under conditional expectations (Assumption 2.1). This assumption should be understood purely as a technical addition to the affiliation assumption. Indeed, Section 2.A shows that the desired beliefs always exist if the prior density is affiliated.

11. The composition $\psi \circ \sigma$ in this expression is understood to be a composition of Markov kernels, i.e., $(\psi \circ \sigma)(\theta) = \int_M \psi(m)\, \sigma(\theta)[dm]$.



(a) GLs geometric characterization.



(b) Characterization in terms of score voting.

Figure 2.1. A posterior implementable social choice function (left) and how it can be understood as the outcome of a score voting mechanism that accepts the alternative if and only if $m_1 + m_2 > 4$, where m_i is the integer message of agent i (right). In black: the step function associated with a posterior implementable social choice function. In grey: the agents' indifference curves.

the state from these reports, truth-telling by every agent is then a posterior equilibrium if and only if it is an ex post equilibrium. A result by Feng, Niemeyer, and Wu (2023) shows that in the present environment, only constant social choice functions are ex post implementable. By contrast, non-constant posterior implementable social choice functions typically exist.

2.3 Posterior Implementation and Score Voting

This section gives a characterization of posterior implementable social choice functions in terms of score voting mechanisms. The first subsection briefly recalls the theorem of Green and Laffont (1987) (GL) for $n = 2$ agents and sketches how their result can be related to score voting. Then, score voting is defined formally, the characterization result is presented, and finally, posterior equilibrium behavior in score voting mechanisms is analyzed.

2.3.1 The Green-Laffont theorem

GL give the following geometric characterization for $n = 2$ agents. Any posterior implementable social choice function is such that there exists (the graph of) a decreasing step function that partitions the type space $[0, 1] \times [0, 1]$ into two components on each of which the social choice is constant; see Figure 2.1a. In addition, any such step function must solve a particular system of non-linear equations that reflects the posterior equilibrium conditions. GLs characterization is based on essentially the same assumptions as those introduced in the previous section. The general case of $n \geq 3$ agents has remained an open question.

A reasonable conjecture for the n-agent case might be that posterior implementable social choice functions are again characterized by higher-dimensional analogues of decreasing step functions. However, it is unclear how to describe these analogues in a tractable way. Therefore, instead of generalizing GLs *geometric* characterization, this paper gives a simple *economic* characterization of posterior implementation that applies to the n-agent case.

To get a sense of this economic characterization, consider again GLs characterization in Figure 2.1a. Imagine projecting the "steps" of the step function onto the axes corresponding to the type space of each agent; there are three steps for agent 1 and four steps for agent 2, and each step yields an interval; see Figure 2.1b. Now imagine the strategy profile where types in the same interval send the same message m_i. Identify these messages with consecutive integers: $m_1 \in \{1,2,3\}$ for agent 1, $m_2 \in \{1,2,3,4\}$ for agent 2. Consider the mechanism that accepts the alternative if and only if the sum $m_1 + m_2$ of both agents' messages exceeds the quota $q = 4$; this mechanism is a particular form of *score voting*. The constructed mechanism and strategy profile implement the social choice function that is characterized by the decreasing step function in Figure 2.1a.

2.3.2 Score voting

In a score voting mechanism, each agent submits a score from a set of consecutive integers, the scores are added, and the alternative is accepted if and only if the sum exceeds a pre-specified quota. The more scores an agent has available, the more influence she can exert over the decision and the more detailed she can transmit her private information to the mechanism.

Definition 2.2. A mechanism (M, ψ) is a *score voting* mechanism if

(1) for all $i \in N$, $M_i = \{1, \dots, |M_i|\}$ is a set of consecutive integers;

(2) there exists a quota $q \in \mathbb{Z}$ and real numbers $0 \leq \alpha < \beta \leq 1$ such that

$$\psi(m) = \begin{cases} \beta & \text{if } \sum_{i \in N} m_i > q, \\ \alpha & \text{else;} \end{cases}$$

(3) each agent $i \in N$ has at most one *veto message* $\bar{m}_i \in M_i$ such that $\psi(\bar{m}_i, \cdot) = \beta$ and at most one *veto message* $\underline{m}_i \in M_i$ such that $\psi(\underline{m}_i, \cdot) = \alpha$. (This last requirement is only for parsimony; there should be no redundant messages.)

The following mechanisms are examples of score voting mechanisms:

- *i-dictatorship*: $M_i = \{0,1\}$, $M_j = \{0\}$ for all $j \neq i$, and $q = 0$;
- *unanimity for acceptance (rejection)*: $M_i = \{0,1\}$ for all i and $q = n-1$ $(q = 0)$;
- *super- or submajority*: $M_i = \{0,1\}$ for all i and $q \in [0, n-1]$;
- *simple majority with abstention*: $M_i = \{-1,0,1\}$ for all i and $q = 0$.

Remark 2.2 (Abstention)**.** Super- or submajority mechanisms *with abstention*—except for simple majority—are generally *not* score voting mechanisms. The reason is that in a super- or submajority mechanism, the number of affirmative votes needed to pass the alternative is proportional to the overall number of votes cast. However, in a score voting mechanism, affirmative votes must exceed negative votes by a pre-specified quota, independently of the number of abstentions.

2.3.3 Characterization theorem

The following property of social choice functions is used in the formulation of the theorem and will be discussed right after. A social choice function ϕ is *responsive* if it can only be implemented by giving each agent at least two messages, i.e., if every implementation (M, ψ, σ) of ϕ satisfies $|M_i| \geq 2$ for all $i \in N$. In other words, a responsive social choice function must take into account the information of every agent.

A pure strategy profile $\sigma : \Theta \to M$ is said to be *surjective* if for every message profile $m \in M$ there exists a type profile $\theta \in \Theta$ such that $\sigma(\theta) = m$.

Theorem 2.1. *Let $n \geq 3$. A responsive social choice function is posterior implementable if and only if it is posterior implementable by score voting in pure surjective strategies.*

Proof. See Section 2.D. □

Score voting only allows information transmission in coarse categories; thus, posterior implementation requires that agents garble their private information. In particular, posterior implementation must always be inefficient. This garbling of information can intuitively be understood by relating posterior implementation to information transmission in sender-receiver games. In the context of posterior implementation, each agent is a sender of information, yet each agent can also be viewed as a receiver of information who can sometimes decide the outcome after receiving the messages of everyone else. Thus, each sender must strike a balance between sharing information for better collective decision-making and withholding information that could be used by the receivers against the sender's own interests. In equilibrium, this trade-off is resolved by the transmission of private information in coarse categories, which is reminiscent of the partition equilibria studied in Crawford and Sobel (1982).

The formal proof of Theorem 2.1 uses a recent result about ex post implementation without transfers by Feng, Niemeyer, and Wu (2023). In the present environment, their result implies that only constant social choice functions are ex post implementable. Since posterior implementation in a direct revelation mechanism is equivalent to ex post implementation, information must be garbled by at least one agent at least somewhere in the type space. This local garbling introduces discontinuities in the other agents' expected valuations, and around such discontinuities, the frontier between acceptance and rejection in type space cannot depend smoothly on the other agents' reports—they must also garble their information. If there are $n \geq 3$ agents, then it can be shown that there are only finitely many information pools for each agent. The exact form of score voting then follows from the monotonicity in the environment.[12]

Remark 2.3 (Responsiveness)**.** Responsiveness is a mild requirement for social choice functions because it merely excludes the presence of dummy agents who can (almost) never affect the decision. Nevertheless, Theorem 2.1 essentially covers non-responsive social choice functions. To see this, note that a dummy agent who can never affect the decision can be removed from the model by modifying the valuation functions of everyone else to be expectations with respect to the dummy's information. In the resulting environment, the assumptions of monotonicity and affiliation are satisfied by standard results about affiliation (see e.g. Milgrom and Weber, 1982), and heterogeneity will be satisfied in generic environments (see Lemma 2.15). Indeed, the "non-generic" environments are precisely why responsiveness is needed in the statement of Theorem 2.1: one can imagine that the environment becomes one of pure common values by integrating out the information of a particular subset of agents. In the resulting common value environment, one can then implement first-best in ex post equilibrium. This is no longer possible as soon as every agent has variation in the information that she transmits in equilibrium.

The statement of Theorem 2.1 holds for $n = 2$ agents as long as their indifference curves do not intersect. If the agents' indifference curves do intersect, then they might be able to transmit information in infinitely many categories; see Figure 2.2 or Green and Laffont (1987, Figure 7). Intuitively, around a point of intersection, the agents' conflict of interest vanishes, allowing information to be transmitted in increasingly finer categories. This echoes how there can be infinite partition equilibria in the model of Crawford and Sobel (1982) if the bias between sender and receiver changes sign as a function of the state; see Gordon (2010) for a detailed treatment. These observations are summarized in the following lemma.

Lemma 2.1. *Let $n = 2$. A social choice function ϕ is posterior implementable with finitely many messages if and only if it is posterior implementable via score voting in pure surjective strategies. Moreover, if there is no $\theta \in \Theta$ such that $v_1(\theta) = v_2(\theta) = 0$, then ϕ is posterior implementable with finitely many messages.*

Proof. See Section 2.D. □

12. Due to the pooling of private information into integer values, one can think about posterior implementation as ex post implementation on a discretized state space, where the discretization is endogenously determined by the posterior implementation under consideration. Indeed, non-trivial ex post implementation is sometimes possible on discrete state spaces; see the discussion in Feng, Niemeyer, and Wu (2023).

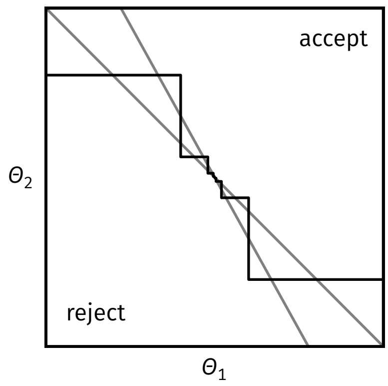


Figure 2.2. Infinitely many steps in GLs geometric characterization. The display is otherwise as in Figure 2.1.

Thus, with $n = 2$ agents, the set of posterior implementable social choice functions might generally be slightly larger than the set of social choice functions that are posterior implementable via score voting. That the the two-agent case is special is no coincidence and will be thoroughly discussed in the next section.

2.3.4 Posterior equilibria of score voting

This last subsection characterizes the pure strategy posterior equilibria of score voting mechanisms in terms of cutoff strategies; together with Theorem 2.1, this completes the characterization of posterior implementation.

In voting games, it is often useful to ponder an agent's pivotal events—the situations in which an agent can affect the collective decision by changing her own vote. The upcoming result for score voting is also conveniently formulated in the language of pivotal events.

For a given score voting mechanism (M, ψ) with quota q, define

$$PIV_i = \left\{ m \in M \;\middle|\; \left(\sum_{j \in N} m_j = q \right) \wedge \left(\, m_i < |M_i| \right) \right\}$$

to be set of messages profiles m for which the alternative is one point short of being accepted and such that agent i can enforce the alternative by submitting a higher score. The message profiles in PIV_i will be called the *pivotal events* of agent i.

A pure surjective strategy σ_i in a score voting mechanism (M, ψ) is a *cutoff strategy* if there exist ordered cutoff types $\theta_i^0 = 0 \leq \theta_i^1 < \ldots < \theta_i^{|M_i|-1} \leq 1 = \theta_i^{|M_i|}$ such that $\operatorname{cl} \sigma_i^{-1}(m_i) = [\theta_i^{m_i - 1}, \theta_i^{m_i}]$. (The behavior of cutoff types is irrelevant for implementation.) In particular, for cutoff strategies, the information encoded in any given message is an interval of types.

Lemma 2.2. *A pure surjective strategy profile $\sigma = (\sigma_i)_{i \in N}$ in a score voting mechanism (M, ψ) is a posterior equilibrium if and only if each σ_i is a cutoff strategy such* *that for all $m \in PIV_i$,*

$$V_i(\theta_i^{m_i} | [\theta_j^{m_j - 1}, \theta_j^{m_j}]_{j \neq i}) \;=\; 0. \tag{2.3}$$

Proof. See Section 2.D. □

Lemma 2.2 says the following. Suppose agent i is pivotal for message profile m. Then the cutoff type $\theta_i^{m_i} = \sup \sigma_i^{-1}(m_i)$ at which agent i switches from submitting m_i to submitting $m_i + 1$ must be indifferent between accepting and rejecting the alternative, given the information that can be inferred from observing the messages m_{-i} of others. Conversely, by the monotonicity in the environment, if cutoff types are indifferent, then each type of each agent gets her preferred choice in every pivotal event.

While posterior equilibrium requires that agents get their preferred choice in *every* situation in which they are pivotal, Bayes-Nash equilibrium requires that agents get their preferred choice *on average* over all the situations in which they are pivotal. This characterization of Bayes-Nash equilibrium—the pivotal voting argument—is familiar from work on majority voting (see Austen-Smith and Banks, 1996; Feddersen and Pesendorfer, 1997). The argument will be relevant for some results in the next section and is thus formally stated in Section 2.D.4; see Lemma 2.14.

Note that Theorem 2.1 and Lemma 2.2 together imply that every posterior implementable social choice function is *deterministic* (in the sense of taking at most two values) and non-decreasing.

2.4 The Limits of Posterior Implementation

This section discusses the existence of posterior implementable social choice functions. There are two main results. First, in generic environments with three or more agents, any responsive posterior implementable social choice function is posterior implementable by unanimity voting (Theorem 2.2). Second, with two agents, any non-decreasing and deterministic social choice function that is Bayesian implementable is also posterior implementable (Theorem 2.3). Thus, with $n = 2$ agents, posterior implementation is not significantly more demanding than Bayesian implementation, whereas with $n \geq 3$ agents, posterior implementation is not significantly more permissive than ex post implementation (which is impossible, even via unanimity voting). Besides the two-agent case, this section also discusses the other exceptions to Theorem 2.2—why posterior implementation by unanimity voting is typically possible and why non-generic environments can allow one to go beyond unanimity voting.

2.4.1 Genericity analysis

Some preliminary definitions are needed in order to make precise what is meant by genericity. Let $C^1(\Theta, \mathbb{R}^n)$ denote the (Banach) space of continuously differentiable functions $\Theta \to \mathbb{R}^n$ equipped with the topology of uniform convergence of functions and their first derivatives.[13] Let $\mathscr{V} \subset C^1(\Theta, \mathbb{R}^n)$ denote the *open* subset of valuation profiles $v = (v_1, \ldots, v_n)$ that satisfy monotonicity (but not necessarily heterogeneity).

The result is shown for two notions of genericity.[14] The first notion is the standard topological one: a subset of $\mathscr{V}$ is *residual* if its complement is a countable union of nowhere dense sets. The second notion—prevalence—is a measure-theoretic one due to Hunt, Sauer, and Yorke (1992). Prevalence generalizes to infinite-dimensional spaces the idea that a subset of $\mathbb{R}^n$ should be considered large if its complement has Lebesgue measure zero. Formally, a Borel subset $S \subset \mathscr{V}$ is *shy* if there exists a Borel measure on $\mathscr{V}$ that is strictly positive and finite on a compact subset of $\mathscr{V}$ and assigns measure zero to every translate of S. A subset of $\mathscr{V}$ is called *prevalent* if its complement is contained in a shy Borel set.

Say that a property holds *in generic environments* if for every given affiliated prior density f there exists a residual and prevalent subset $\mathscr{G} \subset \mathscr{V}$ of valuation profiles such that the property holds whenever $v \in \mathscr{G}$.

Theorem 2.2. *In generic environments with $n \geq 3$ agents, every responsive posterior implementable social choice function is posterior implementable by unanimity voting.*

Proof. See Section 2.E.1. □

The intuition is simple. By Lemma 2.2, the posterior equilibria of score voting can be characterized in terms of ordered cutoff types. The key observation is that in every score voting mechanism—except the unanimity mechanism—there are strictly more pivotal events and hence equilibrium conditions than there are cutoff types. The reason is that for each cutoff type, there are multiple pivotal events for which this cutoff type must be indifferent between accepting and rejecting the alternative. For example, in majority voting, these pivotal events are all the ways in which an equal number of yes and no votes can be distributed across the other voters. Thus, the posterior equilibrium conditions imply an overdetermined system of equations, and this system is shown to not have a solution in generic environments using the transversality theorem.

Remark 2.4 (Responsiveness)**.** A similar remark about responsiveness as for Theorem 2.1 also applies to Theorem 2.2: in generic environments, any posterior implementable social choice function that is responsive to the information of at least three agents is posterior implementable by unanimity voting (immediate corollary from Lemma 2.15). Social choice functions that only respond to the information of two agents are discussed in the next subsection. Moreover, it is always possible to

13. The topology-inducing norm is $||v||_\infty + ||Dv||_\infty$, where Dv is the Jacobian of $v \in C^1(\Theta, \mathbb{R}^n)$.

14. For other applications of these notions see e.g. Heifetz and Neeman (2006), Jehiel et al. (2006), and Reny and Perry (2006).

implement dictatorial social choice functions in posterior equilibrium; such social choice functions only respond to the information of a single agent.

2.4.2 Two agents

Both Theorem 2.1 and Theorem 2.2 assume three agents in the environment, indicating that the two-agent case is special. The peculiarity lies in the following result.

Theorem 2.3. *Let $n = 2$. A non-decreasing and deterministic social choice function is Bayesian implementable if and only if it is posterior implementable.*

Proof. See Section 2.E.2. □

Thus, for $n = 2$ agents, posterior implementable social choice functions as characterized by GL are exactly the non-decreasing and deterministic social choice functions that are Bayesian implementable. These social choice functions are a natural object of study since any remotely efficient social choice function is deterministic and, given monotone preferences, is itself monotone. Thus, even though the set of Bayesian implementable social choice functions may be larger than the set of posterior implementable social choice functions, the two concepts are not too far apart when there are only two agents. In particular, efficient Bayesian implementation, like efficient posterior implementation, is impossible.

The observation at the heart of Theorem 2.3 is the following: if there are only two agents, then the number of pivotal events and cutoff types are the same in every score voting mechanism. The reason is that there is at most one pivotal event associated with any particular message, i.e., for each message $m_i \in M_i$, there is exactly one message $m_{-i} \in M_{-i}$ of the other agent such that $m_i + m_{-i} = q$. As a consequence, every Bayes-Nash equilibrium of score voting is a posterior equilibrium (as long as every message is played with positive probability). This is because an agent understands that the precise message she submits is only relevant in the unique event in which she is pivotal; she conditions her vote on the information contained in the pivotal event, which is the same information as the one derived from actually observing the message of the other agent. Theorem 2.3 shows that this link between posterior implementation and Bayesian implementation also exists outside of score voting mechanisms; the key argument is a suitable generalization of the "pivotal voting argument" from score voting to more general mechanisms.

It remains to discuss formally whether the set of posterior implementable social choice functions for $n = 2$ agents is actually richer than the set of those functions that are implementable by unanimity voting.[15] The fact that the number of pivotal events

15. GL do not provide conditions for the existence of posterior implementable social choice functions. Kawakami (2016) discusses the existence of social choice functions that are posterior implementable by unanimity voting in the two-agent case.

and equilibrium cutoffs is the same—that the system of posterior equilibrium conditions is no longer overdetermined—suggests an affirmative answer. Indeed, under a mild assumption on agents' preferences, one can ensure for *every* score voting mechanism the existence of a pure surjective posterior equilibrium, i.e., the existence of posterior implementable social choice functions with any arbitrary number of steps.

Say that the environment features *partisan types* if for all $i \in N$, type $\theta_i = 0$ prefers rejecting the alternative regardless of other agents' types, and type $\theta_i = 1$ prefers accepting the alternative regardless of other agents' types.

Proposition 2.1. *Let $n = 2$. If the environment features partisan types, then for every score voting mechanism (M, ψ), there exists a social choice function that is posterior implementable via (M, ψ) in cutoff strategies.*

Proof. See Section 2.E.2. □

It may seem paradoxical that partisan types—indicating a strong conflict of interest—are nonetheless a sufficient condition for the conflict of interest to be small enough for meaningful information transmission. This paradox is resolved by noting that for environments with partisan types, the Poincaré-Miranda theorem[16] implies the existence of a type profile where all agents are indifferent between accepting and rejecting the alternative, i.e., where the conflict of interest vanishes locally. The Poincaré-Miranda theorem, which is a higher-dimensional analogue of the intermediate value theorem and equivalent to Brouwer's fixed point theorem, also turns out to be a convenient tool for proving equilibrium existence in (asymmetric) voting games; see the next subsection.

In summary, the previous results show a stark contrast between the two-agent and many-agent cases. The reason for this contrast is the non-equivalence of posterior implementation and monotone Bayesian implementation for more than two agents. Indeed, posterior implementation becomes harder to achieve when there are more than two agents because posterior optimality must be checked against a multi-dimensional rather than a one-dimensional family of pivotal events. In contrast, Bayesian implementation becomes easier to achieve because each agent becomes uncertain about multiple dimensions of information rather than a single dimension. Example 2.1 in Section 2.C shows that this additional uncertainty can be exploited to sometimes (though not typically) implement even efficient social choice functions in Bayes-Nash equilibrium; this is impossible for $n = 2$ agents.

16. The theorem states that for n continuous real-valued functions $f_1, \ldots, f_n$ on the unit cube $[0,1]^n$, if each f_i is non-positive if $x_i = 0$ and non-negative if $x_i = 1$, i.e., changes sign on the corresponding opposite faces of the cube, then there exists a point $x \in [0,1]^n$ such that $f_i(x) = 0$ for all $i = 1, \ldots, n$. Ekmekci, Heese, and Lauermann (2022) provide a generalization that might prove useful in economic applications.

2.4.3 Unanimity voting

What makes the unanimity mechanism special? For $n \geq 3$, unanimity is the only score voting mechanism where the number of pivotal events and cutoff types are exactly equal because an agent is pivotal if and only if all other agents vote in unison. In analogy to the two-agent case, any Bayes-Nash equilibrium of unanimity (in which agents are pivotal with positive probability) is thus a posterior equilibrium because an agent understands that her vote only matters in the event that she is pivotal; she conditions her vote on the information that is contained in the pivotal event, which is the same information as the one contained in the actual messages of others. This is summarized in the following lemma.

Proposition 2.2. *A social choice function is Bayesian implementable by unanimity voting if and only if it is posterior implementable by unanimity voting.*

Proof. See Section 2.E.3. □

When does unanimity admit a Bayes-Nash equilibrium—and thus a posterior equilibrium—that implements a responsive social choice function? The answer clarifies that the impossibility result for ex post implementation by Feng, Niemeyer, and Wu (2023) and Theorem 2.2, which can be viewed as an "almost" impossibility result for posterior implementation, are different. While Feng, Niemeyer, and Wu (2023) show that non-constant ex post implementation is impossible in the environments discussed in the present paper, Theorem 2.2 still leaves room for non-constant posterior implementation in some knife-edge cases and, as the following lemma shows, via the unanimity mechanism under mild assumptions on agents' preferences, e.g., the earlier assumption of partisan types.

Proposition 2.3. *If the environment features partisan types, then there exists a responsive social choice function that is posterior implementable by unanimity voting in cutoff strategies.*

Proof. See Section 2.E.3. □

The lemma is readily proven by applying the Poincaré-Miranda theorem to posterior expected valuations considered as functions on the space of cutoff types. The same technique can be used to show that partisan types imply for *every* super- or submajority mechanism the existence of a responsive Bayes-Nash equilibrium in cutoff strategies.

2.4.4 Symmetric environments

Recall that there are more pivotal events than cutoff types in every score voting mechanism except unanimity. One way in which an environment can nevertheless admit solutions to the (overdetermined) system of posterior equilibrium conditions is for it to render some equilibrium conditions redundant. This is possible within the class of symmetric environments.

Say that the environment is *symmetric* if

(1) the prior density f is symmetric (permutation-invariant) in all of its arguments;

(2) there exists a valuation function $\tilde{v} : \Theta \to \mathbb{R}$ that is symmetric in the last $(n-1)$ arguments such that $v_i(\theta) = \tilde{v}(\theta_i, \theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_n)$ for all $i \in N$.

A social choice function is said to be *anonymous* if it is symmetric in all of its arguments, i.e., processes the agents' information independently of their identities.

Proposition 2.4. *In a symmetric environment, an anonymous social choice function is Bayesian implementable by a super- or submajority mechanism (M, ψ) if and only if it is posterior implementable by (M, ψ).*

Proof. See Section 2.E.4. □

The result is a tale of two symmetries: (1) symmetry of the environment ensures that each agent evaluates information about the other agents' types independently of their identities; (2) anonymity, which requires symmetric strategies, ensures that the same messages sent by different agents reveal the same information about these agents' types. Taken together, the two symmetries make agents evaluate all permutations of a given profile of others' messages equally. Finally, in a majority mechanism, all pivotal events are permutations of a single message profile (where the sum of (0/1)-messages is q); thus, the numerous posterior equilibrium conditions for the cutoff type to be indifferent between acceptance and rejection all collapse into a single Bayes-Nash equilibrium condition.

The existence of a responsive social choice function that is posterior (Bayesian) implementable by majority voting can be guaranteed under suitable assumptions on agents' preferences, e.g., the earlier assumption of partisan types. Thus, symmetric environments are indeed a non-generic exception to Theorem 2.2—the "almost" impossibility result for posterior implementation.

Proposition 2.5. *If the environment is symmetric and features partisan types, then for every super- or submajority mechanism (M, ψ), there exists an anonymous and responsive social choice function that is posterior implementable by (M, ψ) in cutoff strategies.*

Proof. See Section 2.E.4. □

The intermediate value theorem suffices to prove the result since, by symmetry, there is effectively only a single equilibrium cutoff and the single equilibrium condition that the cutoff type be indifferent when pivotal. The simplicity of the proof underscores why the Poincaré-Miranda theorem is a useful tool for proving equilibrium existence in asymmetric voting games.

Remark 2.5 (Beyond majority voting?)**.** The positive results for majority voting in symmetric environments do not extend to more general voting mechanisms. For example, recall that super- or submajority mechanisms with abstention—except for simple majority—are generally not score voting mechanisms. However, even simple majority need no longer admit a pure surjective posterior equilibrium in symmetric environments when abstention is allowed. The reason is that an agent's pivotal events need no longer be informationally equivalent since the information content of two abstentions need not be the same as the information content of a positive and negative vote, even for symmetric strategy profiles in symmetric environments; see Example 2.2 in Section 2.C. Thus, symmetric environments can only eliminate enough equilibrium conditions to allow implementation via super- or submajority mechanisms but not via more general mechanisms.

2.5 Extensive Form Mechanisms and Posterior Equilibrium

This section shows that posterior equilibrium is the exact solution concept that ensures the robustness of agents' equilibrium behavior against the extensive form of the mechanism. As discussed in the introduction, this interpretation of posterior equilibrium is relevant for practical applications of the theory. To formalize the idea, one must first define extensive form mechanisms and define what it means for agents' behavior to not depend on the extensive form.

An *extensive form* of mechanism (M, ψ) is an extensive game form such that

(1) nature draws agents' types at the initial history and privately informs each agent of her type;
(2) each agent $i \in N$ takes exactly one action $m_i \in M_i$;
(3) all actions are publicly observable.[17]

For example, the extensive forms of voting mechanisms correspond to the usual sequential voting games where each agent votes at a predetermined time, with some agents possibly voting simultaneously (see, e.g., Dekel and Piccione, 2000). (More general extensive forms are discussed at the end of this section.)

For a particular extensive form mechanism, let H denote the set of histories (where the move by nature is not included in the description of a history), let $H(i) \subset H$ denote the set of histories up to the move of agent $i \in N$, and let $m(h)$ denote the message profile associated with history $h \in H$, i.e., the associated unordered tuple of messages. Finally, let $\mu(\cdot|\theta_i, h) \in \Delta(\Theta_{-i})$ denote the belief that type θ_i holds about the types of other agents after observing history $h \in H(i)$.

17. Such games are called "multi-stage games with observed actions and incomplete information" in Fudenberg and Tirole (1991, Section 8.2.3).

To formalize the robustness of agents' equilibrium behavior against the extensive form of the mechanism, it is useful to define the following refinement of perfect Bayesian equilibrium (PBE), requiring that the behavior of each agent be independent of how all other agents behave.

A perfect Bayesian equilibrium (PBE) in (the game induced by) an extensive form of mechanism (M, ψ) is *history-independent* if

(1) the strategy profile $\sigma : \Theta \times H \to \Delta(M)$ satisfies for all $i \in N$, $\theta_i \in \Theta_i$, and $h, h' \in H(i)$,
$$\sigma_i(\theta_i, h) = \sigma_i(\theta_i, h') := \sigma_i(\theta_i);$$

(2) for each agent $i \in N$, type $\theta_i \in \Theta_i$, and terminal history $h \in H$, the terminal beliefs $\mu(\cdot|\theta_i, h) \in \Delta(\Theta_{-i})$ coincide with the posterior beliefs of agent i in the normal form mechanism:
$$\mu(\cdot|\theta_i, h) = \mu(\cdot|\theta_i, m_{-i}(h)).$$
(This requirement is automatically satisfied given history-independent strategies whenever terminal beliefs can be derived via Bayes' rule; thus, it constitutes a specific belief refinement for history-independent strategy profiles.)

Lemma 2.3. *A strategy profile σ is a posterior equilibrium in mechanism (M, ψ) if and only if σ is the strategy profile of a history-independent PBE in every extensive form of (M, ψ).*

Proof. See Section 2.F. □

The proof is simple. For history-independent strategy profiles, the beliefs held after terminal histories of an extensive form mechanism coincide with those held in the corresponding normal form of the mechanism after observing the message profile. Thus, an agent's sequential rationality conditions with respect to the information contained in the terminal histories are equivalent to her posterior equilibrium conditions in the normal form mechanism. Consequently, by the law of iterated expectations, posterior equilibrium implies sequential rationality for every agent, each of her types, and every previous history. Conversely, for each agent i, there is an extensive form in which i moves last; the corresponding sequential rationality conditions for agent i in a history-independent PBE then imply her posterior equilibrium conditions.

The key implication of the results from the previous section and Lemma 2.3 is that the precise extensive form of a mechanism, e.g., the order of voting, typically affects equilibrium outcomes and should thus be carefully specified in real-world applications. Unanimity (and trivially, dictatorship) is the unique exception. This implication is discussed in more detail in the next section.

Remark 2.6 (General extensive forms)**.** One can imagine more general extensive forms where agents might have multiple moves and where some actions are unobservable. Although such general extensive forms are tedious to formalize for normal form mechanisms with continuous action and type spaces, there is no conceptual difference from the simple extensive forms considered above.

2.6 Applications

This section relates the findings from the previous sections to work by Dekel and Piccione (2000) on sequential voting games and by Li, Rosen, and Suen (2001) on Bayesian implementation without transfers.

2.6.1 Sequential voting: Dekel and Piccione (2000)

An important result of Dekel and Piccione (2000) (DP) is that in symmetric environments, every symmetric Bayes-Nash equilibrium of a simultaneous majority voting game is also a history-independent perfect Bayesian equilibrium (PBE) for any voting order in the associated sequential voting game.

This result is an immediate corollary from two results in the present paper: (1) Lemma 2.3 shows that "sequential robustness", as studied by DP, is the characterizing property of posterior equilibrium; (2) Proposition 2.4 shows that for symmetric environments, anonymous Bayesian implementation and posterior implementation via majority voting are equivalent. DPs result can then be formulated as follows:[18]

Corollary 2.1. *In a symmetric environment, an anonymous social choice function is Bayesian implementable by a super- or submajority mechanism* (M, ψ) *if and only if it is implementable via a history-independent PBE in every extensive form of* (M, ψ).

As explained in Section 2.4.4, this particular implication of the pivotal voting argument is jointly driven by the symmetries in the environment and strategies. Theorem 2.2 implies that even the slightest asymmetry in the environment typically renders the various pivotal events of an agent informationally distinct. Therefore, it is typically beneficial for an agent to learn over the course of a sequential election about the specific event in which she might become pivotal. In other words, there is generally no strategic equivalence between simultaneous and sequential voting, except in unanimity mechanisms, which are generically the only mechanisms where equilibrium outcomes are robust against extensive-form play. This unique property of unanimity mechanisms may contribute to their widespread use in practice.

18. The result, as stated and proven in DP, is slightly more general in that it assumes neither monotonicity nor affiliation. These assumptions are used in the present paper to characterize posterior implementable social choice functions in terms of score voting.

2.6.2 Bayesian implementation: Li, Rosen, and Suen (2001)

Li, Rosen, and Suen (2001) (LRS) study a specific collective decision problem—the "Condorcet jury model"—which is a special case of the collective decision problems considered in the present paper.

In the Condorcet jury model, there are two states of the world. All agents want to accept the alternative in one state and reject it in the other, but the state is unknown. The agents receive conditionally i.i.d. signals about the state (their types). There is preference interdependence because an agent wants to learn as many signals about the state as possible, but there is also preference heterogeneity because the personal losses associated with false acceptance and rejection vary across agents. The Condorcet jury model can be mapped into the one presented in Section 2.2; the key property of the resulting environment is that the agents' indifference curves (surfaces) do not intersect.

An important result of LRS is the characterization of non-decreasing and deterministic social choice functions that are Bayesian implementable in the Condorcet jury model with $n = 2$ agents: any such social choice function must be a "partition outcome" in the sense of being characterized by a decreasing step function, just as in Green and Laffont (1987). LRS interpret and study these partition outcomes as equilibrium outcomes of voting mechanisms.

This result and interpretation are an immediate corollary from two results in the present paper: (1) Lemma 2.1 shows that in the Condorcet jury model, every posterior implementable social choice function is posterior implementable by score voting; (2) Theorem 2.3 shows that for $n = 2$ agents, monotone Bayesian implementation and posterior implementation are equivalent. The result of LRS can then be formulated as follows:

Corollary 2.2. *In the Condorcet jury model with $n = 2$ agents (Li, Rosen, and Suen, 2001), a non-decreasing and deterministic social choice function is Bayesian implementable if and only if it is Bayesian implementable by score voting.*

The results in this paper point toward the difficulty of generalizing the above characterization to the many-agent case. First, the two-agent case is special: the equivalence of monotone Bayesian implementation and posterior implementation breaks down for $n \geq 3$ agents. Second, Example 2.1 in Section 2.C shows that even efficient Bayesian implementation is sometimes possible for $n \geq 3$ agents when operating outside the Condorcet jury model, and a priori, there is no good reason why a richer set of social choice functions should not also become implementable within the Condorcet jury model.

2.7 Conclusion

Posterior implementation is stronger than Bayesian implementation but weaker than ex post implementation. This paper settles the question of posterior implementability for binary collective decision problems where preferences are interdependent and heterogeneous and where transfers are unavailable. The key message is that posterior implementation becomes very stringent as one moves beyond the two-agent setting of Green and Laffont (1987). One application is that the equivalence result for simultaneous and sequential majority voting in symmetric environments by Dekel and Piccione (2000) is a knife-edge case—the voting order typically matters, where unanimity voting is the unique exception. Another application is the characterization of monotone Bayesian implementation for $n = 2$ agents in terms of score voting mechanisms, extending an earlier characterization by Li, Rosen, and Suen (2001) beyond the "Condorcet jury model".

There are several directions for future research:

(1) An important open question is whether the characterization of monotone Bayesian implementation in terms of score voting mechanisms extends to an arbitrary finite number of agents. This question is important because a suitable generalization would make a strong positive and normative case for the prevalence of voting mechanisms as modes of information exchange and collective decision-making in the real world.

(2) Jehiel et al. (2006) establish the generic impossibility of ex post implementation with transfers in collective choice problems with multi-dimensional types; Jehiel et al. (2007) give a two-agent example showing that non-trivial posterior implementation can still be possible. Does the equivalence of monotone Bayesian implementation and posterior implementation continue to hold in the two-agent case with transfers? Are three agents enough to significantly constrain posterior implementation when types are multi-dimensional?

(3) Ex post implementation is often possible in assignment and matching problems, but efficient ex post implementation may not be feasible (see Che, Kim, and Kojima, 2015). How does posterior implementation fare against ex-post implementation in such settings? It would also be interesting to weaken the concept of posterior implementation in this context by having the mechanism disclose only (parts of) the assignment or matching rather than the entire message profile.

Appendix 2.A Posterior Beliefs and Affiliation

2.A.1 Posterior beliefs

Every strategy profile σ induces a distribution $\lambda \in \Delta(M)$ over messages $m \in M$ as follows: for every measurable $\tilde{M} \subset M$, let

$$\lambda(\tilde{M}) = \int_{\Theta} \sigma(\theta)[\tilde{M}]\,\mu(d\theta). \tag{2.A.1}$$

Analogously, define conditional distributions $\lambda(\cdot|\theta_i)$ over $\Delta(M_{-i})$. These conditional distributions are mutually absolutely continuous because the prior density f is strictly positive. Hence, one can refer to almost every message $m_{-i} \in M_{-i}$ without specifying a particular conditional distribution.

The posterior beliefs $\{\mu(\cdot|\theta_i, m_{-i})\}_{\theta_i \in \Theta_i, m_{-i} \in M_{-i}}$ are *consistent* if (1) for every $\theta_i \in \Theta_i$, they define a Markov kernel $M_{-i} \to \Delta(\Theta_{-i})$, and (2) for every $\theta_i \in \Theta_i$, every measurable $\tilde{\Theta}_{-i} \subset \Theta_{-i}$, and every measurable $\tilde{M}_{-i} \subset M_{-i}$,

$$\int_{\tilde{\Theta}_{-i}} \sigma_{-i}(\theta_{-i})[\tilde{M}_{-i}]\,\mu(d\theta_{-i}|\theta_i) = \int_{\tilde{M}_{-i}} \mu(\tilde{\Theta}_{-i}|\theta_i, m_{-i})\,\lambda(dm_{-i}|\theta_i). \tag{2.A.2}$$

Consistent beliefs are uniquely determined for almost all θ_i and m_{-i}. Existence is guaranteed by Theorem 1 in Nogales (2013).

2.A.2 Affiliation

It is well-known that affiliated distributions preserve monotonicity under conditional expectations as long as one conditions on well-behaved sets, e.g., rectangles or general sublattices of positive measure (see Milgrom and Weber, 1982, Theorem 23). However, such standard results are not sufficient for the purposes of this paper since agents might have to condition on arbitrarily complicated families of null sets when "inverting" the strategies of other agents to form their posterior beliefs $\mu(\cdot|\theta_i, m_{-i})$.[19] This should be considered a purely technical issue, which can be assumed away by postulating that agents only play strategies for which posterior beliefs can be derived via Bayes' rule.[20]

19. The more abstract definition of affiliation in the appendix of Milgrom and Weber (1982)—association conditional on any positive measure sublattice—does not circumvent the problem.

20. An alternative approach to dealing with the undetermined beliefs is to require posterior optimality in the definition of posterior equilibrium only for *almost* all $\theta_i \in \Theta_i$ and $m_{-i} \in M_{-i}$; this is highly inconvenient because one has to wrangle with null sets throughout the proofs. Green and Laffont (1987) seem to follow this approach at first glance, yet they implicitly use that posterior optimality always holds and thereby implicitly make Assumption 2.1 (see the proof of their Lemma 2).

A reasonable approach might be to first consider sets $M'_{-i} \subset M_{-i}$ of positive measure where affiliation does imply the monotonicity of posterior expected valuations

$$V_i(\theta_i|M'_{-i}) = \int_{\Theta_{-i}} v_i(\theta_i, \theta_{-i}) \, \mu(d\theta_{-i}|\theta_i, M'_{-i}) = \int_{M'_{-i}} V_i(\theta_i|m_{-i}) \lambda(dm_{-i}|\theta_i, M'_{-i}) \tag{2.A.3}$$

and to then pass down this monotonicity from sets of messages M'_{-i} of positive measure to null messages $m_{-i} \in M'_{-i}$. This approach works to the extent that one can guarantee $V_i(\theta_i|m_{-i})$ to be single-crossing in θ_i. An intuitive trap is that one might need "non-rectangular" sets M'_{-i}, i.e., $M'_{-i} \neq \prod_{j\neq i} M'_j$ for all $M'_j \subset M_j$, for this approximation approach to work, and for such sets, affiliation cannot guarantee posterior expected valuations to be monotone.

A function $f : \mathbb{R}^d \to \mathbb{R}$ is said to be *strictly single-crossing* (from below) if $f(x) \geq 0$ implies $f(x') > 0$ for all $x' > x$.

Lemma 2.4. *There exist consistent posterior beliefs* $\{\mu(\cdot|\theta_i, m_{-i})\}_{\theta_i \in \Theta_i, m_{-i} \in M_{-i}}$ *such that* $V_i(\theta_i|m_{-i})$ *is strictly single-crossing in* θ_i *for every* $m_{-i} \in M_{-i}$.

Assumption 2.1. Agents' posterior beliefs are as in Lemma 2.4.

Assumption 2.1 is always satisfied whenever posterior beliefs can be derived via Bayes' rule.

The proof of Lemma 2.4 uses the following technical result.

Lemma 2.5. *Let* $(X = \prod_{i=1}^n X_i, \mathscr{A} = \otimes_{i=1}^n \mathscr{A}_i)$ *be a product of measurable spaces* $(X_i, \mathscr{A}_i)$*, and let* μ *be a probability measure on* $\mathscr{A}$*. Then, for every* $Y \in \mathscr{A}$ *of positive measure and every* $\varepsilon > 0$ *there exists a measurable rectangle* R*, i.e.,* $R = \prod_{i=1}^n R_i$ *for* $R_i \in \mathscr{A}_i$*, such that*

$$\frac{\mu(R \setminus Y)}{\mu(R \cap Y)} < \varepsilon. \tag{2.A.4}$$

To see why this result might be useful, note that a set of positive measure need not contain a measurable rectangle of positive measure, e.g., imagine a plane with Lebesgue measure and consider the product of two fat Cantor sets rotated by 45 degrees.

Proof of Lemma 2.5. Let $\mathscr{M}$ denote the set of $Y \in \mathscr{A}$ such that for every $\varepsilon > 0$ there exist finitely many disjoint measurable rectangles $R^j \in \mathscr{A}$ such that

$$\mu(Y \Delta \bigcup_j R^j) < \varepsilon, \tag{2.A.5}$$

where Δ denotes symmetric difference. The following argument shows that $\mathscr{M}$ is a monotone class.

Suppose $Y^1 \subset Y^2 \subset \dots$ is an increasing sequence of sets in $\mathscr{M}$. Let $Y^* = \bigcup_{j\in\mathbb{N}} Y^j$. Fix any $\varepsilon > 0$ and some auxiliary $\delta < \varepsilon$. By the σ-continuity of μ, there exists $k \in \mathbb{N}$ such that

$$\mu(Y^* \setminus Y^k) \ < \ \varepsilon - \delta. \tag{2.A.6}$$

Since $Y^k \in \mathscr{M}$, there exist finitely many disjoint measurable rectangles $R^j \in \mathscr{A}$ such that

$$\mu(Y^k \Delta \bigcup_j R^j) \ < \ \delta. \tag{2.A.7}$$

Note that $Y^* \Delta \bigcup_j R^j \subset (Y^k \Delta \bigcup_j R^j) \cup (Y^* \setminus Y^k)$. Hence, $Y^* \in \mathscr{M}$ because

$$\mu(Y^* \Delta \bigcup_j R^j) \ \leq \ \mu(Y^k \Delta \bigcup_j R^j) + \mu(Y^* \setminus Y^k) \ < \ \varepsilon. \tag{2.A.8}$$

By similar arguments, if $Y^1 \supset Y^2 \supset \dots$ is a decreasing sequence of sets in $\mathscr{M}$, then $Y^* = \bigcap_{j\in\mathbb{N}} Y^j \in \mathscr{M}$. Hence, $\mathscr{M}$ is a monotone class.

Let $\mathscr{R}$ denote the set of all finite unions of disjoint measurable rectangles. Clearly, $\mathscr{R} \subset \mathscr{M}$, and it follows from the definition of $\mathscr{A}$ that $\mathscr{A}$ is generated by $\mathscr{R}$. Moreover, it can be shown that $\mathscr{R}$ is an algebra. By the monotone class theorem, $\mathscr{M} = \mathscr{A}$.

Fix any $\varepsilon > 0$ and any $Y \in \mathscr{A}$ of positive measure. Pick finitely many disjoint measurable rectangles $R^j \in \mathscr{A}$ such that

$$\mu(Y \Delta \bigcup_j R^j) \ < \ \frac{\varepsilon}{1-\varepsilon}\mu(Y). \tag{2.A.9}$$

Finally, there must exist a measurable rectangle R^i such that

$$\begin{aligned} \frac{\mu(R^i \setminus Y)}{\mu(R^i \cap Y)} \leq \frac{\sum_j \mu(R^j \setminus Y)}{\sum_j \mu(R^j \cap Y)} = \frac{\mu(\bigcup_j R^j \setminus Y)}{\mu(\bigcup_j R^j \cap Y)} \\ = \frac{\mu(\bigcup_j R^j \setminus Y)}{\mu(Y) - \mu(Y \setminus \bigcup_j R^j)} < \frac{\mu(Y \Delta \bigcup_j R^j)}{\mu(Y) - \mu(Y \Delta \bigcup_j R^j)} < \varepsilon, \end{aligned} \tag{2.A.10}$$

which completes the proof. □

With the previous lemma, one can show:

Lemma 2.6. *If $M'_{-i} \subset M_{-i}$ is of positive measure, then there exists $M''_{-i} \subset M'_{-i}$ of positive measure such that $V_i(\theta_i|M''_{-i})$ is strictly increasing in θ_i.*

Proof. First note that for any measurable rectangle $R_{-i} = \prod_{j\neq i} R_j \subset M_{-i}$ of positive measure, the density

$$f(\theta|R_{-i}) \ = \ \frac{f(\theta)\prod_{j\neq i}\sigma_j(\theta_j)[R_j]}{\int_{\Theta_{-i}}\prod_{j\neq i}\sigma_j(\theta_j)[R_j]d\mu(\theta_{-i})} \tag{2.A.11}$$

of $\mu(\cdot|R_{-i})$ is affiliated (everything but f cancels in the affiliation inequality).

It is a standard result that $V_i(\theta_i|R_{-i})$ is strictly increasing in θ_i (see Milgrom and Weber, 1982, Theorem 23). Using the same proof technique,[21] one can obtain the following stronger conclusion: for $\theta_i' > \theta_i$,

$$V_i(\theta_i'|R_{-i}) - V_i(\theta_i|R_{-i}) \;\geq\; \inf_{\theta_{-i}\in\Theta_{-i}} \{v_i(\theta_i', \theta_{-i}) - v_i(\theta_i, \theta_{-i})\}. \tag{2.A.12}$$

By the compactness of Θ_{-i}, $V_i(\theta_i'|R_{-i}) - V_i(\theta_i|R_{-i})$ is hence uniformly bounded away from 0 across all measurable rectangles of positive measure.

Note that there exists a bound $b > 0$ such that for all $\theta_i, \theta_i' \in \Theta_i$ and $M_{-i}' \subset M_{-i}$ of positive measure, $1/b < \mu(M_{-i}'|\theta_i')/\mu(M_{-i}'|\theta_i) < b$ because the density f of μ is strictly positive on a compact domain. This observation together with Lemma 2.5 implies that for every $M_{-i}' \subset M_{-i}$ of positive measure and every $\varepsilon > 0$, there exists a measurable rectangle $R_{-i} \subset M_{-i}$ such that for all $\theta_i \in \Theta_i$,

$$\frac{\mu(R_{-i} \setminus M_{-i}'|\theta_i)}{\mu(R_{-i} \cap M_{-i}'|\theta_i)} \;<\; \varepsilon. \tag{2.A.13}$$

By choosing an appropriate R_{-i} and defining $M_{-i}'' = M_{-i}' \cap R_{-i}$, the claim now follows from (2.A.12), (2.A.13), the boundedness of V_i, and the fact that

$$\begin{aligned} V_i(\theta_i|R_{-i} \cap M_{-i}') = V_i(\theta_i|R_{-i}) &+ \frac{\mu(R_{-i} \setminus M_{-i}'|\theta_i)}{\mu(R_{-i} \cap M_{-i}'|\theta_i)} V_i(\theta_i|R_{-i}) \\ &- \frac{\mu(R_{-i} \setminus M_{-i}'|\theta_i)}{\mu(R_{-i} \cap M_{-i}'|\theta_i)} V_i(\theta_i|R_{-i} \setminus M_{-i}'). \end{aligned} \tag{2.A.14}$$

□

Proof of Lemma 2.4. Fix arbitrary consistent posterior beliefs $\{\mu(\cdot|\theta_i, m_{-i})\}_{\theta_i\in\Theta_i, m_{-i}\in M_{-i}}$. In the proof, these beliefs are modified on a null set of messages to obtain the desired monotonicity property.

For each $m_{-i} \in M_{-i}$, define

$$\bar{\theta}_i(m_{-i}) \;=\; \inf\big\{\theta_i \;\in\; \Theta_i \;\big|\; \mu_i(\{\tilde{\theta}_i \;>\; \theta_i \;|\; V_i(\tilde{\theta}_i|m_{-i}) \;\leq\; 0\}) \;=\; 0\big\} \tag{2.A.15}$$

to be the lowest type for which almost all higher types prefer to accept the alternative given m_{-i}. Also define the sets

$$C^{\downarrow}(\theta_i) \;=\; \left\{ m_{-i} \;\in\; M_{-i} \;\middle|\; \begin{array}{l} \theta_i \;<\; \bar{\theta}_i(m_{-i}) \\ V_i(\theta_i|m_{-i}) \;\geq\; 0 \end{array} \right\} \tag{2.A.16}$$

$$C^{\uparrow}(\theta_i) \;=\; \left\{ m_{-i} \;\in\; M_{-i} \;\middle|\; \begin{array}{l} \theta_i \;>\; \bar{\theta}_i(m_{-i}) \\ V_i(\theta_i|m_{-i}) \;\leq\; 0 \end{array} \right\} \tag{2.A.17}$$

21. One writes $V_i(\cdot|R_{-i})$ as an $(n-1)$-fold iterated expectation and then iteratively exploits the fact that given two affiliated random variables, the conditional distributions of one variable given the realizations of the other variable are ordered in the sense of first order stochastic dominance.

in order to represent the type-message combinations (θ_i, m_{-i}) where strict single-crossing of $V_i(\theta_i|m_{-i})$ is violated.

The next step is to establish that, for every $\theta_i \in \Theta_i$, $C^{\downarrow}(\theta_i)$ and $C^{\uparrow}(\theta_i)$ are null sets. Indeed, suppose there exists a type $\theta_i \in \Theta_i$ such that $C^{\downarrow}(\theta_i)$ has positive measure. By the definition of $\bar{\theta}_i(m_{-i})$, for every $m_{-i} \in C^{\downarrow}(\theta_i)$ there exists a positive measure of types $\tilde{\theta}_i > \theta_i$ such that $V_i(\tilde{\theta}_i|m_{-i}) \leq 0$. By Tonelli's theorem, there must exist a type $\tilde{\theta}_i > \theta_i$ such that $V_i(\tilde{\theta}_i|m_{-i}) \leq 0$ for a positive measure of $m_{-i} \in C^{\downarrow}(\theta_i)$. This contradicts Lemma 2.6. By reversing all the previous inequalities, the same order of arguments yields a contradiction if there exists a type $\theta_i \in \Theta_i$ such that $C^{\uparrow}(\theta_i)$ has positive measure.

Now, the proof can be completed as follows. For every $\theta_i \in \Theta_i$ and $m_{-i} \in C^{\downarrow}(\theta_i)$, replace the belief $\mu(\cdot|\theta_i, m_{-i})$ by a belief $\tilde{\mu}(\cdot|\theta_i, m_{-i}) \in \Delta(\Theta_{-i})$ such that $V_i(\theta_i|m_{-i}) < 0$; this is always possible because if $\theta_i < \bar{\theta}_i(m_{-i})$, then there exists $\theta_{-i} \in \Theta_{-i}$ such that $v_i(\theta_i, \theta_{-i}) < 0$. Analogously for every $\theta_i \in \Theta_i$ and $m_{-i} \in C^{\uparrow}(\theta_i)$, replace the belief $\mu(\cdot|\theta_i, m_{-i})$ by a belief $\tilde{\mu}(\cdot|\theta_i, m_{-i}) \in \Delta(\Theta_{-i})$ such that $V_i(\theta_i|m_{-i}) > 0$. By construction, for every $m_{-i} \in M_{-i}$, $V_i(\theta_i|m_{-i})$ is then strictly single-crossing at $\bar{\theta}_i(m_{-i})$. Finally, the resulting system of beliefs is consistent because $C^{\uparrow}(\theta_i)$ and $C^{\downarrow}(\theta_i)$ are null sets. □

Appendix 2.B The Concepts of Bayesian and Ex Post Implementation

Definition 2.3. A strategy profile $\sigma = (\sigma_i)_{i \in N}$ in mechanism (M, ψ) is a *Bayes-Nash equilibrium* if for all $i \in N$, $\theta_i \in \Theta_i$, and $\tilde{m}_i \in M_i$

$$\int_{\Theta_{-i}} v_i(\theta_i, \theta_{-i}) \psi(\sigma_i(\theta_i), \sigma_{-i}(\theta_{-i}))\, \mu(d\theta_{-i}|\theta_i) \\ \geq \int_{\Theta_{-i}} v_i(\theta_i, \theta_{-i}) \psi(\tilde{m}_i, \sigma_{-i}(\theta_{-i}))\, \mu(d\theta_{-i}|\theta_i). \quad (2.B.1)$$

Definition 2.4. A strategy profile $\sigma = (\sigma_i)_{i \in N}$ in mechanism (M, ψ) is an *ex post equilibrium* if for all $i \in N$, $\theta \in \Theta$, and $\tilde{m}_i \in M_i$

$$v_i(\theta_i, \theta_{-i}) \psi(\sigma_i(\theta_i), \sigma_{-i}(\theta_{-i})) \;\geq\; v_i(\theta_i, \theta_{-i}) \psi(\tilde{m}_i, \sigma_{-i}(\theta_{-i})). \quad (2.B.2)$$

The definitions of Bayesian and ex post implementation are analogous to the definition of posterior implementation. The revelation principle applies to both concepts: a social choice function ϕ is Bayesian or ex post implementable if and only if it is Bayesian or ex post incentive compatible, i.e., implementable via the direct implementation $(\Theta, \phi, \mathrm{id}_\Theta)$.

Remark 2.7 (Ex post regret). Note that the no-regret property of ex post equilibrium is required to hold conditional on the true state θ but not conditional on every possible realization of messages in the lottery $\sigma(\theta)$. (This definition is consistent with the literature on robust implementation; see Bergemann and Morris, 2005). Thus, strictly speaking, mixed-strategy ex post equilibrium does not imply mixed-strategy posterior equilibrium. However, one can define a stronger notion of ex post equilibrium that, like posterior equilibrium, maintains the no-regret property conditional on every possible realization of messages. (Only this definition would be consistent with robustness of equilibrium behavior against the extensive form of the mechanism.) Clearly, if agents play pure strategies, then the two notions of ex post equilibrium coincide. It is then a consequence of the revelation principle that the two notions of ex post implementation derived from the two different notions of ex post equilibrium are equivalent (because truthful strategies in direct revelation mechanisms are pure). Consequently, ex post implementation does imply posterior implementation.

Appendix 2.C (Counter-)Examples

This section contains two examples. The first example shows that efficient social choice functions can be Bayesian implementable when there are $n \geq 3$ agents; recall Section 2.4.2. The second example shows that even in a symmetric environment, a Bayes-Nash equilibrium of simple majority with abstention need not be a posterior equilibrium; recall Section 2.4.4.

Example 2.1. Let $n = 3$. For all $i \in N$, let $\Theta_i = [-1, 1]$ and

$$v_i(\theta) = \tfrac{1}{2}(\alpha_i + \beta_i)\theta_i + \alpha_i \theta_{(i \bmod 3)+1} + \beta_i \theta_{(i+1 \bmod 3)+1},$$

where $\alpha_i, \beta_i > 0$. Assume a uniform prior over $\Theta = [-1, 1]^3$. Consider the social choice function

$$\phi(\theta) = \begin{cases} 1 & \text{if } \sum_{i \in N} \theta_i > 0; \\ 0 & \text{else.} \end{cases}$$

It can be verified that ϕ is Bayesian implementable; see below. Moreover, ϕ is efficient if $\alpha_i = \alpha$ and $\beta_i = \beta$ for all $i \in N$. (These particular environments should be considered non-generic because the agents' indifference surfaces are "rotationally symmetric" around the acceptance frontier of the efficient social choice function.)

By the revelation principle, ϕ is Bayesian implementable if ϕ is Bayesian incentive compatible. By the symmetry of the example, it suffices to consider agent 1. Bayesian incentive compatibility requires that for all $\theta_1 \in [-1, 1]$, θ_1 maximizes

$$\int_{-1}^{1} \int_{-1}^{1} v_1(\theta_1, \theta_2, \theta_3) \phi(\theta_1', \theta_2, \theta_3) \, d\theta_3 \, d\theta_2 \qquad (2.C.1)$$

over $\theta_1' \in [-1,1]$. If $\theta_1' \leq 0$, then (2.C.1) can be written as

$$\int_{-1-\theta_1'}^{1} \int_{-\theta_1'-\theta_2}^{1} v_1(\theta_1,\theta_2,\theta_3)\,d\theta_3\,d\theta_2. \tag{2.C.2}$$

If $\theta_1' \geq 0$, then (2.C.1) can be written as

$$\int_{-1}^{1-\theta_1'} \int_{-\theta_1'-\theta_2}^{1} v_1(\theta_1,\theta_2,\theta_3)\,d\theta_3\,d\theta_2 + \int_{1-\theta_1'}^{1} \int_{-1}^{1} v_1(\theta_1,\theta_2,\theta_3)\,d\theta_3\,d\theta_2. \tag{2.C.3}$$

The derivative of (2.C.2) with respect to θ_1' is

$$(\alpha_1+\beta_1)\left(-\tfrac{1}{2}(\theta_1')^2 + \theta_1'(-1+\tfrac{1}{2}\theta_1) + \theta_1\right). \tag{2.C.4}$$

The derivative of (2.C.3) with respect to θ_1' is

$$(\alpha_1+\beta_1)\left(\tfrac{1}{2}(\theta_1')^2 + \theta_1'(-1-\tfrac{1}{2}\theta_1) + \theta_1\right). \tag{2.C.5}$$

Both derivatives, as a function of θ_1', have θ_1 as a root and an additional root outside $[-1,1]$. In either case, the second derivative is negative at $\theta_1' = \theta_1$. Thus, ϕ is Bayesian incentive compatible.

Example 2.2. Consider the following symmetric environment. Let $n = 3$. For all $i \in N$, let $\Theta_i = [-1,1]$ and

$$v_i(\theta) = \theta_i + \sum_{j \neq i} \tfrac{1}{2}\theta_j.$$

Assume a uniform prior over $\Theta = [-1,1]^3$. It can be verified that the symmetric strategies

$$\sigma_i(\theta_i) = \begin{cases} -1 & \text{if } \theta_i \leq \theta^* \\ 0 & \text{if } \theta^* < \theta_i < \theta^{**} \\ 1 & \text{if } \theta_i \geq \theta^{**} \end{cases}$$

where

$\theta^* \approx -0.2181$ is the unique real solution to $289x^3 + 249x^2 + 91x + 11 = 0$ and $\theta^{**} = -\frac{5}{2}\theta^* - \frac{1}{2} \approx 0.0451$

constitute a symmetric Bayes-Nash equilibrium but not a posterior equilibrium of simple majority with abstention; see below.

By the symmetry of the environment, the independence of signals and the monotonicity of valuation functions, the Bayes-Nash equilibrium conditions (2.B.1) require that (1) cutoff type θ^* is indifferent between a negative vote and abstention conditional on the pivotal event that one of the other agents abstains and the other votes positively, and (2) cutoff type θ^* is indifferent between abstention and a positive vote conditional on the pivotal event that (a) both of the other agents abstain OR (b) one of the other agents votes positively and the other votes negatively; formally:

$$\int_{\theta^*}^{1}\int_{\theta^*}^{\theta^{**}} \theta^* + \tfrac{1}{2}\theta_2 + \tfrac{1}{2}\theta_3 \, d\theta_2 \, d\theta_3 \;=\; 0 \tag{2.C.6}$$

$$\int_{\theta^*}^{\theta^{**}}\int_{\theta^*}^{\theta^{**}} \theta^* + \tfrac{1}{2}\theta_2 + \tfrac{1}{2}\theta_3 \, d\theta_2 \, d\theta_3 + 2\int_{\theta^{**}}^{1}\int_{-1}^{\theta^*} \theta^* + \tfrac{1}{2}\theta_2 + \tfrac{1}{2}\theta_3 \, d\theta_2 \, d\theta_3 \;=\; 0. \tag{2.C.7}$$

Tedious calculations show that the system of equations is solved by the cutoffs given in the example.

To show that σ is not a posterior equilibrium, suppose that $\theta_1 = 0.05$ and $m_{-1} = (0,0)$. Under σ_i, agent 1 should submit a positive vote to accept the alternative but she prefers to reject the alternative by abstaining or submitting a negative vote since

$$V_1(\theta_1 \;=\; 0.05|m_{-1} \;=\; (0,0)) \;=\; 0.05 + \frac{\theta^{**} - \theta^*}{2} \;\approx\; -0.0365 \;<\; 0. \tag{2.C.8}$$

Appendix 2.D Proofs for Section 2.3

Almost the entire section is devoted to the proof of Theorem 2.1 and is organized according to the steps that are needed to establish the result. Lemma 2.1 will be an (almost) immediate corollary along the way; see Section 2.D.3. Lemma 2.2 and Lemma 2.14 (the pivotal voting argument for majority voting) are shown in Section 2.D.4.

The proof of Theorem 2.1 proceeds in four steps: (1) if types pool on the same message, then they form an interval; (2) pooling occurs for all types; (3) there are finitely many pools; (4) such pooling can be achieved via score voting.

2.D.1 Convexity

A tuple (M, ψ, σ) is *convex* if (1) σ is pure, (2) for all $i \in N$ and $m_i \in M_i$, $\sigma_i^{-1}(m_i)$ is non-empty and convex, and (3) for all $m_i, m_i' \in M_i$, there exists $m_{-i} \in M_{-i}$ such that $\psi(m_i, m_{-i}) \neq \psi(m_i', m_{-i})$.

Lemma 2.7. *A social choice function ϕ is posterior implementable if and only if there exists a convex posterior implementation of ϕ.*

The following definition as in Green and Laffont (1987) (GL) will be useful. For given (M, ψ, σ), let

$$T_i(m_i) \;=\; \left\{\theta_i \;\in\; \Theta_i \;\middle|\; \forall m_{-i} \;\in\; M_{-i} \;:\; m_i \;\in\; \arg\max_{\tilde{m}_i \in M_i} V_i(\theta_i|m_{-i})\psi(\tilde{m}_i, m_{-i})\right\} \tag{2.D.1}$$

denote the set of types θ_i for which message m_i is posterior optimal. By the definition of posterior equilibrium, each type θ_i must mix over posterior optimal messages, i.e., $\theta_i \in T_i(m_i)$ for $\sigma_i(\theta_i)$-almost every $m_i \in M_i$.

Proof of Lemma 2.7. Let (M, ψ, σ) be a posterior implementation of ϕ. Let $m_i \sim_i m_i'$ if and only if $\psi(m_i, m_{-i}) = \psi(m_i', m_{-i})$ for all $m_{-i} \in M_{-i}$. Let $[m_i]$ denote the equivalence class of m_i. Clearly, if $m_i \sim_i m_i'$, then $T_i(m_i) = T_i(m_i') := T_i([m_i])$.

Step 1. *For all $i \in N$ and $m_i \in M_i$, $T_i(m_i)$ is convex and for $\tilde{m}_i \not\sim_i m_i$, $T_i(m_i)$ and $T_i(\tilde{m}_i)$ intersect at most in their boundaries.*

For the sake of contradiction, suppose $\theta_i, \theta_i'' \in T_i(m_i)$ and $\theta_i' \in T_i(\tilde{m}_i)$, where $\theta_i < \theta_i' < \theta_i''$ and $m_i \not\sim_i \tilde{m}_i$. By the definition of $\sim_i$, there is some $m_{-i} \in M_{-i}$ such that $\psi(m_i, m_{-i}) \neq \psi(\tilde{m}_i, m_{-i})$, and, without loss of generality, suppose $\psi(m_i, m_{-i}) < \psi(\tilde{m}_i, m_{-i})$. The posterior equilibrium conditions for σ_i require that $V_i(\theta_i|m_{-i}) \leq 0$, $V_i(\theta_i'|m_{-i}) \geq 0$, and $V_i(\theta_i''|m_{-i}) \leq 0$, contradicting Assumption 2.1.

Step 2. *Construct a posterior implementation (M^*, ψ^*, σ^*) of ϕ by identifying equivalent messages.*

Let $M_i^* = M_i/\sim_i$ be the set of equivalence classes under $\sim_i$. Let $\psi^* : M^* \to [0,1]$ be such that $\psi^*([m_1], \ldots, [m_n]) = \psi(m_1, \ldots, m_n)$ for all $m \in M$. For every measurable $M_i^{*\prime} \subset M_i^*$, let $\sigma_i^*(\theta_i)[M_i^{*\prime}] = \sigma_i(\theta_i)[\{m_i \in M_i \mid [m_i] \in M_i^{*\prime}\}]$.[22] By construction, $\phi = \psi^* \circ \sigma^*$.

It remains to show that σ^* is a posterior equilibrium in (M^*, ψ^*). Fix any arbitrary $i \in N$, $\theta_i \in \Theta_i$, $m_{-i}^* \in M_{-i}^*$, and $\tilde{m}_i^* \in M_i^*$. The posterior equilibrium conditions for strategy profile σ require that for all $m_{-i} \in m_{-i}^*$ and $\tilde{m}_i \in M_i$,

$$V_i(\theta_i|m_{-i})\psi(\sigma_i(\theta_i), m_{-i}) \;\geq\; V_i(\theta_i|m_{-i})\psi(\tilde{m}_i, m_{-i}). \tag{2.D.2}$$

By construction, for all $m_{-i} \in m_{-i}^*$,

$$V_i(\theta_i|m_{-i})\psi^*(\sigma_i^*(\theta_i), m_{-i}^*) \;\geq\; V_i(\theta_i|m_{-i})\psi^*(\tilde{m}_i^*, m_{-i}^*). \tag{2.D.3}$$

Observe that by the first step, each message m_j^* either perfectly reveals the type θ_j or reveals that the type θ_j lies in some interval. Thus, the conditional distribution $\lambda(\cdot|\theta_i, m_{-i}^*) \in \Delta(m_{-i}^*)$ can be straightforwardly computed (cf. (2.A.1)). Integration yields

$$\int_{m_{-i}^*} V_i(\theta_i|m_{-i})\lambda(dm_{-i}|\theta_i, m_{-i}^*)\,\psi^*(\sigma_i^*(\theta_i), m_{-i}^*)$$
$$\geq \int_{m_{-i}^*} V_i(\theta_i|m_{-i})\lambda(dm_{-i}|\theta_i, m_{-i}^*)\,\psi^*(\tilde{m}_i^*, m_{-i}^*). \tag{2.D.4}$$

22. Endow each M_i^* with the σ-algebra where $M_i^{*\prime} \subset M_i^*$ is measurable if and only if $\cup M_i^{*\prime}$ is a measurable subset of M_i. Then ψ^* is measurable and each σ_i^* is a Markov kernel.

But $\int_{m^*_{-i}} V_i(\theta_i|m_{-i})\,\lambda(dm_{-i}|\theta_i, m^*_{-i}) = V_i(\theta_i|m^*_{-i})$ (cf. (2.A.3)), which concludes the second step.

Step 3. *Construct a convex posterior implementation $(M^*, \psi^\star, \sigma^\star)$ of ϕ by transferring randomization from the agents' strategies into the outcome function.*

Initialize $\sigma^\star = \sigma^*$ and $\psi^\star = \psi^*$. For each agent $i \in N$ and $\theta_i \in \Theta_i$, make adjustments to $\sigma^\star_i(\theta_i)$ and $\psi^\star$ as follows.

By the previous steps, if $\sigma^*_i(\theta_i)$ is non-degenerate, then $\theta_i \in \operatorname{bndr} T_i(m^*_i)$ for some $m^*_i \in M^*_i$. There are two cases: (1) $\theta_i \in T_i(\tilde{m}^*_i)$ for some $\tilde{m}^*_i \in M^*_i$ such that $T_i(\tilde{m}^*_i)$ is non-degenerate; (2) $\{\theta_i\} = T_i(m^*_i)$ for every $m^*_i \in M^*_i$ such that $\theta_i \in T_i(m^*_i)$.

In the first case, let $\sigma^\star_i(\theta_i) = \tilde{m}^*_i$. With this adjustment, $\sigma^\star$ remains a posterior equilibrium because $\theta_i \in T_i(\tilde{m}^*_i)$ and for all other agents, the posterior beliefs given σ^*_i or $\sigma^\star_i$ after observing $\tilde{m}^*_i$ are essentially the same since $T_i(\tilde{m}^*_i)$ is non-degenerate.

In the second case, for every message $m^*_i \in M^*_i$ such that $\{\theta_i\} = T_i(m^*_i)$ and every $m^*_{-i} \in M^*_{-i}$, let $\psi^\star(m^*_i, m^*_{-i}) = \psi^*(\sigma^*_i(\theta_i), m^*_{-i})$. Then, let $\sigma^\star_i(\theta_i) = m^*_i$ for some arbitrary $m^*_i \in M^*_i$ such that $\{\theta_i\} = T_i(m^*_i)$. With this adjustment, $\sigma^\star_i$ remains posterior optimal for agent i because type θ_i receives the same outcome as before and no type $\theta'_i \neq \theta_i$ will find it profitable to deviate to m^*_i (because the acceptance probabilities for m^*_i are less extreme than before). Moreover, for all other agents, m^*_i still perfectly reveals θ_i, and the fact that their messages were posterior optimal against every message over which agent i mixed in $\sigma^*_i(\theta_i)$ implies that their messages are still optimal against the lottery $\sigma^*_i(\theta_i)$; thus against m^*_i given $\psi^\star$.

For the resulting tuple $(M^*, \psi^\star, \sigma^\star)$, $\sigma^\star$ remains a posterior equilibrium in $(M^*, \psi^\star)$ and $(\psi^\star \circ \sigma^\star)(\theta) = \phi(\theta)$ μ-almost everywhere because there are at most countably many types $\theta_i \in \Theta_i$ such that $T_i(\sigma^\star_i(\theta_i))$ is not a singleton and $\theta_i \in \operatorname{bndr} T_i(\sigma^\star_i(\theta_i))$. Thus, $(M^*, \psi^\star, \sigma^\star)$ is a posterior implementation of ϕ.

The posterior implementation $(\operatorname{im} \sigma^\star, \psi^\star_{|\operatorname{im} \sigma^\star}, \sigma^\star)$ of ϕ is convex by construction.

□

2.D.2 Partition equilibria

Although Lemma 2.7 shows that types which send the same message must form an interval, it is not yet clear whether these intervals are non-degenerate, i.e., whether agents garble their private information in that different types pool on the same message. The next step is to show that agents must indeed garble their private information.

A tuple (M, ψ, σ) is *partitional* if it is convex and M_i is countable for all $i \in N$. This is akin to the partition equilibria studied in Crawford and Sobel (1982).

Lemma 2.8. *A responsive social choice function ϕ is posterior implementable if and only if there exists a partitional posterior implementation of ϕ.*

The proof of Lemma 2.8 will need two auxiliary results.

Observe that if (M, ψ, σ) is convex, then each message space M_i inherits a total order from the type space Θ_i: for $m_i \neq m_i'$,

$$m_i > m_i' \iff \inf \sigma_i^{-1}(m_i) \geq \sup \sigma_i^{-1}(m_i'). \tag{2.D.5}$$

In what follows, these orders are extended to product orders on M_{-i} or M.

Lemma 2.9. *If (M, ψ, σ) is convex, then $V_i(\theta_i|m_{-i})$ is continuous and strictly increasing in θ_i for every $m_{-i} \in M_{-i}$ and strictly increasing in m_{-i} for every $\theta_i \in \Theta_i$.*

Proof. Given convexity, the monotonicity assertions are standard (see Milgrom and Weber, 1982, Theorem 5).

It remains to show continuity. Posterior expected utility reads

$$V_i(\theta_i|m_{-i}) = \frac{\int_{\sigma_{-i}^{-1}(m_{-i})} v_i(\theta_i, \theta_{-i}) f(\theta_{-i}|\theta_i)\, d\theta_{-i}}{\int_{\sigma_{-i}^{-1}(m_{-i})} f(\theta_{-i}|\theta_i)\, d\theta_{-i}}.^{23} \tag{2.D.6}$$

Both integrands in the above equation are uniformly continuous because they are (products of) continuous functions on a compact domain Θ. Thus, the integrals are continuous in θ_i. Continuity then follows because the denominator is strictly positive by assumption. □

For convex (M, ψ, σ), it will be useful to say that $\psi(\sigma_i(\cdot), m_{-i})$ *jumps* at θ_i if $\psi(\sigma_i(\cdot), m_{-i})$ is non-constant on every neighborhood of θ_i.

Lemma 2.10. *If (M, ψ, σ) is convex, then, for all $i \in N$ and $m_{-i} \in M_{-i}$, $\psi(\sigma_i(\cdot), m_{-i})$ is non-decreasing and jumps at most once. Moreover, if $\psi(\sigma_i(\cdot), m_{-i})$ jumps at θ_i, then $V_i(\theta_i|m_{-i}) = 0$.*

Proof. Recall that $V_i(\theta_i|m_{-i})$ is continuous and strictly increasing in θ_i by Lemma 2.9. The second claim follows from the continuity of V_i and the posterior equilibrium conditions. Moreover, by the monotonicity of V_i, there is at most one type $\theta_i \in \Theta_i$ such that $V_i(\theta_i|m_{-i}) = 0$. Moreover, $V_i(\theta_i'|m_{-i}) < 0$ for $\theta_i' < \theta_i$ and $V_i(\theta_i'|m_{-i}) > 0$ for $\theta_i' > \theta_i$. Thus, the second claim together with the posterior equilibrium conditions implies the first claim. □

Proof of Lemma 2.8. Let (M, ψ, σ) be a convex posterior implementation of ϕ. For the sake of contradiction, suppose that (M, ψ, σ) is not partitional. Then, there exists an agent $i \in N$ and an open interval $S_i \subset \Theta_i$ such that $\sigma_i^{-1}(\sigma_i(\theta_i)) = \{\theta_i\}$ for every $\theta_i \in S_i$.

23. If a message m_j perfectly reveals the type θ_j, then the above expression is clearly not well-defined. Instead, one would have to integrate over $\sigma_{-ij}^{-1}(m_{-ij})$ and plug θ_j into the integrand. The arguments then go through unchanged; this issue is henceforth ignored.

Pick arbitrary $\theta_i^-, \theta_i^+ \in S_i$. Since (M, ψ, σ) is convex, there must exist $m_{-i} \in M_{-i}$ such that $\psi(\sigma_i(\theta_i^-), m_{-i}) \neq \psi(\sigma_i(\theta_i^+), m_{-i})$. Then, as in Lemma 2.10, there exists a type $\theta_i^* \in [\theta_i^-, \theta_i^+]$ such that $V_i(\theta_i^* | m_{-i}) = 0$ and such that $\psi(\sigma_i(\cdot), m_{-i})$ jumps at θ_i^*.

Fix an arbitrary agent $j \neq i$. Let $\theta_j \in \sigma_j^{-1}(m_j)$. Since ϕ is responsive, there must exist a type $\theta_j' \neq \theta_j$ such that $\sigma_j(\theta_j') \neq m_j$. Let $\theta_j^*(\theta_i) \in \Theta_j$ be the—by Lemma 2.9—unique type such that $V_j(\theta_j^*(\theta_i) | \theta_i, m_{-ij}) = 0$ (if such a type exists). Also by Lemma 2.9, $\theta_j^*(\theta_i)$ is strictly decreasing in θ_i.

Consider the schematic of messages profiles

$$\begin{array}{ccc} (\sigma_i(\theta_i), \sigma_j(\theta_j'), m_{-ij}) & \longrightarrow & (\sigma_i(\theta_i'), \sigma_j(\theta_j'), m_{-ij}) \\ \uparrow & & \uparrow \\ (\sigma_i(\theta_i), \sigma_j(\theta_j), m_{-ij}) & \longrightarrow & (\sigma_i(\theta_i'), \sigma_j(\theta_j), m_{-ij}) \end{array}$$

where $\theta_i, \theta_i' \in S_i$, $\theta_i < \theta_i^* < \theta_i'$. For the sake of the argument, assume that $\theta_j < \theta_j'$; similar arguments apply to the other case.

Recall that ψ jumps along the bottom arrow, i.e., takes different values at the bottom left and bottom right message profiles. As long as θ_i and θ_i' are sufficiently close to θ_i^*, Lemma 2.9 and Lemma 2.10 imply that ψ cannot jump along the top arrow. Thus, as long as θ_i and θ_i' are sufficiently close to θ_i^*, ψ must jump along the left arrow because ψ is non-decreasing by Lemma 2.10.

If θ_i and θ_i' are sufficiently close to θ_i^* and $\theta_j^*(\theta_i^*) \notin [\theta_j, \theta_j')$, then $\theta_j^*(\theta_i), \theta_j^*(\theta_i') \notin [\theta_j, \theta_j']$ by Lemma 2.9. The implication would contradict that ψ jumps along the left arrow by Lemma 2.10; thus, $\theta_j^*(\theta_i^*) \in [\theta_j, \theta_j')$ as long as θ_i and θ_i' are sufficiently close to θ_i^*.

Suppose for all sufficiently small $\varepsilon > 0$ there exists a type $\theta_j'' \in (\theta_j^*(\theta_i^*), \theta_j')$ such that $|\theta_j'' - \theta_j^*(\theta_i^*)| < \varepsilon$ and such that $\sigma_j^{-1}(\sigma_j(\theta_j''))$ is not a singleton. Thus, by Lemma 2.9, there exist types $\theta_i < \theta_i^*$ arbitrarily close to θ_i^* such that $\theta_j^*(\theta_i) \in \operatorname{int} \sigma_j^{-1}(\sigma_j(\theta_j^*(\theta_i)))$. Consequently, $\psi(\sigma_j(\cdot), \sigma_i(\theta_i), m_{-j})$ cannot jump at $\theta_j^*(\theta_i)$, which contradicts Lemma 2.10. Thus, there must exist an open interval $S_j \subset (\theta_j^*(\theta_i^*), \theta_j')$ such that for all $\theta_j'' \in S_j$, $\sigma_j^{-1}(\sigma_j(\theta_j''))$ is a singleton and such that for some $(\theta_i, \theta_j) \in S_i \times S_j$, $\psi(\sigma_j(\cdot), \sigma_i(\theta_i), m_{-ij})$ jumps at θ_j.

Now, iterate on the previous arguments with all the remaining agents one-by-one, at each step using for the arguments the just-constructed message profile at which ψ jumps, agent j in place of agent i, and some new agent k in place of agent j. This procedure yields a type profile $\theta \in \Theta$ such that for all $j \in N$, there exists an open interval $S_j \subset \Theta_j$ such that for all $\theta_j \in S_j$, $\sigma_j^{-1}(\sigma_j(\theta_j))$ is a singleton and, moreover, there is some agent $k \in N$ such that $\psi(\sigma_k(\cdot), \sigma_{-k}(\theta_{-k}))$ jumps at θ_k.

Finally, observe that σ is an ex post equilibrium on the restricted state space $S = \prod_{j \in N} S_j$ because messages perfectly reveal the types in S. A result by Feng, Niemeyer, and Wu (2023, Theorem 1) implies that ϕ must be constant on S, which contradicts that $\psi(\sigma_k(\cdot), \sigma_{-k}(\theta_{-k}))$ jumps at θ_k. $\square$

2.D.3 Finite partitions

Any posterior equilibrium of any score voting mechanism yields a partitional implementation of some social choice function. The next step is to show that—as in score voting—a partitional posterior implementation only needs finitely many messages (if $n \geq 3$).

A tuple (M, ψ, σ) is *finite* if M is finite.

Lemma 2.11. *Let $n \geq 3$. A responsive social choice function ϕ is posterior implementable if and only if there exists a finite partitional posterior implementation of ϕ.*

Again, the following auxiliary result will be useful.

For convex (M, ψ, σ), say that two messages $m_i, m_i' \in M_i$ are *adjacent* if $\sigma_i^{-1}(m_i) \cup \sigma_i^{-1}(m_i')$ is an interval. If m_i and m_i' are adjacent and $m_i' > m_i$, abuse notation by writing $m_i' = m_i + 1$ and $m_i = m_i' - 1$. (Once Lemma 2.11 is established, one can identify agents' messages with consecutive integers so that there will no longer be any abuse of notation.)

Lemma 2.12. *For every $i, j \in N$ and $m \in M$ such that $m_i + 1 \in M_i$ and $m_j + 1 \in M_j$,*

$$\psi(m_i, m_{-i}) = \alpha < \beta = \psi(m_i + 1, m_{-i}) \\ \implies \psi(m_j, m_{-j}) = \alpha < \beta = \psi(m_j + 1, m_{-j}). \quad (2.D.7)$$

Moreover, for every $i, j \in N$ and $m \in M$ such that $m_i + 1 \in M_i$ and $m_j - 1 \in M_j$,

$$\psi(m_i, m_{-i}) = \alpha < \beta = \psi(m_i + 1, m_{-i}) \\ \implies \psi(m_i + 1, m_j - 1, m_{-j}) = \alpha < \beta = \psi(m_i + 1, m_j, m_{-j}). \quad (2.D.8)$$

Proof. To show (2.D.7), consider the schematic of message profiles

$$\begin{array}{ccc} (m_i, m_j + 1, m_{-ij}) & \longrightarrow & (m_i + 1, m_j + 1, m_{-ij}) \\ \uparrow & & \uparrow \\ (m_i, m_j, m_{-ij}) & \longrightarrow & (m_i + 1, m_j, m_{-ij}). \end{array}$$

By Lemma 2.10, ψ jumping along the bottom arrow, i.e., taking different values at the bottom left and bottom right message profiles, implies that $V_i(\theta_i | m_j, m_{-ij}) = 0$, where $\theta_i = \sup \sigma_i^{-1}(m_i) = \inf \sigma_i^{-1}(m_i + 1)$. By Lemma 2.9, $V_i(\theta_i | m_j + 1, m_{-ij}) > 0$. Hence, ψ cannot jump along the top arrow. By the same argument, ψ can only either jump along the left or along the right arrow. If ψ jumps along the right arrow, then $\psi(m_i, m_j, m_{-ij}) = \psi(m_i + 1, m_j + 1, m_{-ij})$ by going along the left and top arrow and $\psi(m_i, m_j, m_{-ij}) < \psi(m_i + 1, m_j + 1, m_{-ij})$ by going along the bottom and right arrow due to the monotonicity of ψ (Lemma 2.10)—a contradiction. Hence, ψ must jump along the left arrow, which proves (2.D.7). Finally, (2.D.8) follows from similar arguments. $\square$

Proof of Lemma 2.11. Let (M,ψ,σ) be a partitional posterior implementation of ϕ. For the sake of contradiction, suppose that M_i is infinite. Let $m_i \in M_i$ be such that $m_i + \ell \in M_i$ for all $\ell \in \mathbb{N}$; the case where there are only $m_i \in M_i$ such that $m_i - \ell \in M_i$ for all $\ell \in \mathbb{N}$ is analogous. By convexity, there exists $m_{-i} \in M_{-i}$ such that $\psi(m_i, m_{-i}) \neq \psi(m_i + 1, m_{-i})$.

Step 1. *For every $j \neq i$, $m_j - 1 \in M_j$ unless $m_j = \min M_j$.*

For the sake of contradiction, suppose $m_j > \min M_j$ but $m_j - 1$ does not exist. Then, there exist infinitely many $m_j' < m_j$, and $|\inf \sigma_j^{-1}(m_j) - \sup \sigma_j^{-1}(m_j')|$ can be chosen arbitrarily small.

Consider the schematic of message profiles

$$\begin{array}{ccccc} (m_i, m_j, m_{-ij}) & \longrightarrow & (m_i+1, m_j, m_{-ij}) & \longrightarrow & (m_i+2, m_j, m_{-ij}) \\ \downarrow & & \downarrow & & \downarrow \\ (m_i, m_j', m_{-ij}) & \longrightarrow & (m_i+1, m_j', m_{-ij}) & \longrightarrow & (m_i+2, m_j', m_{-ij}). \end{array}$$

By construction, ψ jumps along the top left arrow, i.e., takes different values along the top left and top center message profiles. By the monotonicity of V_i (Lemma 2.9), ψ cannot jump along the bottom-left and top-right arrows. By the continuity and monotonicity of V_j (Lemma 2.9), for $|\inf \sigma_j^{-1}(m_j) - \sup \sigma_j^{-1}(m_j')|$ small enough, ψ can either only jump along the left, center, or right arrow. Thus, for $|\inf \sigma_j^{-1}(m_j) - \sup \sigma_j^{-1}(m_j')|$ small enough, ψ must jump along the center arrow for otherwise, going down-right or right-down starting from the top left message profile would yield different values for ψ—a contradiction. By the same argument, ψ must then jump along the bottom right arrow. But ψ must jump along the bottom right arrow for any $m_j' < m_j$ such that $|\inf \sigma_j^{-1}(m_j) - \sup \sigma_j^{-1}(m_j')|$ is sufficiently small, which is impossible by the monotonicity of V_i. ∎

Step 2. *There exists $j \neq i$ such that $m_j - \ell \in M_j$ for all $\ell \in \mathbb{N}$.*

For the sake of contradiction, suppose for every $j \neq i$ there exists $\ell_j \in \mathbb{N} \cup \{0\}$ such that $m_j - \ell_j = \min M_j$. By iteratively applying (2.D.7) and (2.D.8), $\psi(m_i + \sum_{j\neq i} \ell_j, \min M_{-i}) \neq \psi(m_i + \sum_{j \neq i} \ell_j + 1, \min M_{-i})$. By the monotonicity of V_i, $\psi(m_i + \ell_i, m_{-i}) = \psi(m_i + \ell_i + 1, m_{-i})$ for all $\ell_i > \sum_{j \neq i} \ell_j$ and $m_{-i} \in M_{-i}$, which contradicts that (M,ψ,σ) is convex. ∎

Now, the proof of Lemma 2.11 can be completed as follows. Note that for every $k \neq i,j$, $m_k \neq \min M_k$ or $m_k \neq \max M_k$ since ϕ is responsive. Suppose first that there is some $k \neq i,j$ such that $m_k \neq \min M_k$.

Consider the following schematic of messages profiles for any $\ell \in \mathbb{N}$

$$\begin{array}{ccc} (m_i+\ell, m_j-\ell, m_k, m_{-ijk}) & \longrightarrow & (m_i+\ell+2, m_j-\ell, m_k, m_{-ijk}) \\ \downarrow & & \downarrow \\ (m_i+\ell, m_j-\ell, m_k-1, m_{-ijk}) & \longrightarrow & (m_i+\ell+2, m_j-\ell, m_k-1, m_{-ijk}). \end{array}$$

By iteratively applying (2.D.7) and (2.D.8), ψ must jump along the top arrow. Also by (2.D.8), $\psi(m_i+\ell+1, m_j-\ell, m_k, m_{-ijk}) \neq \psi(m_i+\ell+1, m_j-\ell, m_k-1, m_{-ijk})$, and thus, by the monotonicity of V_k, ψ cannot jump along the right arrow. By the monotonicity of ψ (Lemma 2.10), ψ must then jump along the bottom arrow.

By Lemma 2.10, the jump at the top arrow requires $V_i(\sup\sigma_i^{-1}(m_i+\ell)|m_j-\ell, m_k, m_{-ijk}) = 0$. Also by Lemma 2.10, the change at the bottom arrow requires $V_i(\sup\sigma_i^{-1}(m_i+\ell+1)|m_j-\ell, m_k-1, m_{-ijk}) = 0$. However, this is impossible for large enough $\ell \in \mathbb{N}$ by the continuity and monotonicity of V_i together with the fact that $|\sup\sigma_i^{-1}(m_i+\ell+1) - \sup\sigma_i^{-1}(m_i+\ell)|$ becomes arbitrarily small.

If $m_k \neq \max M_k$ for all $k \neq i, j$, then the argument is similar with agent j in place of agent i (and "-" and "+" exchanged in the argument above). □

Proof of Lemma 2.1. If $n = 2$, then responsiveness is not needed for the arguments because any non-responsive posterior implementable social choice function is trivially posterior implementable by dictatorship, which is a score voting mechanism.

The first part of the claim immediately follows from Lemma 2.13 in the next subsection.

For the second part, recall that in the proof of Lemma 2.11, it is shown that whenever some partitional posterior implementation (M, ψ, σ) is not finite, then there are messages (m_1, m_2) such that $m_1 + \ell \in M_1$ for all $\ell \in \mathbb{N}$, $m_2 - \ell \in M_2$ for all $\ell \in \mathbb{N}$, and $\psi(m_1, m_2) \neq \psi(m_1+1, m_2)$. By iteratively applying (2.D.7) and (2.D.8), $\psi(m_1+\ell, m_2-\ell) \neq \psi(m_1+\ell+1, m_2-\ell)$ and $\psi(m_1+\ell+1, m_2-\ell-1) \neq \psi(m_1+\ell+1, m_2-\ell)$ for all $\ell \in \mathbb{N}$. By Lemma 2.10, $V_1(\sup\sigma_1^{-1}(m_1+\ell)|m_2-\ell) = 0$ and $V_2(\sup\sigma_2^{-1}(m_2-\ell-1)|m_1+\ell+1) = 0$ for all $\ell \in \mathbb{N}$. Since Θ is compact, there must be a type profile $\theta \in \Theta$ such that for every neighborhood $B(\theta)$ of θ there is an $\underline{\ell} \in \mathbb{N}$ such that for all $\ell > \underline{\ell}$, $\sigma_1^{-1}(m_1+\ell) \times \sigma_2^{-1}(m_2-\ell-1) \subset B(\theta)$. By the continuity of valuation functions, $v_1(\theta) = v_2(\theta) = 0$ (see also p.84 in Green and Laffont, 1987). □

2.D.4 Score Voting

Lemma 2.13. *If (M, ψ, σ) is a finite partitional posterior implementation, then (M, ψ) is a score voting mechanism.*

Proof. Let (M, ψ, σ) be a finite partitional posterior implementation of ϕ. For each $i \in N$, identify the messages in M_i with consecutive integers as in the definition of score voting, following the total order on M_i induced by σ_i.

If ϕ is constant, then (M, ψ) is trivially a score voting mechanism. Thus suppose ϕ is not constant. Then there exists a message profile $m \in M$ and an agent $i \in N$ such that $\psi(m_i, m_{-i}) = \alpha \neq \beta = \psi(m_i+1, m_{-i})$. Let $q = \sum_{j\in N} m_j$. Iteratively applying (2.D.7) and (2.D.8) yields $\psi(m_j, m_{-j}) = \alpha < \beta = \psi(m_j+1, m_{-j})$ for all $m \in M$ and $j \in N$ such that $\sum_{k\in N} m_k = q$ and $m_j + 1 \in M_j$. The monotonicity of V_i

(Lemma 2.9) and Lemma 2.10 imply that ψ jumps between aggregate scores q and $q+1$ only; thus (M,ψ) is a score voting mechanism. $\square$

Proof of Lemma 2.2. Let (M,ψ) be a score voting mechanism. Let σ be a pure surjective posterior equilibrium in (M,ψ).

The first step is to show that σ has a cutoff structure. For the sake of contradiction, suppose $\theta_i' > \theta_i$ but $\sigma_i(\theta_i') < \sigma_i(\theta_i)$. The third requirement in the definition of score voting implies that there exists $m_{-i} \in M_{-i}$ such that $\psi(\sigma_i(\theta_i'), m_{-i}) = 0$ and $\psi(\sigma_i(\theta_i), m_{-i}) = 1$. The monotonicity of V_i (Lemma 2.9) rules out that σ is a posterior equilibrium. Hence σ is non-decreasing with weakly ordered cutoffs $(\theta_i^1, \ldots, \theta_i^{|M_i|-1})$ for each agent $i \in N$.

Now, suppose for the sake of contradiction that there exists an agent $i \in N$ and an integer $0 < k < |M_i| - 1$ such that $\theta_i^k = \theta_i^{k+1}$. Then, there are pivotal events $m, m' \in PIV_i$ such that $m_i = k$, $m_i' = k+1$, $m_j' = m_j - 1$, and $m_{-ij} = m'_{-ij}$. Posterior equilibrium requires in particular that

$$V_i(\theta_i^k | [\theta_j^{m_j-1}, \theta_j^{m_j}]_{j\neq i}) \;=\; 0 \tag{2.D.9}$$

$$V_i(\theta_i^{k+1} | [\theta_j^{m_j'-1}, \theta_j^{m_j'}]_{j\neq i}) \;=\; 0, \tag{2.D.10}$$

which, by the monotonicity of V_i (Lemma 2.9), implies that $\theta_j^{m_j-2} = \theta_j^{m_j}$, contradicting that σ_j is a well-defined strategy.

The second step to show that pivotal types are indifferent. For the sake of contradiction, suppose $\sum_{j\in N} m_j = q$ and $m_i + 1 \in M_i$ but $V_i\left(\sup \sigma_i^{-1}(m_i) | m_{-i}\right) \neq 0$. Since σ is non-decreasing and surjective and V_i is continuous (Lemma 2.9), there exist $\theta_i \in \sigma_i^{-1}(m_i)$ and $\theta_i' \in \sigma_i^{-1}(m_i+1)$ such that $V_i(\theta_i|m_{-i}) V_i(\theta_i'|m_{-i}) > 0$, which rules out that σ is a posterior equilibrium.

For the converse, let σ be a pure, surjective and non-decreasing. Let $i \in N$ and $\theta \in \Theta$ be arbitrary. Posterior equilibrium of a score voting mechanism (M,ψ) (cf. Condition (2.2)) requires that, for all $m_i \in M_i$,

$$V_i(\theta_i|\sigma_{-i}(\theta_{-i}))\mathbb{1}\{\sigma_i(\theta_i) + \sum_{j\neq i} \sigma_j(\theta_j) > q\} \geq V_i(\theta_i|\sigma_{-i}(\theta_{-i}))\mathbb{1}\{m_i + \sum_{j\neq i} \sigma_j(\theta_j) > q\}. \tag{2.D.11}$$

By distinguishing the following two cases, it can be shown that (2.3) implies (2.D.11).

First, if $V_i(\theta_i|\sigma_{-i}(\theta_{-i})) < 0$ and $\sigma_i(\theta_i) + \sum_{j\neq i} \sigma_j(\theta_j) > q$, then $m_i + \sum_{j\neq i} \sigma_j(\theta_j) > q$ for all $m_i \in M_i$. Otherwise, if $m_i + \sum_{j\neq i} \sigma_j(\theta_j) = q$, then, since σ is non-decreasing and V_i is monotone, $V_i(\sup \sigma_i^{-1}(m_i)|\sigma_{-i}(\theta_{-i})) < 0$, which contradicts (2.3).

Second, if $V_i(\theta_i|\sigma_{-i}(\theta_{-i})) > 0$ and $\sigma_i(\theta_i) + \sum_{j\neq i} \sigma_j(\theta_j) \leq q$, then $m_i + \sum_{j\neq i} \sigma_j(\theta_j) \leq q$ for all $m_i \in M_i$. Otherwise, if $m_i + \sum_{j\neq i} \sigma_j(\theta_j) = q+1$, then, since σ is non-decreasing and V_i is monotone, $V_i(\sup \sigma_i^{-1}(m_i - 1)|\sigma_{-i}(\theta_{-i})) > 0$, which contradicts (2.3). □

The following pivotal voting argument for majority voting shows that the Bayes-Nash equilibrium conditions are convex combinations of the posterior equilibrium conditions.

Lemma 2.14. *A strategy profile σ in a super- or submajority mechanism (M, ψ) is Bayes-Nash equilibrium if and only if for all $i \in N$ and $\theta_i \in \Theta_i$,*

$$\sum_{m_{-i} \in M_{-i}:\, \sum_{j \neq i} m_j = q} V_i(\theta_i | m_{-i}) \lambda(m_{-i}|\theta_i) \begin{cases} \geq 0 & \text{if } \sigma_i(\theta_i) = 1 \\ = 0 & \text{if } \sigma_i(\theta_i) \in (0,1) \\ \leq 0 & \text{if } \sigma_i(\theta_i) = 0, \end{cases} \tag{2.D.12}$$

where $\lambda(m_{-i}|\theta_i) = \int_{\Theta_{-i}} \mathbb{1}\{\sigma_{-i}(\theta_{-i}) = m_{-i}\} \, \mu(d\theta_{-i}|\theta_i)$ is the probability with which type θ_i expects others to play m_{-i} given σ_{-i} (cf. (2.A.1)).

Proof. By the definition of a super- or submajority mechanism (M, ψ), the Bayes-Nash equilibrium conditions (2.B.1) require that for all $i \in N$ and $\theta_i \in \Theta_i$,

$$\int_{\Theta_{-i}} v_i(\theta_i, \theta_{-i}) \mathbb{1}\left\{\sum_{j \neq i} \sigma_{-i}(\theta_{-i}) = q\right\} \mu(d\theta_{-i}|\theta_i) \tag{2.D.13}$$

$$= \sum_{m_{-i} \in M_{-i}:\, \sum_{j \neq i} m_j = q} \int_{\Theta_{-i}} v_i(\theta_i, \theta_{-i}) \mathbb{1}\{\sigma_{-i}(\theta_{-i}) = m_{-i}\} \, \mu(d\theta_{-i}|\theta_i) \tag{2.D.14}$$

$$= \sum_{m_{-i} \in M_{-i}:\, \sum_{j \neq i} m_j = q} V_i(\theta_i | m_{-i}) \lambda(m_{-i}|\theta_i) \begin{cases} \geq 0 & \text{if } \sigma_i(\theta_i) = 1 \\ = 0 & \text{if } \sigma_i(\theta_i) \in (0,1) \\ \leq 0 & \text{if } \sigma_i(\theta_i) = 0. \end{cases} \tag{2.D.15}$$

□

As a side remark, if types are independent, then the left-hand side of (2.D.12) is strictly increasing; consequently, any Bayes-Nash equilibrium of majority voting must feature cutoff strategies if types are independent.

Appendix 2.E Proofs for Section 2.4

2.E.1 Proofs for Section 2.4.1

The following Lemma 2.15 shows that heterogeneity is a generic property; a result that will be needed in the proof of Theorem 2.2. Indeed, Lemma 2.15 shows that heterogeneity is a generic property even if it is required to hold in all environments that can be obtained by removing a subset of agents, yet not revealing anything about these agents' information. Thus, Lemma 2.15 backs up the claim in Remark 2.3 that considering responsive social choice functions is essentially without loss of generality.

Lemma 2.15. *Fix any prior density f. There is a subset $\mathscr{G}$ of valuation profiles that is open, dense, and prevalent in $\mathscr{V}$ such that if $v \in \mathscr{G}$, then for every subset $J \subset N$ of agents, the valuation profile $v^* = (v_i^*)_{i \in J}$ on Θ_J, where*

$$v_i^*(\theta_J) = \int_{\Theta_{-J}} v(\theta_J, \theta_{-J}) f(\theta_{-J} | \theta_J)\, d\theta_{-J},$$

satisfies heterogeneity.

The proof is deferred to the very end of this subsection since it uses arguments that are similar to—yet much simpler than—those in the proof of Theorem 2.2.

The proof of Theorem 2.2 needs two technical auxiliary results. The first result is a version of Lemma 2.9.

Lemma 2.16. *For all $j \neq i$, let $0 \leq \underline{\theta}_j \leq \bar{\theta}_j \leq 1$. The posterior expected valuation $V_i(\theta_i | [\underline{\theta}_j, \bar{\theta}_j]_{j \neq i})$ is continuous and strictly increasing in all $2n-1$ arguments and continuously differentiable in the interior of the domain.*

Proof. The monotonicity assertion is again standard (e.g. Milgrom and Weber, 1982, Theorem 5). Continuous differentiability follows from similar arguments as in the proof of Lemma 2.9 by applying the Leibniz integration rule and then using the assumption that the valuation function v_i and the prior density f are continuously differentiable over a compact domain. Continuity at the boundary follows from standard limit arguments. □

Lemma 2.17. *Let $(I_1, \ldots, I_K)$ be a partition of $[0, 1]$ into non-degenerate subintervals. Moreover, for each subinterval $k \in \{1, \ldots, K\}$, let X_k be a real-valued random variable with distribution μ_k that is supported on I_k. Then, the matrix of moments*

$$E = \begin{bmatrix} \mathbb{E}(X_1) & \ldots & \mathbb{E}(X_1^K) \\ \vdots & \ddots & \vdots \\ \mathbb{E}(X_K) & \ldots & \mathbb{E}(X_K^K) \end{bmatrix} \tag{2.E.1}$$

has full rank K.

Proof. For the sake of contradiction, suppose there exist coefficients $(\alpha_1, \ldots, \alpha_K) \in \mathbb{R}^K \setminus \{0\}$ such that

$$\alpha_1 \begin{bmatrix} \mathbb{E}(X_1) \\ \vdots \\ \mathbb{E}(X_K) \end{bmatrix} + \ldots + \alpha_K \begin{bmatrix} \mathbb{E}(X_1^K) \\ \vdots \\ \mathbb{E}(X_K^K) \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{2.E.2}$$

By the fundamental theorem of algebra, the polynomial $p : x \mapsto \alpha_1 x + \ldots + \alpha_k x^K$ has at most $K-1$ non-zero roots. Thus, there must exist a subinterval I_k without a root of p in its interior. Consequently, by the continuity of p,

$$0 \neq \int_{I_k} p(x)\, \mu_k(dx) \;=\; \alpha_1 \mathbb{E}(X_k) + \ldots + \alpha_K \mathbb{E}(X_k^K), \tag{2.E.3}$$

which completes the proof by contradiction. □

Theorem 2.2 will follow almost immediately from the following result.

Lemma 2.18. *Let (M, ψ) be a score voting mechanism except unanimity. Then, there is a subset of valuation profiles $\mathcal{G}$ that is open, dense, and prevalent in $\mathcal{V}$ such that if $v \in \mathcal{G}$, then there exists no pure surjective posterior equilibrium in (M, ψ).*

Proof. Let (M, ψ) be a score voting mechanism with quota q except unanimity. Say that (M, ψ) is a *veto mechanism* if $|M_i| \geq 2$ for all $i \in N$ and every agent has a veto message as in the definition of score voting (Definition 2.2). In particular, unanimity is a veto mechanism. Moreover, note that if (M, ψ) is a veto mechanism, then $|M_i| = |M_j|$ for all $i, j \in N$ because either $q = 1 + \sum_{j \neq i} |M_j|$ for all $i \in N$ or $q = (n-1) + |M_i|$ for all $i \in N$.

Let $PIV = \sqcup_{j \in N} PIV_j$ denote the set of all agents' pivotal events.

Step 1. *Construct a specific family of pivotal events $PIV^* \subset PIV$ such that for every agent $i \in N$, there exists an agent $j \neq i$ such that the message profiles in PIV_i^* admit a strict total order along the messages of agent j. If (M, ψ) is a veto mechanism, then $|PIV^*| = \sum_{i \in N} |M_i| - n$. If (M, ψ) is not a veto mechanism, then $|PIV^*| > \sum_{i \in N} |M_i| - n$.*

It is immediate from the definitions of score voting and pivotal events that for any distinct agents $i, j \in N$, $|\operatorname{proj}_j(PIV_i)|$ equals the number of non-veto messages of agent j. Then, for any agent $i \in N$, let $j \in \arg\max_{k \in N \setminus \{i\}} |M_k|$. Moreover, let $\ell \in \arg\min_{k \in N} |M_k|$.

Let $PIV_i^* \subset PIV_i$ be a subset such that every non-veto message of agent j appears exactly once in PIV_i^*. In addition, if i has a veto message for rejection, then let $PIV_i^* \subset PIV_i$ be such that $\min \operatorname{proj}_\ell PIV_i^*$ is as large as possible; if i has a veto message for acceptance, then let $PIV_i^* \subset PIV_i$ be such that $\max \operatorname{proj}_\ell PIV_i^*$ is as small as possible. Let $PIV^* = \sqcup_{k \in N} PIV_k^*$.

It remains to show the claims about the cardinality of PIV^*. Let $j \in \arg\max_{i \in N} |M_i|$, $k \in \arg\max_{i \in N \setminus \{j\}} |M_i|$, and $b = |M_j| - |M_k|$. Distinguish three cases.

First, if j has no veto message, then no other agent has a veto message. Thus,

$$\begin{aligned} |PIV^*| = (n-1)|M_j| + |M_k| \geq \left(\sum_{i \neq j} |M_i| + (n-1)b \right) + \big(|M_j| - b\big) \\ \geq \sum_{i \in N} |M_i| - n + 1. \end{aligned} \tag{2.E.4}$$

Second, if j has two veto messages, then no other agent has a veto message and $b \geq 2$. Thus,

$$|PIV^*| = (n-1)(|M_j|-2) + |M_k| \geq \left(\sum_{i \neq j} |M_i| + (n-1)(b-2)\right) + \left(|M_j| - b\right)$$
$$\geq \sum_{i \in N} |M_i| - n + 1. \quad (2.E.5)$$

Third, if j has exactly one veto message, then

$$|PIV^*| \;=\; (n-1)(|M_j|-1) + |M_k| - \mathbb{1}\{b \;=\; 0\}. \qquad (2.E.6)$$

If (M, ψ) is a veto mechanism, then $|M_i| = |M_j|$ for all $i \in N$; thus $|PIV^*| = \sum_{i \in N} |M_i| - n$. Otherwise, there exists an agent $i \in N$ such that $|M_i| < |M_j|$; thus $|PIV^*| > \sum_{i \in N} |M_i| - n$.

■

For the next step, define a mapping

$$T_v : [0,1]^{\sum_{i \in N} |M_i| - n} \to \mathbb{R}^{|PIV|}$$

where for $v \in \mathscr{V}$ and arguments $0 \leq \theta_1^1 \leq \ldots \leq \theta_1^{|M_1|-1} \leq 1, \ldots, 0 \leq \theta_n^1 \leq \ldots \leq \theta_n^{|M_n|-1} \leq 1$,

$$T_v(\theta_1^1, \ldots, \theta_1^{|M_1|-1}, \ldots, \theta_n^1, \ldots, \theta_n^{|M_n|-1}) \;=\; \left(V_i(\theta_i^{m_i} | [\theta_j^{m_j-1}, \theta_j^{m_j}]_{j \neq i})\right)_{(m,i) \in PIV}. \qquad (2.E.7)$$

By Lemma 2.2, a root of T_v exactly corresponds to a pure surjective posterior equilibrium.

Also define $S \subset [0,1]^{\sum_{i \in N} |M_i| - n}$ to be the subset of strictly ordered cutoffs $0 < \theta_1^1 < \ldots < \theta_1^{|M_1|-1} < 1, \ldots, 0 < \theta_n^1 < \ldots < \theta_n^{|M_n|-1} < 1$, i.e., those cutoffs that correspond to posterior equilibria where strategies are non-degenerate at the boundary of the type space.

Finally, let $\mathscr{P}^d \subset C^1(\Theta, \mathbb{R}^n)$ denote the finite-dimensional subspace of (n-tuples of multivariate) polynomials of degree at most d. (Recall that every polynomial can be identified with a point in some finite-dimensional Euclidean space via the coefficients with respect to its monomial basis.)

Step 2. *For every $v \in \mathscr{V}$, $T_{v+p} = 0$ has no solution over S for (Lebesgue-)almost all $p \in \mathscr{P}^{|PIV^*|}$.*[24]

24. It is consistent to refer to Lebesgue null sets on a finite-dimensional subspace $E \subset \mathscr{V}$ because linear isomorphisms preserve null sets. An exact characterization of Lebesgue measure on the subspace E depends on the choice of basis for E, i.e., the specific linear isomorphism between E and Euclidean space.

Let $T_v^*: S \to \mathbb{R}^{|PIV^*|}$ result from T_v by restricting T_v to S and then considering the subset of maps in T_v corresponding to the pivotal events in PIV^*. Clearly, if $T_v^* = 0$ has no solution, then $T_v = 0$ has no solution over S. Also define

$$\mathscr{T}_v^*: S \times \mathscr{P}^{|PIV^*|} \to \mathbb{R}^{|PIV^*|}$$

by setting $\mathscr{T}_v^*(\cdot, p) = T_{v+p}^*(\cdot) = T_v^*(\cdot) + T_p^*(\cdot)$ for $v \in \mathscr{V}$ and $p \in \mathscr{P}^{|PIV^*|}$. Note that $\mathscr{T}_v^*$ is continuously differentiable in the cutoffs $(\theta_1^1, \ldots, \theta_1^{|M_1|-1}, \ldots, \theta_n^1, \ldots, \theta_n^{|M_n|-1})$ as well as the polynomial coefficients of p by similar arguments as in Lemma 2.16 (see also the fourth step below).

First, suppose (M, ψ) is not a veto mechanism. By the preimage theorem and the fact that $|PIV^*| > \sum_{i \in N} |M_i| - n$, if 0 is a regular value of T_{v+p}^*, then $T_{v+p}^* = 0$ has no solution. By a simple parametric transversality theorem (see e.g. Mas-Colell, 1985, Proposition 8.3.1.), if 0 is a regular value of $\mathscr{T}_v^*$, then 0 is a regular value of T_{v+p}^* for almost all $p \in \mathscr{P}^{|PIV^*|}$. A sufficient condition for 0 to be a regular value of $\mathscr{T}_v^*$ is that the Jacobian $D\mathscr{T}_v^*$ of $\mathscr{T}_v^*$ has full rank $|PIV^*|$, which is shown in the next paragraph. (Each row of the Jacobian $D\mathscr{T}_v^*$ contains the derivatives of a posterior expected valuation corresponding to a pivotal event in PIV^* with respect to all equilibrium cutoffs and polynomial coefficients.)

For each agent i, consider the columns of $D\mathscr{T}_v^*$ with respect to the coefficients of the monomials $\theta_j, \theta_j^2, \ldots, \theta_j^{|PIV_i^*|}$ in p_i, where agent $j \neq i$ is the agent from the previous step along whose messages the pivotal events in PIV_i^* can be linearly ordered. By differentiating under the integral sign, the entry corresponding to pivotal event $m \in PIV_i^*$ in the ℓ-th column, $1 \leq \ell \leq |PIV_j^*|$, reads

$$\int_{\Theta_{-i}} \theta_j^\ell f(\theta_{-i} | \theta_i^{m_i}, [\theta_k^{m_k-1}, \theta_k^{m_k}]_{k \neq i})\, d\theta_{-i} = \mathbb{E}[\theta_j^\ell | \theta_i^{m_i}, [\theta_k^{m_k-1}, \theta_k^{m_k}]_{k \neq i}]. \tag{2.E.8}$$

For all pivotal events $(m, k) \in PIV^*$ where $k \neq i$, the entry is 0. By Lemma 2.17, the columns must be linearly independent. The collection of all such columns across all agents $i \in N$ must also be linearly independent because the corresponding submatrix of $D\mathscr{T}_v^*$ is block diagonal. Hence $D\mathscr{T}_v^*$ has rank $|PIV^*|$.

Now, suppose (M, ψ) is a veto mechanism except unanimity. Suppose that $\underline{m}_i = 1 \in M_i$ is the veto message of every agent $i \in N$. (The case where $\bar{m}_i = |M_i| \in M_i$ is the veto message of every agent $i \in N$ is symmetric.) Let (M', ψ') be the non-veto mechanism that results from (M, ψ) by removing the veto message of the agent $\ell \in N$ from the previous step, i.e., $M'_{-\ell} = M_{-\ell}$, $M'_\ell = M_\ell \setminus \{1\}$, and $\psi' = \psi|_{M'}$.[25]

The claim follows because if $T_{v+p}^{*\prime} = 0$ has no solution for the mechanism (M', ψ'), then $T_{v+p}^* = 0$ has no solution for the mechanism (M, ψ). Indeed, note that T_{v+p}^*

25. (M', ψ') is equivalent to a score voting mechanism with quota $q - 1$ by translating the message space of agent ℓ by -1. However, it will be convenient to work with the particular isomorphism (M', ψ').

differs from $T^{*\prime}_{v+p}$ only in the addition of a new cutoff variable θ^1_ℓ and an additional pivotal event associated to the message $m_\ell = 1$. However, by the construction of PIV^*, the equilibrium conditions in T^*_{v+p} that are also in $T^{*\prime}_{v+p}$ are not a function of the additional cutoff type θ^1_ℓ because for every $i \neq \ell$, the message $m_\ell = 2$ does not appear in PIV^*_i whenever (M, ψ) is not the unanimity mechanism. ■

Step 3. *For every $v \in \mathscr{V}$, $T_{v+p} = 0$ has no solution for (Lebesgue-)almost all $p \in \mathscr{P}^{|PIV^*|}$.*

The only solutions to $T_{v+p} = 0$ that were not considered in the previous step are those that correspond to posterior equilibria where strategies are degenerate at the boundary of the type space. The reason is that continuous differentiability of V_i is not well-defined at the boundary of S.

To circumvent the problem, consider for any non-empty subsets $J, J' \subset N$ of agents the subset $S_{J,J'} \subset [0,1]^{\sum_{i\in N}|M_i|-n}$ where $\theta^1_i = 0$ for all $i \in J$, $\theta^1_i \neq 0$ for all $i \notin J$, $\theta^{|M_i|-1}_i = 1$ for $i \in J'$, and $\theta^{|M_i|-1}_i \neq 1$ for $i \notin J'$. By the same arguments as in Lemma 2.16, V_i is continuously differentiable over the relatively open set $S_{J,J'}$. The arguments from the previous step now go through unchanged.

Finally, there are only finitely many subsets $J, J' \subset N$, and the union of finitely many Lebesgue nullsets is still Lebesgue null; hence the claim. ■

Let $\mathscr{G} \subset \mathscr{V}$ denote the set of valuation profiles $v \in \mathscr{V}$ for which $T_v = 0$ has no solution.

Step 4. *$\mathscr{G}$ is open.*

As a preliminary step, note that $T_v(x)$ (as a functional) is continuous in (v, x); for $(v^\varepsilon, x^\varepsilon) \to (v, x)$, write

$$\begin{aligned}|T_v(x) - T_{v^\varepsilon}(x^\varepsilon)| &= |T_v(x) - T_v(x^\varepsilon)| + |T_v(x^\varepsilon) - T_{v^\varepsilon}(x^\varepsilon)| \\ &\leq |T_v(x) - T_v(x^\varepsilon)| + (||v_i - v^\varepsilon_i||)_{i\in N}. \qquad (2.E.9)\end{aligned}$$

The first summand converges to zero by Lemma 2.16, and the second summand converges to zero by assumption.

For the sake of contradiction, suppose there exists $v \in \mathscr{G}$ such that for every $\varepsilon > 0$, there exists $v^\varepsilon \in \mathscr{V}$ and $x^\varepsilon \in [0,1]^{\sum_{i\in N}|M_i|-n}$ such that $||v - v^\varepsilon|| < \varepsilon$ and $T_{v^\varepsilon}(x^\varepsilon) = 0$. Since $[0,1]^{\sum_{i\in N}|M_i|-n}$ is compact, it is without loss of generality to assume that x^ε converges to $x \in [0,1]^{\sum_{i\in N}|M_i|-n}$ as $\varepsilon \to 0$. By the continuity of $T_v(x)$ in (v,x), $T_v(x) = 0$, which contradicts that $v \in \mathscr{G}$. ■

Step 5. *$\mathscr{G}$ is dense and prevalent.*

By the previous step, $\mathscr{G}$ is Borel. Consider Lebesgue measure on the the finite-dimensional subspace $\mathscr{P}^{|PIV^*|}$. By the third step, $\mathscr{V} \setminus \mathscr{G}$ meets every translate of $\mathscr{P}^{|PIV^*|}$ in a set of measure zero. This is equivalent to $\mathscr{P}^{|PIV^*|}$ meeting every translate of $\mathscr{V} \setminus \mathscr{G}$ in a set of measure zero. Thus, $\mathscr{G}$ is prevalent. Moreover, any prevalent set is dense (Hunt, Sauer, and Yorke, 1992, Fact 2'). ■

By construction, if $T_v = 0$ has no solution, then there exists no pure surjective posterior equilibrium in (M, ψ). $\square$

Remark 2.8 (Finite prevalence)**.** Prevalent sets that are constructed as in the proof of Lemma 2.18 are called *finitely prevalent* in Anderson and Zame (2001).

Proof of Theorem 2.2. The assertion follows immediately follows from Theorem 2.1, Lemma 2.15, Lemma 2.18, and the facts that the number of score voting mechanisms is countable (up to the choice of α and β, which are irrelevant for incentives), that the countable intersection of open dense sets is residual, and that the countable intersection of prevalent sets is prevalent (Hunt, Sauer, and Yorke, 1992, Fact 3”). $\square$

Proof of Lemma 2.15. The proof shows that for any arbitrary subset $J \subset N$ of agents, the set of valuation profiles $v \in \mathscr{V}$ such that 0 is regular value of v^*—i.e., such that the Jacobian Dv^* of v^* has full rank J whenever $v^* = 0$—is open, dense, and prevalent. This assertion immediately implies Lemma 2.15 because (1) that 0 is regular value of v^* implies heterogeneity, (2) there are only finitely many subsets of N, and (3) the three desired properties are preserved under finite intersections.

Note that $Dv^*(\theta_J)$ is continuous (as a functional) in (v, θ_J) by similar arguments as in the fourth step in the proof of Lemma 2.18.

The argument for openness is again similar to the fourth step in the proof of Lemma 2.18. Consider a sequence of valuation profiles $v^\varepsilon \to v \in \mathscr{V}$ and critical points $\theta_J^\varepsilon \to \theta_J$ such that $v^{\varepsilon *}(\theta_J^\varepsilon) = 0$ and $Dv^{\varepsilon *}(\theta_J^\varepsilon)$ is singular. (Convergence of θ_J^ε in Θ_J is without loss of generality by compactness.) Then, $v^*(\theta_J) = 0$ and $Dv^{\varepsilon *}(\theta_J)$ is singular by continuity and the fact that linear dependence is a closed property. Thus, the complement of the set under consideration is closed.

The argument for prevalence (hence denseness) is much simpler than in the proof of Lemma 2.18. Consider any $v \in \mathscr{V}$. Claim that 0 is a regular value of $(v + p)^*$ for almost all $p \in \mathscr{P}^0$, where $\mathscr{P}^0 \subset \mathscr{V}$ is the one-dimensional subspace of constant functions. The claim and hence prevalence follows immediately from Sard’s theorem by noting that $(v + p)^* = v^* + p^*$, where $p^* = \mathrm{proj}_J(p)$. $\square$

2.E.2 Proofs for Section 2.4.2

Proof of Theorem 2.3. Let $n = 2$. As an intermediate step, the proof establishes that all non-decreasing and deterministic Bayesian implementable social choice functions admit a characterization that is reminiscent of the pivotal voting argument.

For a given non-decreasing and deterministic social choice function ϕ, define $\theta_{-i}^*(\theta_i)$ to be the unique type in Θ_{-i} such that

$$\phi(\theta_i, \theta_{-i}) = \begin{cases} 0 & \text{if } \theta_{-i} < \theta_{-i}^*; \\ 1 & \text{if } \theta_{-i} > \theta_{-i}^*. \end{cases}$$

Note that $\theta^*_{-i}(\theta_i)$ is non-increasing because ϕ is non-decreasing.

For $\theta_i \in (0,1)$, define

$$C_i(\theta_i) = \lim_{\varepsilon \downarrow 0} [\theta^*_{-i}(\theta_i + \varepsilon), \theta^*_{-i}(\theta_i - \varepsilon)]$$

and also define

$$C_i(0) = \lim_{\varepsilon \downarrow 0} [\theta^*_{-i}(\varepsilon), \theta^*_{-i}(0)] \qquad C_i(1) = \lim_{\varepsilon \downarrow 0} [\theta^*_{-i}(1), \theta^*_{-i}(1-\varepsilon)].$$

These sets are well-defined because a monotone function admits both left and right limits. Henceforth, perturbations for boundary types are assumed to be carried out as above without further mention.

Say that type $\theta_i \in \Theta_i$ is *critical* if $\theta^*_{-i}(\theta_i)$ is not locally constant at θ_i. In plain words, slightly perturbing a critical type θ_i affects the decision for a strictly positive measure of types of the other agent. Let $\Theta^c_i \subset \Theta_i$ denote the set of critical types of agent $i \in \{1,2\}$.

Note that if $\theta_i \in \Theta_i \setminus \Theta^c_i$ is non-critical, then there is a neighborhood of θ_i in which all types are non-critical. Thus, the set of non-critical types must be a countable union of intervals that are open except at the boundaries of the type space. Moreover, on each interval, $\theta^*_{-i}(\theta_i)$ is constant.

Step 1. *If Θ^c_i is non-empty, then $\{C_i(\theta_i)\}_{\theta_i \in \Theta^c_i}$ is a cover of $[\theta^*_{-i}(1), \theta^*_{-i}(0)] \subset \Theta_{-i}$ such that for all distinct $\theta_i, \theta'_i \in \Theta^c_i$, $C_i(\theta_i)$ and $C_i(\theta'_i)$ intersect at most in their boundaries.*

From the definition of C_i, it is immediate that $\{C_i(\theta_i)\}_{\theta_i \in \Theta_i}$ is a cover of $[\theta^*_{-i}(1), \theta^*_{-i}(0)]$.

Note that the endpoints of each interval of non-critical types must be critical by definition, except perhaps on the boundary of the type space. Thus, by the definition of C_i and the assumption that Θ^c_i is non-empty, for every $\theta_i \in \Theta_i \setminus \Theta^c_i$ there exists an endpoint $\theta'_i \in \Theta^c_i$ such that $C_i(\theta_i) \subset C_i(\theta'_i)$. Consequently, $\{C_i(\theta_i)\}_{\theta_i \in \Theta^c_i}$ is a cover of $[\theta^*_{-i}(1), \theta^*_{-i}(0)]$.

Finally, that $C_i(\theta_i)$ and $C_i(\theta'_i)$ intersect at most in their boundaries follows immediately from the definition of C_i and the fact that $\theta^*_{-i}(\theta_i)$ is non-increasing. ■

For $\Theta'_i \subset \Theta_i$, define $C_i(\Theta'_i) = \bigcup_{\theta_i \in \Theta'_i} C_i(\theta_i)$. By the previous step, $\mu(\cdot|\theta_i) \circ C_i \in \Delta(\Theta_i)$ is a well-defined probability measure.

Step 2. *A non-decreasing and deterministic social choice function ϕ is Bayesian implementable if and only if for both $i \in \{1,2\}$ and their critical types $\theta_i \in \Theta^c_i$,*

$$V_i(\theta_i|C_i(\theta_i)) \;=\; \int_{\Theta_{-i}} v_i(\theta_i, \theta_{-i}) f(\theta_{-i}|\theta_i, C_i(\theta_i))\, d\theta_{-i} \;=\; 0. \tag{2.E.10}$$

By the revelation principle, ϕ is Bayesian implementable if and only if it is Bayesian incentive compatible: for all $i \in N$ and $\theta_i, \theta_i' \in \Theta_i$,

$$\int_{\Theta_{-i}} v_i(\theta_i, \theta_{-i}) f(\theta_{-i}|\theta_i) \left[\phi(\theta_i, \theta_{-i}) - \phi(\theta_i', \theta_{-i})\right] d\theta_{-i} \geq 0. \tag{2.E.11}$$

Since ϕ is non-decreasing and deterministic, instead of (2.E.11), one can equivalently require for all $i \in N$ and $\theta_i, \theta_i' \in \Theta_i$ that

$$\int_{\theta^*_{-i}(\theta_i)}^{\theta^*_{-i}(\theta_i')} v_i(\theta_i, \theta_{-i}) f(\theta_{-i}|\theta_i)\, d\theta_{-i} \geq 0. \tag{2.E.12}$$

First, suppose ϕ is Bayesian incentive compatible. Then, in particular for all $\theta_i \in \Theta_i^c$ and $\varepsilon > 0$,

$$V_i(\theta_i - \varepsilon | [\theta^*_{-i}(\theta_i + \varepsilon), \theta^*_{-i}(\theta_i - \varepsilon)]) \leq 0 \tag{2.E.13}$$
$$V_i(\theta_i + \varepsilon | [\theta^*_{-i}(\theta_i + \varepsilon), \theta^*_{-i}(\theta_i - \varepsilon)]) \geq 0 \tag{2.E.14}$$

by using $\theta_i - \varepsilon$ and $\theta_i + \varepsilon$ in (2.E.12), each once as a true type and once as a deviation.

By the continuity of V_i (Lemma 2.16), taking limits yields

$$V_i(\theta_i | C_i(\theta_i)) = 0. \tag{2.E.15}$$

For the converse, suppose (2.E.10) holds for all $i \in \{1, 2\}$ and $\theta_i \in \Theta_i^c$. Let $i \in N$ and $\theta_i, \theta_i' \in \Theta_i$ be arbitrary. Suppose $\theta_i > \theta_i'$. One has

$$\int_{\theta^*_{-i}(\theta_i)}^{\theta^*_{-i}(\theta_i')} v_i(\theta_i, \theta_{-i})\, \mu(d\theta_{-i}|\theta_i) \tag{2.E.16}$$
$$= \int_{\theta_i'}^{\theta_i} \int_{\theta^*_{-i}(\theta_i)}^{\theta^*_{-i}(\theta_i')} v_i(\theta_i, \theta_{-i})\, \mu(d\theta_{-i}|\theta_i, C_i(\tilde{\theta}_i))\, \mu(\cdot|\theta_i) \circ C_i(d\tilde{\theta}_i) \tag{2.E.17}$$
$$= \int_{\theta_i'}^{\theta_i} V_i(\theta_i | C_i(\tilde{\theta}_i))\, \mu(\cdot|\theta_i) \circ C_i(d\tilde{\theta}_i) \tag{2.E.18}$$
$$\geq \int_{\theta_i'}^{\theta_i} V_i(\tilde{\theta}_i | C_i(\tilde{\theta}_i))\, \mu(\cdot|\theta_i) \circ C_i(d\tilde{\theta}_i) = 0. \tag{2.E.19}$$

The first equality is the law of iterated expectations. The second equality is a definition. The inequality follows from Lemma 2.16. The final equality follows because $\mu(\cdot|\theta_i) \circ C_i$-almost every type $\theta_i \in \Theta_i$ is critical (the function $\theta^*_{-i}(\cdot)$ takes at most countably many values over the set of non-critical types). If $\theta_i < \theta_i'$, then the inequality is reversed, as desired. ■

Step 3. *If ϕ is non-decreasing, deterministic, and Bayesian implementable, then ϕ is posterior implementable.*

For each interval of non-critical types, select a representative type. For each non-critical type $\theta_i \in \Theta_i \setminus \Theta_i^c$, let $\sigma_i(\theta_i) \in \Theta_i$ denote the associated representative. If $\theta_i \in \Theta_i^c$ is on the boundary of an interval of non-critical types, then let $\sigma_i(\theta_i) = \sigma_i(\theta_i')$ for θ_i' in one of the intervals where θ_i is on the boundary. If $\theta_i \in \Theta_i^c$ is not on the boundary of an interval of non-critical types, then let $\sigma_i(\theta_i) = \theta_i$.

Claim that $(\operatorname{im}\sigma, \phi|_{\operatorname{im}\sigma}, \sigma)$ is an implementation of ϕ. Indeed, there are at most countably many types that are on the boundary of an interval of non-critical types. Moreover, for $\theta_i, \theta_i' \in \Theta_i \setminus \Theta_i^c$ in the same interval of non-critical types, $\phi(\theta_i, \theta_{-i}) = \phi(\theta_i', \theta_{-i})$ for all $\theta_{-i} \in \Theta_{-i} \setminus \{\theta_{-i}^*(\theta_i)\}$. Thus, $\phi = \phi \circ \sigma$ μ-almost everywhere.

It remains to show that σ is a posterior equilibrium in $(\operatorname{im}\sigma, \phi|_{\operatorname{im}\sigma})$. Posterior equilibrium (cf. (2.2)) requires that for all $i \in N$, $\theta_i \in \Theta_i$ $\theta_i' \in \operatorname{im}\sigma_i$ and $\theta_{-i} \in \operatorname{im}\sigma_{-i}$,

$$V_i(\theta_i|\theta_{-i})\phi(\sigma_i(\theta_i), \theta_{-i}) \;\geq\; V_i(\theta_i|\theta_{-i})\phi(\theta_i', \theta_{-i}). \tag{2.E.20}$$

If $\phi(\sigma_i(\theta_i), \theta_{-i}) = \phi(\theta_i', \theta_{-i})$, then the equilibrium condition is trivially satisfied; thus, suppose $\phi(\sigma_i(\theta_i), \theta_{-i}) \neq \phi(\theta_i', \theta_{-i})$. Thus, $\theta_i^*(\theta_{-i})$ is in between θ_i and θ_i'. Moreover, by the definition of σ_i, $\theta_i^*(\theta_{-i}) \in \Theta_i^c$.

Note that

$$\theta_{-i} \;\in\; C_i(\theta_i^*(\theta_{-i})) \;=\; \lim_{\varepsilon \downarrow 0}[\theta_{-i}^*(\theta_i^*(\theta_{-i}) + \varepsilon), \theta_{-i}^*(\theta_i^*(\theta_{-i}) - \varepsilon)]. \tag{2.E.21}$$

If not, then there exists $\varepsilon > 0$ such that $\theta_{-i} \notin [\theta_{-i}^*(\theta_i^*(\theta_{-i}) + \varepsilon), \theta_{-i}^*(\theta_i^*(\theta_{-i}) - \varepsilon)]$. By the definition of θ_{-i}^*, $\phi(\theta_i^*(\theta_{-i}) + \varepsilon, \theta_{-i}) = 0$ or $\phi(\theta_i^*(\theta_{-i}) - \varepsilon, \theta_{-i}) = 1$, which contradicts the definition of θ_i^*.

Claim that $C_i(\theta_i^*(\theta_{-i})) = \operatorname{cl}\sigma_{-i}^{-1}(\theta_{-i})$. First, suppose that $C_i(\theta_i^*(\theta_{-i}))$ is not a singleton. The proof proceeds by showing that $\theta_i^*(\theta_{-i}') = \theta_i^*(\theta_{-i})$ for all $\theta_{-i}' \in \operatorname{int} C_i(\theta_i^*(\theta_{-i}))$. Indeed, if there exists $\theta_{-i}' \in \operatorname{int} C_i(\theta_i^*(\theta_{-i}))$ such that $\theta_i^*(\theta_{-i}') < \theta_i^*(\theta_{-i})$, then $\phi(\theta_i^*(\theta_{-i}) - \varepsilon, \theta_{-i}') = 1$ for $\varepsilon > 0$ sufficiently small. Thus, $\theta_{-i}^*(\theta_i^*(\theta_{-i}) - \varepsilon) \leq \theta_{-i}'$ for all $\varepsilon > 0$ sufficiently small, which contradicts that $\theta_{-i}' \in \operatorname{int} C_i(\theta_i^*(\theta_{-i}))$. If there only exists $\theta_{-i}' \in C_i(\theta_i^*(\theta_{-i}))$ such that $\theta_i^*(\theta_{-i}') > \theta_i^*(\theta_{-i})$, then the argument is analogous. By the definition of σ_i, one can conclude from the previous argument that $\operatorname{int} C_i(\theta_i^*(\theta_{-i})) \subset \operatorname{int}\sigma_{-i}^{-1}(\theta_{-i})$. If $\theta_{-i}' \in \operatorname{int}\sigma_{-i}^{-1}(\theta_{-i})$, then $\theta_i^*(\theta_{-i}') = \theta_i^*(\theta_{-i})$. By the previous paragraph, $\theta_{-i}' \in C_i(\theta_i^*(\theta_{-i}')) = C_i(\theta_i^*(\theta_{-i}))$. Thus, $\operatorname{int} C_i(\theta_i^*(\theta_{-i})) = \operatorname{int}\sigma_{-i}^{-1}(\theta_{-i})$. In particular, it has been shown that if $\sigma_{-i}^{-1}(\theta_{-i})$ is not a singleton, then $C_i(\theta_i^*(\theta_{-i}))$ is not a singleton. Finally then, $C_i(\theta_i^*(\theta_{-i}))$ is a singleton if and only if $\sigma_{-i}^{-1}(\theta_{-i})$ is a singleton; thus, it follows from the previous paragraph and the definition of σ_i that $C_i(\theta_i^*(\theta_{-i})) = \sigma_{-i}^{-1}(\theta_{-i})$.

Since ϕ is Bayesian implementable,

$$V_i(\theta_i''|C_i(\theta_i'')) \;=\; 0. \tag{2.E.22}$$

The claim then follows by the fact that $C_i(\theta_i'')$ and $\sigma_{-i}^{-1}(\theta_{-i})$ coincide except on their boundaries and the monotonicity of V_i (Lemma 2.16). □

Proof of Proposition 2.1. Let $n=2$. Let (M,ψ) be any score voting mechanism with quota q. By the third requirement in the definition of score voting (Definition 2.2), for every agent $i \in \{1,2\}$ and every $m_i \in M_i$, there is exactly one message $m_{-i} \in M_{-i}$ such that $m_i + m_{-i} = q$. Then, by Lemma 2.2, the cutoff profile $(\theta_1^1, \dots, \theta_1^{|M_1|-1}, \theta_2^1, \dots, \theta_2^{|M_2|-1})$ is a pure surjective posterior equilibrium in (M,ψ) if for all $i \in N$ and $m_i < |M_i|$,

$$V_i(\theta_i^{m_i} | [\theta_{-i}^{q-m_i-1}, \theta_{-i}^{q-m_i}]) = 0. \tag{2.E.23}$$

Define $S^i \subset [0,1]^{|M_i|-1}$ to be the subset of strictly ordered cutoffs $0 < \theta_i^1 < \dots < \theta_i^{|M_i|-1} < 1$. Define mappings $T^i : \operatorname{cl} S^i \to \operatorname{cl} S^{-i}$ as follows: for each $k = 1, \dots, |M_{-i}|-1$, let the value $T_k^i(\theta_i^1, \dots, \theta_i^{|M_i|-1}) \in (0,1)$ of the k-th mapping be the solution to

$$V_{-i}(\cdot | [\theta_i^{q-k-1}, \theta_i^{q-k}]) = 0; \tag{2.E.24}$$

this solution is uniquely defined by the assumption of partisan types and the monotonicity and continuity of V_{-i} (Lemma 2.16). It is an immediate consequence of Lemma 2.16 that T_i is continuous.

The composite mapping $T_2 \circ T_1$ is a continuous self-map on the compact convex set $\operatorname{cl} S^1$, thus has a fixed point $(\theta_1^1, \dots, \theta_1^{|M_1|-1})$ by Brouwer's fixed point theorem. By construction, if the cutoff profile $(\theta_1^1, \dots, \theta_1^{|M_1|-1}, T_1(\theta_2^1, \dots, \theta_2^{|M_2|-1}))$ is in $S_1 \times S_2$, then it corresponds to a pure surjective posterior equilibrium in (M,ψ). Thus, it remains to show that every fixed point of $T_2 \circ T_1$ yields a cutoff profile in $S_1 \times S_2$. Indeed, note that if $(\theta_1^1, \dots, \theta_1^{|M_1|-1}) \in S^1$, then $T_1(\theta_2^1, \dots, \theta_2^{|M_2|-1}) \in S^2$ by the monotonicity of V_2 (Lemma 2.16).

For the sake of contradiction, suppose $(\theta_1^1, \dots, \theta_1^{|M_1|-1}) \in \operatorname{bndr} S^1$ is a fixed point of $T_2 \circ T_1$. Let $(\theta_2^1, \dots, \theta_2^{|M_2|-1}) = T_1(\theta_1^1, \dots, \theta_1^{|M_1|-1})$. By the assumption of partisan types, $\theta_1^1, \theta_2^1 > 0$ and $\theta_1^{|M_1-1|}, \theta_2^{|M_2-1|} < 1$. Thus, there must exist $k = 1, \dots, |M_1|-2$ such that $\theta_1^k = \theta_1^{k+1}$. It must hold that

$$V_1(\theta_1^k | [\theta_2^{q-k-1}, \theta_2^{q-k}]) = V_1(\theta_1^{k+1} | [\theta_2^{q-k-2}, \theta_2^{q-k-1}]) = 0. \tag{2.E.25}$$

By the monotonicity of V_1 (Lemma 2.16), $\theta_2^{q-k-2} = \theta_2^{q-k}$. Similarly, it must hold that

$$V_2(\theta_2^{q-k-2} | [\theta_1^{k+1}, \theta_1^{k+2}]) = V_2(\theta_2^{q-k} | [\theta_1^{k-1}, \theta_1^k]) = 0. \tag{2.E.26}$$

By the monotonicity of V_2 (Lemma 2.16), $\theta_1^{k-1} = \theta_1^{k+2}$. Iterating this argument forward eventually yields $0 = \theta_0^1 = \theta_1^1 = \dots = \theta_1^{M_1-1} = \theta_1^{M_1} = 1$, a contradiction. This completes the proof.

□

2.E.3 Proofs for Section 2.4.3

Proof of Proposition 2.2. Let σ be a a strategy profile in a unanimity mechanism (M, ψ). Let $\phi = \psi \circ \sigma$. In either unanimity mechanism, there is exactly one pivotal event $m^i \in PIV_i$ for each agent $i \in N$. If there exists an agent $i \in N$ such that $\lambda(\{m^i_{-i}\}) = 0$ (cf. (2.A.1)), i.e., i is never pivotal, then ϕ is μ-almost everywhere constant and thus both Bayesian and posterior implementable. If $\lambda(\{m^i_{-i}\}) > 0$ for all agents $i \in N$, then the Bayes-Nash equilibrium condition (2.D.12) requires that for all $i \in N$ and $\theta_i \in \Theta_i$,

$$V_i(\theta_i|m^*_{-i}) \begin{cases} \geq 0 & \text{if } \sigma_i(\theta_i) \;=\; 1 \\ = 0 & \text{if } \sigma_i(\theta_i) \;\in\; (0,1) \\ \leq 0 & \text{if } \sigma_i(\theta_i) \;=\; 0. \end{cases} \tag{2.E.27}$$

Given the monotonicity of $V_i(\theta_i|m^*_{-i})$ (Lemma 2.9), σ is characterized by cutoffs as in Lemma 2.2, and thus the Bayes-Nash equilibrium condition (2.E.27) is equivalent to the posterior equilibrium condition (2.3). This completes the proof. □

Proof of Proposition 2.3. By Lemma 2.2, there exists a responsive social choice function that is posterior implementable by unanimity for acceptance if and only if there exists a pure surjective posterior equilibrium σ of unanimity voting such that the equilibrium cutoffs are interior, i.e., $\theta^*_i = \theta^1_i \in (0,1)$ for all $i \in N$.

Consider the mapping $T : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$T(\theta^*_1, \ldots, \theta^*_n) \;=\; \left(V_i(\theta^*_i|[\theta^*_j, 1]_{j \neq i})\right)_{i \in N}. \tag{2.E.28}$$

By definition, there exist a pure surjective posterior equilibrium of unanimity if and only if $T = 0$ has an interior solution.

Note that T is continuous by Lemma 2.16. That $T = 0$ has an interior solution under the assumption of partisan types follows immediately from the Poincaré-Miranda theorem. In fact, it is enough to assume that for each agent $i \in N$, type $\theta_i = 0$ prefers rejecting the alternative regardless of others' types, and type $\theta_i = 1$ prefers accepting the alternative ex-interim.

The proof for the unanimity for rejection mechanism is symmetric with the assumption that for each agent $i \in N$, type $\theta_i = 1$ prefers accepting the alternative regardless of others' types, and type $\theta_i = 0$ prefers rejecting the alternative ex-interim. □

2.E.4 Proofs for Section 2.4.4

Proof of Proposition 2.4. Let σ be a symmetric strategy profile, i.e., $\sigma_i = \sigma_j$ for all $i, j \in N$, in a super- or submajority mechanism (M, ψ) with quota q. Clearly, an anonymous social choice function is implementable by (M, ψ) if and only if it is implementable by (M, ψ) in symmetric strategies.

For a symmetric strategy profile σ and a symmetric density f, the distribution over others' messages $\lambda(\cdot|\theta_i) \in \Delta(M_{-i})$ is invariant under permutations of the coordinates. Consequently, the left-hand side of the Bayes-Nash equilibrium condition (2.D.12) is strictly increasing by Lemma 2.9; thus, σ is a cutoff strategy profile as in Lemma 2.2 that is characterized by the same cutoff $\theta^* \in [0,1]$ for all agents.

Note that in a symmetric environment, there exists a posterior expected valuation function $\tilde{V}(\theta_i|S_{-i})$ that is (1) symmetric in the subsets $S_j \subset [0,1]$ for $j \neq i$ and such that (2) for every agent $i \in N$, $V_i(\theta_i|S_{-i}) = \tilde{V}(\theta_i|S_{-i})$.

Thus, in a symmetric environment, the Bayes-Nash equilibrium condition (2.D.12) simply requires that if $\theta^* \in (0,1)$, then

$$\tilde{V}(\theta^*|[0,\theta^*],\dots,[0,\theta^*],[\theta^*,1],\dots,[\theta^*,1]) \;=\; 0, \tag{2.E.29}$$

where the number of intervals $[\theta^*,1]$ equals q, i.e., the number of affirmative votes of others needed for any given agent to be pivotal in (M,ψ). The previous display is also exactly the posterior equilibrium condition (cf. (2.3)).

If $\theta^* \in \{0,1\}$, then ϕ is μ-almost everywhere constant and thus both Bayesian and posterior implementable. This completes the proof. □

Proof of Proposition 2.5. The assumption of partisan types ensures that the posterior expected valuation $\tilde{V}(\theta^*|[0,\theta^*],\dots,[0,\theta^*],[\theta^*,1],\dots,[\theta^*,1])$ from the proof of Proposition 2.4 is strictly negative at $\theta^* = 0$ and strictly positive at $\theta^* = 1$. Moreover, it is continuous by Lemma 2.16. Hence, there exists $\theta^* \in (0,1)$ such that (2.E.29) is satisfied, which completes the proof. □

Appendix 2.F Proofs for Section 2.5

Proof of Lemma 2.3. Consider any mechanism (M,ψ) and strategy profile σ. Let $B(i) \subset N$ denote the set of agents that move before agent $i \in N$, and let $A(i) = N \setminus (B(i) \cup \{i\})$ denote the set of agents that move simultaneously with or after agent i.

First, suppose that σ is a posterior equilibrium in (M,ψ). Consider any extensive form of (M,ψ). Fix any belief system $\{\mu(\cdot|\theta_i,h)\}_{i\in N,\theta_i\in\Theta_i,h\in H}$ that is derived from μ and σ according to Bayes' rule whenever possible and that satisfies

(1) for all terminal histories $h \in H$, $\mu(\cdot|\theta_i,h) = \mu(\cdot|\theta_i,m_{-i}(h))$;
(2) for all $i \in N$, $h \in H(i)$, every measurable $\tilde{\Theta}_{-i} \in \Theta_{-i}$, and every measurable $\tilde{M}_{-i} \in M_{-i}$, the belief system is consistent with the terminal beliefs:

$$\int_{\tilde{\Theta}_{-i}} \sigma_{A(i)}(\theta_{A(i)})[\tilde{M}_{A(i)}]\mathbb{1}\{m_{B(i)}(h) \in \tilde{M}_{B(i)}\}\,\mu(d\theta_{-i}|\theta_i,h) = \int_{\tilde{M}_{-i}} \mu(\tilde{\Theta}_{-i}|\theta_i,m_{-i})\,\lambda(dm_{-i}|\theta_i,h), \tag{2.F.1}$$

where (cf. (2.A.1))

$$\lambda(\tilde{M}_{-i}|\theta_i, h) = \int_{\Theta_{-i}} \sigma_{A(i)}(\theta_{A(i)})[\tilde{M}_{A(i)}]\mathbb{1}(m_{B(i)}(h) \in \tilde{M}_{B(i)})\, \mu(d\theta_{-i}|\theta_i, h). \tag{2.F.2}$$

In particular, conditions (1) and (2) are always satisfied for beliefs that can be derived via Bayes' rule.

It can now be verified that σ together with the belief system $\{\mu(\cdot|\theta_i, h)\}_{i\in N, \theta_i \in \Theta_i, h\in H}$ is a history-independent PBE. Sequential rationality requires that for all $i \in N$, $h \in H(i)$, $\theta_i \in \Theta_i$, and $\tilde{m}_i \in M_i$,

$$\begin{aligned}\int_{\Theta_{-i}} & v_i(\theta_i, \theta_{-i})\psi(\sigma_i(\theta_i), h, \sigma_{A(i)}(\theta_{A(i)}))\, \mu(d\theta_{-i}|\theta_i, h) \\ &\geq \int_{\Theta_{-i}} v_i(\theta_i, \theta_{-i})\psi(\tilde{m}_i, h, \sigma_{A(i)}(\theta_{A(i)}))\, \mu(d\theta_{-i}|\theta_i, h). \end{aligned} \tag{2.F.3}$$

By using definitions and (2.F.1), this condition can be rewritten as follows:

$$\begin{aligned} \iff & \int_{\Theta_{-i}}\int_{M_{A(i)}} v_i(\theta_i, \theta_{-i})\psi(\sigma_i(\theta_i), h, m_{A(i)})\, \sigma_{A(i)}(\theta_{A(i)})[dm_{A(i)}]\, \mu(d\theta_{-i}|\theta_i, h) \\ & \geq \int_{\Theta_{-i}}\int_{M_{A(i)}} v_i(\theta_i, \theta_{-i})\psi(\tilde{m}_i, h, m_{A(i)})\, \sigma_{A(i)}(\theta_{A(i)})[dm_{A(i)}]\, \mu(d\theta_{-i}|\theta_i, h) \end{aligned} \tag{2.F.4}$$

$$\begin{aligned} \iff & \int_{M_{-i}}\int_{\Theta_{-i}} v_i(\theta_i, \theta_{-i})\psi(\sigma_i(\theta_i), m_{-i})\, \mu(d\theta_{-i}|\theta_i, m_{-i})\, \lambda(dm_{-i}|\theta_i, h) \\ & \geq \int_{M_{-i}}\int_{\Theta_{-i}} v_i(\theta_i, \theta_{-i})\psi(\tilde{m}_i, m_{-i})\, \mu(d\theta_{-i}|\theta_i, m_{-i})\, \lambda(dm_{-i}|\theta_i, h) \end{aligned} \tag{2.F.5}$$

$$\begin{aligned} \iff & \int_{M_{-i}} V_i(\theta_i|m_{-i})\psi(\sigma_i(\theta_i), m_{-i})\, \lambda(dm_{-i}|\theta_i, h) \\ & \geq \int_{M_{-i}} V_i(\theta_i|m_{-i})\psi(\tilde{m}_i, m_{-i})\, \lambda(dm_{-i}|\theta_i, h). \end{aligned} \tag{2.F.6}$$

The last line is satisfied because it is an average over the posterior equilibrium conditions. Thus, σ together with the belief system $\{\mu(\cdot|\theta_i, h)\}_{i\in N, \theta_i\in\Theta_i, h\in H}$ is a history-independent PBE.

Second, suppose that σ is the strategy profile of a history-independent PBE in every extensive form of (M, ψ). In particular, for every agent $i \in N$ there is an extensive form where i is the unique last mover, i.e., $A(i) = \emptyset$. Since for every history $h \in H(i)$ it holds that $\mu(\cdot|\theta_i, h) = \mu(\cdot|\theta_i, m_{-i}(h))$, sequential rationality requires that for all $h \in H(i)$, $\theta_i \in \Theta_i$, and $\tilde{m}_i \in M_i$

$$\int_{\Theta_{-i}} v_i(\theta_i, \theta_{-i}) \psi(\sigma_i(\theta_i), m_{-i}) \, \mu(d\theta_{-i} | \theta_i, m_{-i}(h))$$
$$\geq \int_{\Theta_{-i}} v_i(\theta_i, \theta_{-i}) \psi(\tilde{m}_i, m_{-i}) \, \mu(d\theta_{-i} | \theta_i, m_{-i}(h)), \quad (2.F.7)$$

which are exactly the posterior equilibrium conditions for agent i. Thus, σ is a posterior equilibrium in (M, ψ). $\square$

References

Ali, S Nageeb, and Navin Kartik. 2012. "Herding with collective preferences." *Economic Theory* 51 (3): 601–26. [27]

Anderson, Robert M, and William R Zame. 2001. "Genericity with infinitely many parameters." *BE Journal of Theoretical Economics* 1 (1). [68]

Austen-Smith, David, and Jeffrey S Banks. 1996. "Information aggregation, rationality, and the Condorcet jury theorem." *American Political Science Review* 90 (1): 34–45. [35]

Battaglini, Marco. 2005. "Sequential voting with abstention." *Games and Economic Behavior* 51 (2): 445–63. [27]

Bergemann, Dirk, and Stephen Morris. 2005. "Robust mechanism design." *Econometrica* 73 (6): 1771–813. [51]

Callander, Steven. 2007. "Bandwagons and momentum in sequential voting." *Review of Economic Studies* 74 (3): 653–84. [27]

Che, Yeon-Koo, Jinwoo Kim, and Fuhito Kojima. 2015. "Efficient assignment with interdependent values." *Journal of Economic Theory* 158: 54–86. [45]

Cramton, Peter C, and Thomas R Palfrey. 1995. "Ratifiable mechanisms: learning from disagreement." *Games and Economic Behavior* 10 (2): 255–83. [27]

Crawford, Vincent P. 1985. "Efficient and durable decision rules: a reformulation." *Econometrica: Journal of the Econometric Society,* 817–35. [27]

Crawford, Vincent P, and Joel Sobel. 1982. "Strategic information transmission." *Econometrica: Journal of the Econometric Society,* 1431–51. [32, 33, 55]

Dekel, Eddie, and Michele Piccione. 2000. "Sequential voting procedures in symmetric binary elections." *Journal of Political Economy* 108 (1): 34–55. [25–27, 41, 43, 45]

Ekmekci, Mehmet, Carl Heese, and Stephan Lauermann. 2022. "A generalized intermediate value theorem." Mimeo. [38]

Feddersen, Timothy, and Wolfgang Pesendorfer. 1997. "Voting behavior and information aggregation in elections with private information." *Econometrica: Journal of the Econometric Society,* 1029–58. [35]

Feng, Tangren, Axel Niemeyer, and Qinggong Wu. 2023. "The limits of ex post implementation without transfers." *Theoretical Economics* 18 (2): 463–79. [23, 27, 30, 32, 33, 39, 57]

Feng, Tangren, and Qinggong Wu. 2020. "Robust Binary Voting." Mimeo. [27]

Forges, Françoise. 1994. "Posterior efficiency." *Games and Economic Behavior* 6 (2): 238–61. [27]

Fudenberg, Drew, and Jean Tirole. 1991. *Game theory.* MIT press. [41]

Gaertner, Wulf, and Yongsheng Xu. 2012. "A general scoring rule." *Mathematical Social Sciences* 63 (3): 193–96. [27]

Gordon, Sidartha. 2010. "On infinite cheap talk equilibria." Mimeo. [33]

Green, Jerry R, and Jean-Jacques Laffont. 1987. "Posterior implementability in a two-person decision problem." *Econometrica: Journal of the Econometric Society,* 69–94. [24–26, 28, 30, 33, 44–46, 53, 60]

Heifetz, Aviad, and Zvika Neeman. 2006. "On the generic (im)-possibility of full surplus extraction in mechanism design." *Econometrica* 74 (1): 213–33. [36]

Holmström, Bengt, and Roger B Myerson. 1983. "Efficient and durable decision rules with incomplete information." *Econometrica: Journal of the Econometric Society,* 1799–819. [27]

Hunt, Brian R, Tim Sauer, and James A Yorke. 1992. "Prevalence: a translation-invariant "almost every" on infinite-dimensional spaces." *Bulletin of the American Mathematical Society* 27 (2): 217–38. [36, 67, 68]

Jehiel, Philippe, Moritz Meyer-ter-Vehn, Benny Moldovanu, and William R Zame. 2006. "The limits of ex post implementation." *Econometrica* 74 (3): 585–610. [23, 36, 45]

Jehiel, Philippe, Moritz Meyer-ter-Vehn, Benny Moldovanu, and William R Zame. 2007. "Posterior implementation vs ex-post implementation." *Economics Letters* 97 (1): 70–73. [26, 45]

Kawakami, Kei. 2016. "Posterior renegotiation-proofness in a two-person decision problem." *International Journal of Game Theory* 45 (4): 893–931. [26, 27, 37]

Li, Hao, Sherwin Rosen, and Wing Suen. 2001. "Conflicts and common interests in committees." *American Economic Review* 91 (5): 1478–97. [26, 27, 43–45]

Lopomo, Giuseppe. 2001. "Optimality and robustness of the English auction." *Games and Economic Behavior* 36 (2): 219–40. [26]

Macé, Antonin. 2018. "Voting with evaluations: characterizations of evaluative voting and range voting." *Journal of Mathematical Economics* 79: 10–17. [27]

Mas-Colell, Andreu. 1985. *The Theory of General Economic Equilibrium.* Cambridge University Press. [66]

Milgrom, Paul R, and Robert J Weber. 1982. "A theory of auctions and competitive bidding." *Econometrica: Journal of the Econometric Society,* 1089–122. [33, 46, 49, 56, 63]

Myerson, Roger B. 1995. "Axiomatic derivation of scoring rules without the ordering assumption." *Social Choice and Welfare* 12 (1): 59–74. [27]

Newham, Melissa, and Rune Midjord. 2020. "Do expert panelists herd? Evidence from FDA committees," DIW Berlin Discussion Paper. [24]

Niemeyer, Axel. 2022. "Posterior Implementability in an N-Person Decision Problem." Available at SSRN 4270283. [23]

Nogales, Agustin G. 2013. "The existence of regular conditional probabilities for Markov kernels." *Statistics & Probability Letters* 83 (3): 891–97. [46]

Reny, Philip J, and Motty Perry. 2006. "Toward a strategic foundation for rational expectations equilibrium." *Econometrica* 74 (5): 1231–69. [36]

Urfalino, Philippe, and Pascaline Costa. 2015. "Secret-public voting in FDA advisory committees." *Secrecy and Publicity in Votes and Debates,* 165–95. [24]

Chapter 3

Simple Allocation with Correlated Types*

3.1 Introduction

We consider environments where an object is allocated among a number of agents. The efficient allocation depends on how the agents evaluate their peers, but monetary transfers are not used to elicit this information. A number of environments fit this description:

(1) A group has to elect one of its members to a prestigious post. The group as whole benefits from selecting a qualified candidate, and each agent knows the qualities of their friends in the group. Monetary transfers would naturally be excluded in such an election.

(2) A community of households has to distribute a good among its members. Each member can vouch for the needs and valuations of their friends or neighbors. If some members are financially constrained, it may be infeasible or undesirable to have members compete for the good via bids.

(3) A funding agency splits a budget across researchers. Each researcher can evaluate others in their field. If all parties are risk neutral, the allocated share of the budget can be interpreted as the probability of being allocated the object. Additional monetary transfers would be self-defeating.

In these environments, asking the agents straightforwardly who "should" get the object does not guarantee satisfactory outcomes. In particular, if agents are primarily concerned with their own winning chances, they may exaggerate their individual qualities instead of impartially disclosing their peer information.

To better understand good allocation rules, we take a mechanism design approach and consider the following model. Each agent wants to win the object and

* This chapter is based on Niemeyer and Preusser (2022).

is indifferent to which of the others wins. Allocating to an agent generates a social value. The agents have private information about these values—their *types*. We model peer information by allowing for an arbitrary joint distribution of types and values. Hence an agent's type may be informative about the types and values of all others.

We study mechanisms for maximizing the expected value of the allocation. In a mechanism, each agent is asked to report their type. We focus on mechanisms where truthfully reporting one's type is a dominant strategy; that is, we focus on dominant-strategy incentive-compatible (DIC) mechanisms. For the assumed preferences of the agents, DIC requires that one's report never influences one's own winning probability.

Let us highlight some of the differences to existing models (a detailed review follows later). Alon et al. (2011) and Holzman and Moulin (2013) consider DIC mechanisms (there called strategyproof or impartial) where the agents nominate one another to win the object. These nominations do not arise from some ground truth. By contrast, we fix a general joint distribution of types and values. This lets us study mechanisms where, say, two agents can share their private information and form a consensus about which of the others to nominate. Other work considers settings where non-monetary instruments for screening the agents are available, but where the agents have no peer information (for example, Ben-Porath, Dekel, and Lipman, 2014, 2019).

We contribute two results demonstrating the difficulty of designing "simple" mechanisms for this problem: deterministic DIC mechanisms are not without loss, and anonymous DIC mechanism cannot meaningfully elicit information. We further contribute three positive results on so-called *jury mechanisms*. These mechanisms, described in detail below, solve the problem with three agents, are approximately optimal in symmetric environments with many agents, and are the only deterministic DIC mechanisms satisfying a relaxed notion of anonymity. Let us elaborate.

For each agent, there is a trade-off between allocating to the agent and using the agent's peer information. This trade-off arises since, on the one hand, DIC demands that a change in an agent's type does not affect that agent's own winning probability, but, on the other hand, the change in the type reveals information about the values from allocating to the others.

Optimally resolving this trade-off may require the use of stochastic mechanisms that cannot be implemented by randomizing over deterministic ones. That is, the set of DIC mechanisms may admit stochastic extreme points, and these can be uniquely optimal. Stochastic extreme points exist if and only if there are at least four agents and the type spaces are not "too small." The typical view in the literature is that one should use mechanisms that can be implemented by randomizing over deterministic ones (for example, Pycia and Ünver, 2015; Chen et al., 2019). We find that doing so is not generally without loss in the present problem.

Our next result is that all anonymous DIC mechanisms must ignore the reports of the agents. Here, anonymity means that all agents can make the same reports and that an agent's winning probability does not change when one permutes the reports of the others. We view anonymity as the familiar axiom from social choice theory that no agent play a special role in determining the chosen social alternative; that is, in determining who wins the object. As such, anonymity helps reduce the complexity of the mechanism, protects agents' privacy when evaluating their peers, and ensures that agents have the same rights as voters. Our negative result also sheds new light on a characterization due to Holzman and Moulin (2013) and Mackenzie (2015) of a slightly different notion of anonymity.

Our positive results concern the following class of mechanisms. In a *jury mechanism*, each agent is either a *juror* or a *candidate*. The allocation only depends on the reports of the jurors, and the object is always allocated to a candidate. Given that jurors cannot win, all jury mechanisms are DIC.

If there are three agents, then all DIC mechanisms are randomizations over deterministic jury mechanisms. In particular, a deterministic jury mechanism is optimal. This generalizes a known result for deterministic DIC mechanisms due to Holzman and Moulin (2013). Our key insight is that in the three-agent case all DIC mechanisms are actually randomizations over deterministic ones.

Next, we identify a condition on the environment under which deterministic jury mechanisms are approximately optimal with many agents. By "approximately optimal" we mean that the difference in expected values between an optimal deterministic jury mechanism and an optimal DIC mechanism vanishes as the number of agents diverges. The condition on the environment is that agents are exchangeable in terms of supplying information about the vector of values. Intuitively, when agents are exchangeable, increasing their number relaxes the aforementioned trade-off. In particular, there is essentially no loss from ignoring the reports of those agents who are sometimes allocated the object—this is the defining property of a jury mechanism.

For the last result, we consider a relaxed notion of anonymity—*partial anonymity*. Whereas the earlier notion of anonymity demands that an agent's winning probability be invariant with respect to *all* permutations of the others, partial anonymity only considers permutations of those agents that in the given mechanism *actually* influence the agent's winning probability. We show that all deterministic partially anonymous DIC mechanism are jury mechanisms.

The paper is organized as follows. We next discuss related work (Section 3.2) and present the model (Section 3.3). In Section 3.4, we introduce jury mechanisms and present the results for the three- and many-agent cases. In Section 3.5, we characterize when stochastic extreme points exist. In Section 3.6, we study anonymous mechanisms, presenting the two notions and the associated characterizations side-by-side. We conclude by discussing open questions (Section 3.7). All omitted proofs are in Section 3.A. Supplementary material is collected in Section 3.B and Section 3.C.

3.2 Related Literature

Holzman and Moulin (2013) study axioms for peer nomination rules. In such a rule, agents nominate one another to receive a prize. Their central axiom—*impartiality*—is equivalent to DIC when each agent cares only about their own winning probability. As Holzman and Moulin note, many of their axioms have no obvious counterparts in a model with abstract types. Most relevant for us is their notion and characterization of anonymity, as well subsequent results due to Mackenzie (2015, 2020). We discuss the differences to our characterization in detail in Section 3.6.4.[1]

Alon et al. (2011) initiated a literature on optimal DIC mechanisms (there called *strategyproof* mechanisms) in a model where each agent nominates a subset of the others, and the aim is to select an agent nominated by many. Mechanisms are ranked according to approximation ratios[2] rather than according to expected values, and this leads to qualitatively different optimal mechanisms. For example, while jury mechanisms can be optimal in our model, the 2-*partition mechanism* of Alon et al. (2011), which is a natural analogue of jury mechanisms, is not optimal in their model.[3,4]

See Olckers and Walsh (2022) for a survey of the literature following Holzman and Moulin (2013) and Alon et al. (2011). Olckers and Walsh also report on some related empirical studies.

Other work in mechanism design focuses on non-monetary instruments for eliciting information For example, in the aforementioned paper of Ben-Porath, Dekel, and Lipman (2014), the agents' types can be verified at a cost.[5] The typical assumption in

1. Further contributions to the literature following Holzman and Moulin (2013) include Tamura and Ohseto (2014), Tamura (2016), and Edelman and Por (2021). See also de Clippel, Moulin, and Tideman (2008).

2. Given $\alpha \in [0, 1]$, a mechanism has an *approximation ratio* of α if it guarantees a fraction α of some benchmark value. The guarantee is computed across all realizations of the type profile; that is, across all possible approval sets. The benchmark value at a particular realization is the maximal number of approvals across agents.

3. The 2-partition mechanism randomly splits the agents into two subsets, and then selects an agent from the first subset with the most approvals from agents in the second subset. Alon et al. (2011, Theorem 4.1) show that the 2-partition mechanism has an approximation ratio of $\frac{1}{4}$. Fischer and Klimm (2015) present a mechanism that achieves the strictly higher and optimal ratio of $\frac{1}{2}$.

4. Further contributions to this literature include Bousquet, Norin, and Vetta (2014), Aziz et al. (2016), Bjelde, Fischer, and Klimm (2017), Aziz et al. (2019), Mattei, Turrini, and Zhydkov (2020), and Lev et al. (2021). See also Caragiannis, Christodoulou, and Protopapas (2019, 2021), who consider additive approximations rather than approximation ratios.

5. See Epitropou and Vohra (2019), Erlanson and Kleiner (2019), and Li (2020) for further work with costly verification. Other examples of non-monetary instruments include promises of fu-

this literature is that the agents do not have information about their peers. Most relevant for us are papers that study how a *Bayesian* incentive-compatible mechanism may use agents' peer information to incentivize truthtelling (Kattwinkel, 2019; Kattwinkel and Knoepfle, 2021; Bloch, Dutta, and Dziubiński, 2022; Kattwinkel et al., 2022). The idea is that when agents have information about their peers, one can detect lies by cross-checking the agents' reports. We observe that the *dominant-strategy* incentive-compatible mechanisms that we consider do not use peer information in this manner. While DIC thus shuts down a screening channel, it leads to mechanisms that are far simpler for the agents to play. Relatedly, the fundamental insights of Crémer and Richard P McLean (1985, 1988) and McAfee and Reny (1992) on mechanisms *with transfers* do not apply here.

The papers of Baumann (2018) and Bloch and Olckers (2021, 2022) study related settings but focus on different questions. For instance, Bloch and Olckers (2022) study whether it is possible to reconstruct the ordinal ranking of agents from their reports when agents prefer a high rank.

We also contribute to the literature on the gap between stochastic and deterministic mechanisms[6] by fully characterizing when deterministic DIC mechanisms suffice for describing the set of DIC mechanisms in the present model. Methodologically, we show that here the existence of stochastic extreme points can be understood via a graph-theoretic result due to Chvátal (1975). We elaborate in Section 3.B.

3.3 Model

A single indivisible object is to be allocated to one of n agents, where $n \geq 2$. For each agent i, let Ω_i be a finite set of reals representing the possible social values from allocating to agent i, and let Θ_i be a finite set representing agent i's possible private types. Let $\Omega = \times_{i=1}^{n} \Omega_i$ and $\Theta = \times_{i=1}^{n} \Theta_i$. Values and types are distributed according to a joint distribution μ over $\Omega \times \Theta$. At all type profiles, agent i strictly prefers winning the object to not winning it; agent i is indifferent to which of the others is allocated the object.

In a *(direct) mechanism*, each agent reports a type, and then the object is allocated to one of the agents according to some lottery. Formally, a mechanism is a function $\phi : \Theta \to [0,1]^n$ satisfying $\sum_{i=1}^{n} \phi_i = 1$. Here $\phi_i : \Theta \to [0,1]$ denotes the winning probability of agent i. Since the object is allocated to one of the agents, these probabilities sum to 1. The requirement that the object is always allocated

ture allocations (Guo and Hörner, 2021), costly signaling (Condorelli, 2012; Chakravarty and Kaplan, 2013), allocative externalities (Bhaskar and Sadler, 2019; Goldlücke and Tröger, 2020), or ex-post punishments (Mylovanov and Zapechelnyuk, 2017; Li, 2020).

6. See, for example, Budish et al. (2013), Pycia and Ünver (2015), Jarman and Meisner (2017), Chen et al. (2019), and Rivera Mora (2022).

keeps with some earlier work (for example, Alon et al. (2011) and Holzman and Moulin (2013)). In Section 3.B, we discuss mechanisms that do not always allocate.

A mechanism ϕ is *dominant-strategy incentive-compatible (DIC)* if truthfully reporting one's type is a dominant strategy. For the assumed preferences of the agents, a mechanism is DIC if and only if one's report never affects one's own winning probability.

To see the previous point in detail, let $u_i(\theta)$ denote the payoff to an agent i when i is allocated the object at a type profile θ. We normalize i's payoff when not allocated the object to 0, and we assume $u_i > 0$. DIC for a mechanism ϕ requires that all $i, \theta_i, \theta_i', \theta_{-i}$, and θ_{-i}' satisfy $u_i(\theta_i, \theta_{-i})\phi_i(\theta_i, \theta_{-i}') \geq u_i(\theta_i, \theta_{-i})\phi_i(\theta_i', \theta_{-i}')$. Since $u_i > 0$ and since θ_i and θ_i' are arbitrary, we must have $\phi_i(\theta_i, \theta_{-i}') = \phi_i(\theta_i', \theta_{-i}')$. That is, agent i's report never affects ϕ_i. Observe that nothing in this argument changes if $u_i < 0$. Hence we can equally model cases where some agents prefer not to be allocated the object.

We evaluate a DIC mechanism ϕ via the expected value of the allocation, which is given by $\mathbb{E}_{\omega,\theta}\left[\sum_{i=1}^n \phi_i(\theta)\omega_i\right]$. When we say a DIC mechanism is *optimal*, we mean it maximizes the expected value among all DIC mechanisms. The Revelation Principle implies that DIC mechanisms are without loss: if a mechanism can be implemented in some dominant-strategy equilibrium of some game, then it is DIC.

Lastly, we define the following: A mechanism is *deterministic* if it maps to a subset of $\{0,1\}^n$. A mechanism is *stochastic* if it is not deterministic.

3.4 Jury Mechanisms

In this section, we focus on the following class of mechanisms.

Definition 3.1. A mechanism ϕ is a *jury mechanism* if for all agents i we have the following: if the mechanism is non-constant in agent i's report, then agent i never wins, meaning $\phi_i = 0$.

Given a jury mechanism, we refer to an agent as a *juror* if the mechanism is non-constant in their report. The set of jurors is called the *jury*, and the remaining agents are called *candidates*. All jury mechanisms are DIC since jurors never win.

The most natural jury mechanisms are those that allocate to the top candidate conditional on the jurors' reports. That is, when the set of jurors is J and jurors report types $(\theta_i)_{i\in J}$, the object is allocated to one of the candidates in

$$\underset{k\in\{1,\ldots,n\}\setminus J}{\arg\max}\ \mathbb{E}_{\omega_k}[\omega_k|(\theta_i)_{i\in J}].$$

Assuming a common prior, this mechanism would be implemented by having the jurors share their private information via cheap-talk messages, update their beliefs about the candidates, and then award the object to the top candidate given their shared posterior belief. (For our proofs, however, it is convenient to allow the jurors to select a suboptimal candidate.)

A priori, all agents in the model are candidates for winning and suppliers of information. Jury mechanisms are special since the roles of candidates and jurors are assigned *before* the agents are consulted. There are more complicated mechanisms where an agent's "role" varies across type profiles, and we shall encounter such mechanisms later. As such, it is remarkable that there are situations where jury mechanisms are (approximately) optimal, as we discuss next.

3.4.1 Jury mechanisms solve the three-agent case

Theorem 3.1. *Let $n \leq 3$. A mechanism is DIC if and only if it is a convex combination of deterministic jury mechanisms. In particular, there is an optimal DIC mechanism that is a deterministic jury mechanism.*

With three agents, a jury mechanism admits at most one juror who deliberates between the other two. Therefore, all DIC mechanisms with three agents can be implemented by nominating a juror (according to some distribution over the set of agents), and then asking the juror to pick one of the others as a winner of the object. Optimally, the information of at least two of the agents is ignored. (With only two agents, all DIC mechanisms are constant.)

In the remainder of this subsection, we explain the steps in the proof of Theorem 3.1. We begin with a known result (Holzman and Moulin, 2013, Proposition 2.i).

Lemma 3.1. *If $n \leq 3$, then all deterministic DIC mechanisms are jury mechanisms.*

In the language of Section 5 of Holzman and Moulin (2013), a deterministic DIC mechanism is an *impartial award rule*. Their Proposition 2.i implies that, if $n \leq 3$, then in each impartial award rule there is at most one agent whose report influences the allocation, and this influential agent never wins. Such a rule is a jury mechanism.[7]

To the best of our knowledge, Lemma 3.1 has so far been limited to deterministic DIC mechanisms. We now close the gap to stochastic ones.

Lemma 3.2. *If $n \leq 3$, then all DIC mechanisms are convex combinations of deterministic DIC mechanisms.*

Lemma 3.2 completes the proof of Theorem 3.1. Indeed, Lemma 3.2 and Lemma 3.1 immediately imply that all DIC mechanisms are convex combinations

7. Holzman and Moulin (2013) note that the result is essentially due to Kato and Ohseto (2002), who study pure exchange economics. For a discussion of this relationship, we refer to Section 1.4 of Holzman and Moulin (2013).

of deterministic jury mechanisms. Since the expected value is a linear function of the mechanism, at least one deterministic jury mechanism must be optimal.

To prove Lemma 3.2 we consider the extreme points of the set of DIC mechanisms. A routine argument shows that the set of DIC mechanisms is convex and compact (as a subset of Euclidean space). Hence, by the Krein-Milman theorem (Aliprantis and Border, 2006, Theorem 7.68), the set is given by the convex hull of its extreme points.

We show that all stochastic DIC mechanisms fail to be extreme points. Specifically, given an arbitrary stochastic DIC mechanism ϕ we construct a non-zero function f such that $\phi + f$ and $\phi - f$ are two other DIC mechanisms. To understand this construction, recall that a stochastic mechanism is one where, for at least one type profile, at least one agent enjoys an interior winning probability. Since the object is always allocated, some other agent must also enjoy an interior winning probability at the same profile. The function f represents a shift of a small probability mass between these two agents. This shift should be consistent with DIC (since we want $\phi + f$ and $\phi - f$ to be DIC), and hence we have to shift masses at multiple type profiles. What makes the construction of f difficult is that changing one agent's type may change which of the others enjoys an interior winning probability. Our argument thus intuitively leans on there only being three agents. Indeed, we shall later see that the argument does not go through with four or more agents.

3.4.2 Approximate optimality of jury mechanisms

In this subsection, we identify environments in which jury mechanisms are approximately optimal if the number n of agents is large. As suggested in the introduction, DIC creates a tension between allocating to an agent and using the agent's peer information. This tension becomes easier to resolve with many agents. Indeed, we intuit that many DIC mechanisms become approximately optimal as $n \to \infty$. The insight of the upcoming result is that this includes the DIC mechanisms that resolve the tension in the most straightforward way—jury mechanisms.

The following example conveys the basic idea.

Example 3.1. For each agent i, the value ω_i of allocating to i depends on some common component s and some private component t_i. Specifically, for some function $\hat{\omega}_i$ we have $\omega_i = \hat{\omega}_i(s, t_i)$ with probability 1. The agents observe their private components, which are independently and identically distributed across agents and independent of s. All agents observe s. (So, agent i's type is $\theta_i = (s, t_i)$.) Let ϕ be an arbitrary DIC mechanism for these n agents. Now suppose a new agent $n+1$, who also observes the common component s, joins the group. Agent $n+1$ may observe some additional information, but this will not be relevant. We claim that there is a jury mechanism that only uses agent $n+1$ as a single juror and that does as well as ϕ. Note that, by ignoring the reports of agents 1 to n, the information contained in the public component s is not lost. The only information that is potentially lost is the first n agents' knowledge of their private components $t_1, \dots, t_n$. Each agent i's private component t_i is informative only about i's own value (by independence). However, DIC of the original mechanism ϕ implies that t_i could not have been used to determine i's own allocation. Thus one does not actually lose any information when ignoring the reports of agents 1 to n.

The main result of this section generalizes the previous example as follows. Under an assumption on the distribution of types and values, an arbitrary DIC mechanism with n agents can be replicated by a jury mechanism when additional agents are around. If values remain bounded in n, an implication is that the loss from using an optimal jury mechanism vanishes as $n \to \infty$.

We introduce new notation to accommodate the growing number of agents. The agents share a common finite type space ($\Theta_1 = \Theta_i$ for all i). The prior distribution of values and types is now a Borel-probability measure μ on $\times_{i\in\mathbb{N}}(\Omega_i \times \Theta_i)$,where each Ω_i is a finite set of reals.[8]

The following assumption captures the idea that if i, j, and k are three distinct agents, then i and j have access to the same sources of information about ω_k.

Assumption 3.1. For all $n \in \mathbb{N}$, all $i \in \{1, \dots, n\}$, and all $\omega_i \in \Omega_i$, we have the following: Conditional on the value of agent i being equal to ω_i, the distribution of $(\theta_j)_{j\in\{1,\dots,n\}\setminus\{i\}}$ is invariant with respect to permutations of $\{1, \dots, n\} \setminus \{i\}$.

We are not assuming that i and j have the same information as k about ω_k. For example, in Example 3.1, the common component is the only information that i and j have about ω_k, but agent k actually observes ω_k.

When there are n agents (meaning that mechanisms only consult and allocate to the first n agents), let V_n denote the expected value from an optimal DIC mechanism. Let V_n^J denote the expected value from a jury mechanism with n agents that is optimal among jury mechanisms with n agents.

Theorem 3.2. *Let Assumption 3.1 hold. For all $n \in \mathbb{N}$ there exists $m \in \mathbb{N}$ such that $V_n \le V_{n+m}^J$. If, additionally, the sequence $\{V_n\}_{n\in\mathbb{N}}$ is bounded,*[9] *then* $\lim_{n\to\infty}(V_n - V_n^J) = 0$.

In plain words, if m new agents are added to the group, a jury mechanism with $n + m$ agents does as well as an with an arbitrary DIC mechanism with n agents. The proof shows this claim for a jury mechanism that has the new m agents as jurors, and the old n agents as candidates, and where $m = n$. That is, a jury mechanism with the desired properties exists as soon as the number of agents is doubled. Depending

8. Each of the finite sets Ω_i and Θ_i is equipped with the discrete metric. The product $\times_{i\in\mathbb{N}}(\Omega_i \times \Theta_i)$ is equipped with the product metric.

9. A sufficient condition for boundedness of the sequence $\{V_n\}_{n\in\mathbb{N}}$ is that the values ω_i are bounded across agents. For example, suppose with μ-probability 1 we have $\omega_i \in [0, 1]$ for all $i \in \mathbb{N}$.

on the exact distribution μ, a much smaller number of new agents may be needed; in Example 3.1, one new agent suffices.

Assumption 3.1 is stronger than what we really need. It suffices if, informally speaking, for all groups of agents $\{1, \dots, n\}$ there eventually comes a disjoint group of agents that is at least as well informed as $\{1, \dots, n\}$ about each other. Assumption 3.2 in Section 3.A.1.2 formalizes this idea.

Remark 3.1. Theorem 3.2 does not assert that DIC mechanisms become approximately ex-post optimal conditional on the type profile. In Example 3.1, the only information that is used in the allocation is the common component. The common component need not pin down the entire profile of values.

3.5 Random Allocations

In this section, we show that it typically does not suffice to consider deterministic mechanisms. This fact sheds light on the fundamental economic forces of the model and has practical implications for implementation, as we explain below.

3.5.1 Stochastic extreme points

One of way constructing a stochastic DIC mechanism is by randomizing over deterministic ones; that is, by taking a convex combination of deterministic DIC mechanisms. In this case, one of the deterministic mechanisms from the combination must generate a weakly higher expected value than the stochastic mechanism.

We therefore ask whether all stochastic DIC mechanisms can be represented as convex combinations of deterministic ones; that is, whether all extreme points of the set of DIC mechanisms are deterministic. In a nutshell, this is true if and only if there are at most three agents or the agents' type spaces are small.

Theorem 3.3. *All extreme points of the set of DIC mechanisms are deterministic if and only if at least one of the following is true:*

(1) There are at most three agents; that is, we have $n \leq 3$.

(2) All agents have at most two types; that is, for all i *we have* $|\Theta_i| \leq 2$.

(3) At least $(n-2)$*-many agents have a degenerate type; that is, we have*

$$|\{i \in \{1, \dots, n\} : |\Theta_i| = 1\}| \geq n-2.$$

We already know from Lemma 3.2 that (1) is sufficient for all extreme points to be deterministic. Sufficiency of (2) is related to a generalization of the well-known Birkhoff-von Neumann theorem; sufficiency of (3) is economically and technically uninteresting, but must be included for completeness.[10] As for the other direction: we momentarily give an example of a stochastic extreme point. The general claim that a stochastic extreme point exists when (1) to (3) all fail follows readily by extending this example.

An implication of Theorem 3.3 is that deterministic DIC mechanisms do not suffice for optimality. Indeed, for each extreme point there exists at least one distribution of types and values where the extreme point is the unique optimal DIC mechanisms.[11]

We do not expect stochastic extreme points to closely resemble mechanisms observed in practice. The literature discusses several issues. First, to reduce complexity and opaqueness, it is appealing to implement a mechanism by randomizing over deterministic mechanisms, announcing the selected mechanism, and only then collecting the agents' reports (see, for example, Pycia and Ünver (2015)). A stochastic extreme point is precisely a DIC mechanism that cannot be implemented in this way.[12] Second, to implement a stochastic extreme point, the designer must commit to honoring the outcome of a stochastic process (see, for example, Chen et al. (2019)). A commitment issue arises if the agents' collective information identifies a unique qualified agent but the mechanism nevertheless promises to flip a coin between this agent and a less qualified one.

Despite the above points, it may be acceptable to randomize if this happens "rarely" or is used to break ties between "similar" agents. As it happens, the optimality of stochastic extreme points is not limited to such cases. We next present an example where a stochastic extreme point is uniquely optimal. This stochastic extreme point "frequently" randomizes between "dissimilar" agents.

10. The reader may wonder whether one can prove sufficiency of (1) to (3) by viewing the set of DIC mechanisms as the set of solutions to a linear system of inequalities, checking for total unimodularity of the constraint matrix, and then invoking the Hoffman-Kruskal theorem (Korte and Vygen, 2018, Theorem 5.21). In the mechanism design literature, this approach is discussed in Pycia and Ünver (2015), for example. Here the approach works for the case where all type spaces are binary; our proof uses a result which can itself be derived from the Hoffman-Kruskal theorem. However, in the difficult case with three agents, the constraint matrix is not generally totally unimodular (see Section 3.C.3).

11. The argument is as follows. The set of DIC mechanisms is a polytope in Euclidean space that does not depend on the distribution. All extreme points of the polytope are exposed. Since all linear functionals on this polytope can be represented via some distribution, the claim follows. See Section 3.C.1 for the formalities.

12. In fact, in our model, stochastic extreme points cannot be implemented via any dominant-strategy equilibrium of any deterministic *indirect* mechanism. See Section 3.C.2. We note, however, a result of Rivera Mora (2022) implying the following (for our model): *Given an arbitrary DIC direct mechanism, there is an* ex-post *equilibrium of a deterministic indirect mechanism that implements the given DIC direct mechanism.* In this ex-post equilibrium, the agents play mixed strategies that emulate the randomization on the part of the given DIC mechanism. These mixed strategies do not generally form a dominant-strategy equilibrium.

3.5.2 An example of a stochastic extreme point

There are four agents, and their types are as follows:

$$\Theta_1 = \{\ell, r\}, \quad \Theta_2 = \{u, d\}, \quad \Theta_3 = \{f, c, b\}, \quad \Theta_4 = \{0\}. \tag{3.1}$$

Figure 3.1 shows (among other things that are not yet relevant) the type profiles of agents 1, 2, and 3; the degenerate type of agent 4 is omitted. The types of agents 1, 2, and 3 span a three-dimensional hyperrectangle. (Mnemonically, their types mean left, right, up, down, front, center, and back.) Each edge of the hyperrectangle represents a set of type profiles along which exactly one agent's type is changing. Hence DIC requires that the winning probability of this agent be constant along the edge. We identify such an edge by a pair (i, θ_{-i}), where i indicates the agent whose type is changing, and θ_{-i} indicates the fixed types of the others.

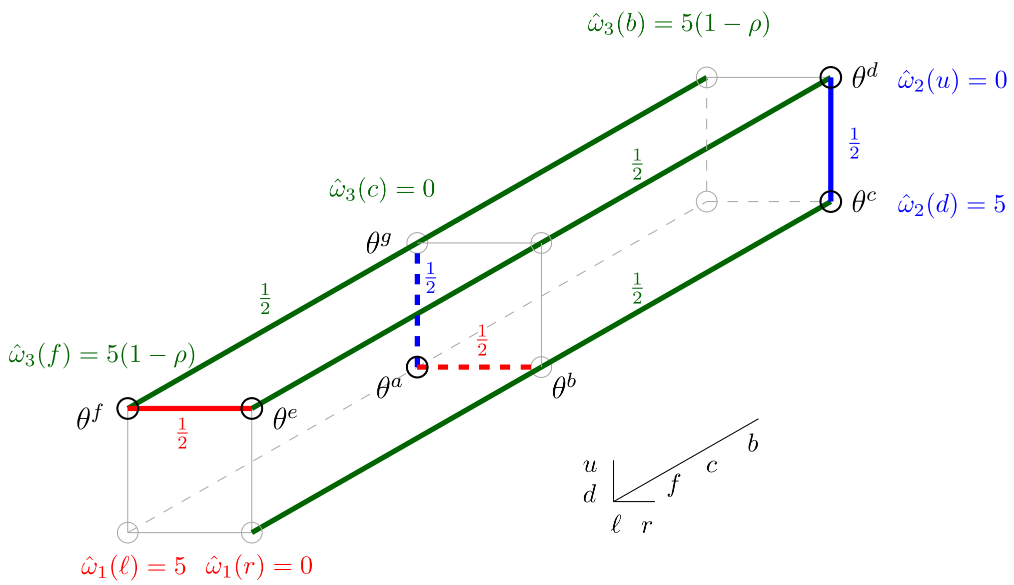


Figure 3.1. The set of types of agents 1, 2, and 3. The probabilities $\frac{1}{2}$ attached to the edges of the hyperrectangle represent the relevant values of the mechanism ϕ^*. The values from the allocation are as defined in (3.5). The distribution μ assigns probability $\frac{1}{5}$ to the profiles $\{\theta^a, \theta^c, \theta^d, \theta^e, \theta^f\}$. All other profiles have probability 0.

Let $\Theta^* = \{\theta^a, \theta^b, \theta^c, \theta^d, \theta^e, \theta^f, \theta^g\}$ be the set of labeled type profiles in Figure 3.1; these are the profiles

$$\begin{aligned} \theta^a &= (\ell, d, c, 0), & \theta^b &= (r, d, c, 0), & \theta^c &= (r, d, b, 0), \\ \theta^d &= (r, u, b, 0), & \theta^e &= (r, u, f, 0), & \theta^f &= (\ell, u, f, 0), \\ \theta^g &= (\ell, u, c, 0). \end{aligned} \tag{3.2}$$

Let V^* denote the set of bold edges in Figure 3.1 that connect the profiles in Θ^*; these are the edges

$$V^* = \{(1, \theta^a_{-1}), (3, \theta^c_{-3}), (2, \theta^c_{-2}), (3, \theta^e_{-3}), (1, \theta^e_{-1}), (3, \theta^f_{-3}), (2, \theta^a_{-2})\}.$$

Our candidate stochastic extreme point ϕ^* is defined as follows (see Figure 3.1): For all $i \in \{1,2,3\}$ and $\theta \in \Theta$, let

$$\phi_i^*(\theta) = \begin{cases} \frac{1}{2}, & \text{if } (i, \theta_{-i}) \in V^*, \\ 0, & \text{otherwise.} \end{cases}$$

Further, for all $\theta \in \Theta$ let $\phi_4^*(\theta) = 1 - \sum_{i \in \{1,2,3\}} \phi_i^*(\theta)$. In plain words, at all profiles in Θ^*, exactly two bold edges of the hyperrectangle intersect at the profile; the mechanism ϕ^* randomizes evenly between the two agents of these edges. All remaining probability mass is assigned to agent 4. It is easy to verify from Figure 3.1 that ϕ^* is a well-defined DIC mechanism.

Further below we specify values Ω and a distribution μ such that ϕ^* is the unique optimal DIC mechanism. This implies that ϕ^* is an extreme point of the set of DIC mechanisms. Since the proof for uniqueness is somewhat involved, we next present a simple self-contained argument showing that ϕ^* is an extreme point.

Let ϕ be a DIC mechanism that receives non-zero weight in a convex combination that equals ϕ^*. We show $\phi = \phi^*$. For all profiles $\theta \in \Theta^*$, there are exactly two agents i and j such that (i, θ_{-i}) and (j, θ_{-j}) both belong to V^*; these are the two bold edges of the hyperrectangle that intersect at θ. Hence at θ the mechanism ϕ^* randomizes evenly between i and j. Since ϕ is part of a convex combination that equals ϕ^*, it follows that at θ the mechanism ϕ only randomizes between i and j, meaning $\phi_i(\theta) = 1 - \phi_j(\theta)$. Since ϕ is DIC, repeatedly applying this observation shows:

$$\begin{aligned} \phi_1(\theta^a) = 1 - \phi_3(\theta^c) = \phi_2(\theta^c) &= 1 - \phi_3(\theta^e) \\ &= \phi_1(\theta^e) \\ &= 1 - \phi_3(\theta^f) = \phi_2(\theta^a) = 1 - \phi_1(\theta^a). \end{aligned} \tag{3.3}$$

In particular, we have $\phi_1(\theta^a) = 1 - \phi_1(\theta^a)$, implying $\phi_1(\theta^a) = \frac{1}{2}$. Hence all probabilities in (3.3) equal $\frac{1}{2}$. Hence ϕ agrees with ϕ^* at all profiles in Θ^*. By inspecting $\Theta \setminus \Theta^*$, we may easily convince ourselves that ϕ and ϕ^* also agree on $\Theta \setminus \Theta^*$. Thus ϕ^* is an extreme point.

We next construct an environment in which ϕ^* is uniquely optimal. We could do so by invoking a separating hyperplane theorem. However, this would be unsatisfying since we would gain no intuition for why randomization helps or for whether ϕ^* is uniquely optimal in a restricted class of environments. We shall gain both by considering environments in which values are *privately known*, in the following sense: for all agents i, the value of allocating to i is pinned down by a function $\hat{\omega}_i$ that depends only on θ_i.

We can describe an environment with privately known values by specifying a distribution μ over type profiles and, for all agents i, a function $\hat{\omega}_i \colon \Theta_i \to \mathbb{R}$ that governs the value of allocating to i. Our candidate distribution μ is given by (see Figure 3.1)

$$\forall_{\theta\in\Theta}, \quad \mu(\theta) = \begin{cases} \frac{1}{5}, & \text{if } \theta \in \{\theta^a, \theta^c, \theta^d, \theta^e, \theta^f\} \\ 0, & \text{else.} \end{cases} \tag{3.4}$$

Our candidates for $\hat{\omega}_1, \dots \hat{\omega}_4$ are parametrized by $\rho \in [0, \frac{1}{2}]$ and given by

$$\begin{aligned} \hat{\omega}_1(r) = \hat{\omega}_2(u) = \hat{\omega}_3(c) &= 0 \\ \hat{\omega}_1(\ell) = \hat{\omega}_2(d) &= 5 \\ \hat{\omega}_3(f) = \hat{\omega}_3(b) &= 5(1-\rho) \\ \hat{\omega}_4 &= 0. \end{aligned} \tag{3.5}$$

Proposition 3.1. *The mechanism ϕ^* is an optimal DIC mechanism if and only if $\rho \in [0, \frac{1}{2}]$, and it is uniquely optimal if and only if $\rho \in (0, \frac{1}{2})$.*

In the introduction, we intuited that there is a trade-off between allocating to an agent and using that agent's information about others. In the present example, this trade-off involves agent 3 and depends on ρ.

To gain an intuition for the trade-off and the result, consider the case $\rho = 0$. Allocating to agent 3 is now ex-post optimal at *all except one* of the five profiles in the support of μ. Indeed, one optimal DIC mechanisms is the constant one that always allocates to agent 3. The mechanism ϕ^* is another optimal mechanism for $\rho = 0$, which is intuitively explained by agent 3's type being informative: if $\theta_3 = c$ realizes, the type profile must be θ^a, where θ^a is the unique type profile in the support of μ at which allocating to agents 1 or 2 is better than allocating to agent 3. The mechanism ϕ^* indeed allocates to agents 1 and 2 at θ^a.

Since ρ decreases the value from allocating to agent 3, it is now intuitive that ϕ^* does strictly better than always allocating to agent 3 for small but strictly positive values of ρ. In the formal proof, most of our effort goes towards showing that ϕ^* is in fact uniquely optimal for small but strictly positive values of ρ. The idea is that, among all DIC mechanisms that are optimal for $\rho = 0$, the mechanism ϕ^* is the unique one minimizing agent 3's overall winning probability.

If we increase ρ further, it eventually becomes optimal to use agent 3 as a source of information and never allocate to agent 3. The critical value turns out to be $\rho = \frac{1}{2}$. The intuition is confirmed by the fact that, if $\rho = \frac{1}{2}$, the following *jury* mechanism with agent 3 as a juror is optimal: if agent 3 reports f, agent 1 wins; if agent 3 reports c, a coin flip determines whether agent 1 or 2 wins; if agent 3 reports b, agent 2 wins.

Proposition 3.1 also helps illustrate the commitment issue discussed in the paragraphs following Theorem 3.3. At the profile θ^e, a coin flip determines whether agent 1 or 3 wins the object. Yet, at this profile, the value from allocating to agent 3 is strictly higher than the value from allocating to agent 1. In fact, a coin is flipped at all type profiles in the support of the distribution. For $\rho \in (0, \frac{1}{2})$, the mechanism designer is indifferent to the outcome of the coin flip at only one of these profiles.

Remark 3.2. Chen et al. (2019) show that, in certain mechanism design problems, given any stochastic mechanism there is a deterministic one that induces the same interim-expected allocations. Since the deterministic mechanism is not guaranteed to be DIC, their result does not contradict the suboptimality of deterministic DIC mechanisms in our model.

Remark 3.3. An alternative approach to showing the existence of a stochastic extreme point uses a graph-theoretic result due to Chvátal (1975), as we explain in Section 3.B.2. For a certain graph G that we define in Section 3.B.2, Chvátal's theorem implies that all extreme points are deterministic if and only if G is perfect. To be precise, the results of this appendix concern the related problem where the mechanism may dispose the object instead of allocating it to the agents. The associated characterization of extreme points is implied by Theorem 3.3, but not vice versa.

3.6 Anonymous Juries

In this section, we study anonymous DIC mechanisms. Anonymity, formally defined below, is roughly the requirement that any two agents exert the same influence with their reports on the winning probability of any third agent. This is a desirable property as it helps protect the agents' privacy when they evaluate their peers, reduces the complexity of the mechanism, and ensures that the agents have the same voting rights.

We offer two insights. First, all anonymous DIC mechanisms ignore the reports of the agents. Second, we consider a relaxed notion of anonymity—*partial anonymity*—and show that all deterministic partially anonymous DIC mechanisms are jury mechanisms.

Throughout, we assume that the agents share a common type space, meaning $\Theta_1 = \ldots = \Theta_n$. In an equally valid interpretation, we can consider indirect mechanisms where all agents have the same message space and cannot influence their own winning probabilities.

3.6.1 Notions of anonymity

Anonymity and partial anonymity are defined next. Anonymity requires that, for all k, the winning probability of agent k does not change if one permutes the reports of the agents other than k. Partial anonymity relaxes anonymity as follows: When testing whether k's winning probability is affected by permutations, we only consider permutations of those agents who actually influence agent k. In particular, partial anonymity permits the set of agents who influence k to be a proper subset of $\{1, \ldots, n\} \setminus \{k\}$.

Definition 3.2. Let the agents have a common type space. Let ϕ be a mechanism.

(1) Given i, j, and k that are all distinct, agents i and j are *exchangeable for* k if ϕ_k is invariant with respect to permutations of i's and j's reports; that is, for all profiles θ and θ' such that θ is obtained from θ' by permuting the types of i and j we have $\phi_k(\theta) = \phi_k(\theta')$.
(2) Given distinct i and k, agent i *influences* k if ϕ_k is non-constant in i's report; that is, there exist type profiles θ and θ' that differ only in i's type and satisfy $\phi_k(\theta) \neq \phi_k(\theta')$.
(3) The mechanism is *anonymous* if for all i, j, and k that are all distinct, agents i and j are exchangeable for k.
(4) The mechanism is *partially anonymous* if for all i, j, and k that are all distinct we have the following: if i and j both influence k, then i and j are exchangeable for k.

To state the upcoming characterization of partial anonymity, we also define what we mean by an anonymous jury.

Definition 3.3. Let the agents have a common type space. A jury mechanism has an *anonymous jury* if all jurors i and j are exchangeable for all agents k.

Remark 3.4. If Assumption 3.1 holds, then among jury mechanisms it is without loss to use one with an anonymous jury. Indeed, consider the jury mechanism that selects the candidate that is best conditional on the types of the jurors (breaking ties in some fixed order). Under Assumption 3.1, the identity of the favored candidate does not change when one permutes the jurors' types.

3.6.2 Anonymous DIC mechanisms ignore all reports

Theorem 3.4. *Let the agents have a common type space. All anonymous DIC mechanisms are constant.*

Note well that anonymity does *not* demand that i and j be exchangeable for i's own winning probability. If anonymity did demand this, the theorem would follow rather trivially from DIC.

The theorem is more subtly related to the requirement that the mechanism always allocates the object, as we explain next. This requirement lets us link the influence that two agents i and j exert on others to the influence that they exert on each other.

More concretely, assume towards a contradiction that at some profile θ agent i can increase ϕ_j by changing their report from θ_i to some θ_i'. By DIC and since the object is always allocated, this change in i's report decreases $\sum_{k:\, i\neq k\neq j} \phi_k$. Now consider the profile that is obtained from θ by permuting the reports of i and j. By anonymity, agent j can change their report from θ_i to θ_i' to decrease $\sum_{k:\, i\neq k\neq j} \phi_k$.

Using again that the mechanism is DIC and that the object must be allocated, it follows that the change in agent j's report increases ϕ_i. In summary, if i can increase j's winning probability at some profile, then j must also be able to increase i's winning probability at a permuted profile. This observation suggests that i and j both win with "high" probability when both report θ_i'. In a deterministic mechanism, where winning probabilities are either 0 or 1, we thus arrive at a contradiction to there being only one object to allocate. We address stochastic mechanisms via a substantially more complex summation over winning probabilities across all pairs (i,j).

Remark 3.5. Theorem 3.4 implies that all DIC mechanisms satisfying the following stronger notion of anonymity are constant: Whenever the set of reports is permuted, then the same permutation is applied to the vector of winning probabilities. This stronger notion captures a sense in which agents are treated equally both as voters and winners.

Remark 3.6. An implication of Theorem 3.4 is that it is impossible to elicit information in environments where anonymity is without loss. Indeed, if the joint distribution of types and values is invariant with respect to all permutations of the agents, then it is without loss to use a DIC mechanism that satisfies the strong notion of anonymity from Remark 3.5. Hence in this case it is without loss to use a constant mechanism.

3.6.3 Partial anonymity

Theorem 3.4 implies that a non-constant DIC mechanism must admit some asymmetry in how it processes the reports of different agents. This brings us to partial anonymity. We offer the following characterization for deterministic mechanisms.

Theorem 3.5. *Let the agents have a common type space. A mechanism is deterministic, partially anonymous, and DIC if and only if it is a deterministic jury mechanism with an anonymous jury.*

To better understand the theorem, consider how a partially anonymous jury mechanism could fail to admit an anonymous jury. Given agents i and j, partial anonymity is silent on the winning probabilities of those agents k who are influenced by either i or j but not by both. By contrast, anonymity of the jury requires that all candidates are either influenced by all or none of the jurors. Accordingly, most of our effort goes towards proving that, in a deterministic partially anonymous DIC mechanism, if i and j influence *some* third agent k, then i and j influence exactly the same set of agents. Equipped with this fact, we show that the agents can be partitioned into equivalence classes with the following property: two agents in the same class do not influence one another, but influence the same (possibly empty) set of agents outside the class. Lastly, there cannot be multiple classes; indeed, else there is a profile where two distinct classes allocate the object to two distinct agents, which is impossible. The unique class defines an anonymous jury.

3.6.4 Discussion of Theorem 3.4 and Theorem 3.5

We conclude by discussing limitations of Theorem 3.4 and Theorem 3.5.

3.6.4.1 Disposal and randomization

The following definition will be useful: A *mechanism with disposal* is a function $\phi : \Theta \to [0,1]^n$ satisfying $\sum_{i=1}^n \phi_i \leq 1$. In plain words, this is a mechanism that does not necessarily always allocate the object to the agents. For a mechanism with disposal, DIC and anonymity are defined as above.

The next result shows via an example that Theorem 3.4 does not extend to mechanisms with disposal, and that Theorem 3.5 does not extend to stochastic mechanisms (without disposal).

Proposition 3.2. *Let the agents have a common type space T such that $|T| = 7$.*

(1) If $n = 3$, then the set of DIC mechanisms with disposal admits an extreme point that is stochastic and anonymous.

(2) If $n = 4$, then the set of DIC mechanisms (without disposal) admits an extreme point that is stochastic and partially anonymous.

The extreme point in (1) is non-constant (else it would be a convex combination of deterministic constant mechanisms). The extreme point in (2) is not a jury mechanism (else it would be a convex combination of deterministic jury mechanisms). The idea of the proof is to "symmetrize" the stochastic extreme point ϕ^* from Section 3.5.2. See Section 3.A.3.3 for the proof and an informal sketch.

3.6.4.2 Anonymous ballots

Lastly, we discuss the assumption that all agents can make the same reports. Indeed, a third escape route from Theorem 3.4 (besides partial anonymity and disposal) entails message spaces with some inherent asymmetry across agents. This brings us to the results of Holzman and Moulin (2013) and Mackenzie (2015, 2020). They consider DIC mechanisms where agents nominate one another. Let us keep with the terminology of Holzman and Moulin by referring to these mechanisms as *impartial nomination rules*. This is the same mathematical object as a DIC mechanism when each agent i's type space is $\{1, \ldots, n\} \setminus \{i\}$. Their notion of anonymity—*anonymous ballots*—requires that the winning probabilities depend only on the number of nominations received by each agent.[13] Importantly, in a nomination rule agents cannot nominate themselves, and hence they all have distinct message spaces. By contrast,

13. Equivalently, the allocation is unchanged if one permutes the profile in a way that does not yield self-nominations (Mackenzie, 2015, Lemma 1.1). Mackenzie uses the name *voter anonymity* instead of anonymous ballots.

we have assumed that the agents have the same type space. Hence our notion of anonymity neither nests nor is nested by anonymous ballots.

Contrasting Theorem 3.4, there are *non-constant* impartial nomination rules with anonymous ballots. For one example, suppose one of the agents is selected uniformly at random as a juror, following which the juror's nomination determines a winner. See Mackenzie (2015, Theorem 1) for a full characterization of anonymous ballots. Mackenzie's result generalizes Theorem 3 of Holzman and Moulin (2013), who had previously shown that all *deterministic* impartial nomination rules with anonymous ballots are constant.

Mackenzie (2020) shows that impartiality and anonymous ballots are compatible for deterministic nomination rules with disposal.[14] This parallels our discussion from Section 3.6.4 and contrasts the aforementioned Theorem 3 of Holzman and Moulin (2013). Mackenzie (2020, Theorem 1) also shows that when agents can nominate themselves, then deterministic impartial nomination rules with anonymous ballots must be constant. This is a special case of our Theorem 3.4 as anonymous ballots with self-nominations is stronger than anonymity.

3.7 Conclusion

We saw that jury mechanisms are optimal with three agents, and approximately-optimal when there are many exchangeable agents in the sense of Assumption 3.1. While DIC mechanisms cannot process all reports anonymously, jury mechanisms are the only deterministic partially anonymous DIC mechanisms. Lastly, outside of special cases of the model, the set of DIC mechanisms admits stochastic extreme points.

We conclude by discussing some interesting open problems.

The discussion on stochastic extreme points (Section 3.5.1) motivates restricting attention to deterministic mechanisms. We observe in Section 3.C.4 that finding an optimal deterministic DIC mechanism can be cast as the problem of finding a maximum weight perfect matching in a certain hypergraph. If we relax the requirement that the object is always allocated, the problem can also be cast as finding a maximum weight independent set in another graph. Both of these problems are known to be NP-hard when general (hyper-)graphs and weights are considered. As such, it is interesting to investigate the hardness of the problem for the particular family of (hyper-)graphs that emerge from our model. (All weights can emerge via a suitable choice of the distribution of types and values.) If we include stochastic mechanisms in our search, finding an optimal DIC mechanism is a linear program and hence computationally tractable.

14. In fact, Mackenzie (2020, Theorem 2) shows that impartiality, anonymous ballots, and some other desirable axioms together characterize supermajority.

It is naturally interesting to extend the analysis to settings with multiple objects, allocated simultaneously or over many periods.[15] If the mechanism designer can commit to future allocations, this should lead to stronger foundations for jury mechanisms. Agents serving as jurors today can be promised a future spot as candidates, which may help justify excluding jurors as potential winners in the present. Alternatively, past winners may be expected to volunteer as jurors in the future.

The problem of finding an optimal composition of the jury is an interesting problem in itself. We expect interesting comparative statics when agents who are likely to have good information are also likely to yield a high value. In the example from the introduction where a group selects a president, say, an agent who is popular with others may be a suitable candidate (being well-liked for their pleasant qualities) but also have good information about others (being well-acquainted with everyone).

An important line of future research concerns optimal DIC mechanisms when agents care about the allocation to their peers. While DIC has different implications in such a model, our results provide insight in at least two cases. Firstly, in situations where agents evaluate their peers, it is seems inherently interesting to use a mechanism where agents cannot influence their individual chances of winning; that is, to impose the *impartiality* axiom of Holzman and Moulin (2013). Secondly, suppose agents have the following lexicographic preferences: each agent i strictly prefers one allocation to another if the former has i winning with strictly higher probability; if two allocations have the same winning probability for i, agent i ranks them according to some type-dependent preference. In some applications, this preference could reasonably capture i's opinion about who is the most deserving winner if it cannot be i themself. In particular, it could coincide with the preference of the mechanism designer. In this case, optimal jury mechanism are ex-post incentive compatible. However, an agent's preferences may also differ from those of the designer. This is plausibly the case when agents are biased in favor of friends or family, biased against minorities, or simply have a different notion of who deserves to win.[16] Fixing a jury of agents, the designer therefore also has to design a voting rule for eliciting the jurors' information.

15. See Guo and Hörner (2021) for recent work in this direction with a single agent. The literature following Alon et al. (2011) has also studied settings with multiple objects. Lipnowski and Ramos (2020) and de Clippel et al. (2021) study settings with limited or no commitment.

16. For example, Alatas et al. (2012), reporting on a field experiment on selecting beneficiaries of aid programs in Indonesian communities, find evidence of nepotism, though the welfare impact may be small relative to other upsides from involving the community in the decision. They also find evidence that community members have a poverty notion that differs from poverty as defined by per capita income. In this sense, if the central government wishes to select beneficiaries on the basis of per capita income, agents indeed hold a different notion of who deserves to win.

Appendix 3.A Omitted Proofs

In Section 3.A.1, Section 3.A.2, and Section 3.A.3, we present the omitted proofs for Section 3.4, Section 3.5, and Section 3.6, respectively.

3.A.1 Jury mechanisms

3.A.1.1 Proof of Lemma 3.2

Proof of Lemma 3.2. If $n = 1$ or $n = 2$, it is easy to verify that all DIC mechanisms are constant. All constant mechanisms are convex combination of deterministic constant mechanisms, proving the claim. In what follows, let $n = 3$. Given an arbitrary stochastic DIC mechanism ϕ, we will find a non-zero function f such that $\phi + f$ and $\phi - f$ are two other DIC mechanisms. This shows that all extreme points of the set of DIC mechanisms are deterministic. Since this set is non-empty, convex and compact as a subset of Euclidean space, the claim follows from the Krein-Milman theorem.

In what follows, we fix a stochastic DIC mechanisms ϕ. Let us agree to the following terminology. In view of DIC, we drop i's type from ϕ_i. Given a profile θ, we refer to the equation $\sum_{i\in\{1,2,3\}} \phi_i(\theta_{-i}) = 1$ as the *feasibility constraint* at profile θ. We refer to (i, θ_{-i}) as the *node of agent i with coordinates* θ_{-i}. Lastly, when we say $\phi_i(\theta_{-i})$ is interior we naturally mean $\phi_i(\theta_{-i}) \in (0, 1)$.

Most of the work will go towards proving the following auxiliary claim.

Claim 3.1. *There are non-empty disjoint subsets R and B ("red" and "blue") of* $\cup_{i\in\{1,2,3\}}(\{i\} \times \Theta_{-i})$ *such that all of the following are true:*

(1) If $(i, \theta_{-i}) \in R \cup B$*, then* $\phi_i(\theta_{-i})$ *is interior.*

(2) For all $\theta \in \Theta$*, exactly one of the following is true:*

   a. There does not exist $i \in \{1, 2, 3\}$ *such that* $(i, \theta_{-i}) \in R \cup B$.

   b. There exists exactly one $i \in \{1, 2, 3\}$ *such that* $(i, \theta_{-i}) \in R$*, exactly one* $j \in \{1, 2, 3\}$ *such that* $(j, \theta_{-j}) \in B$*, and exactly one* $k \in \{1, 2, 3\}$ *such that* $(k, \theta_{-k}) \notin R \cup B$.

Before proving Claim 3.1, let us use it to complete the proof of Lemma 3.2. For a number ε to be chosen in a moment, let $f : \Theta \to \{-\varepsilon, 0, \varepsilon\}^3$ be defined as follows:

$$\forall_{\theta\in\Theta}, \quad f_i(\theta) = \begin{cases} -\varepsilon, & \text{if } (i, \theta_{-i}) \in R, \\ \varepsilon, & \text{if } (i, \theta_{-i}) \in B, \\ 0, & \text{if } (i, \theta_{-i}) \notin R \cup B. \end{cases}$$

By finiteness of Θ and Claim 3.1, if we choose $\varepsilon > 0$ sufficiently close to 0, then $\phi + f$ and $\phi - f$ are two DIC mechanisms. Since f is non-zero, it follows that ϕ is not an extreme point. It remains to prove Claim 3.1.

Proof of Claim 3.1. Given candidate sets R and B, let us say a profile θ is *uncolored* if it falls into case (2.a) of Claim 3.1. A profile *two-colored* if it falls into case (2.a) of Claim 3.1. In this terminology, our goal is to construct sets R and B such that all $(i, \theta_{-i}) \in R \cup B$ satisfy $\phi_i(\theta_{-i}) \in (0, 1)$, and such that all type profiles are either uncolored or two-colored.

Since ϕ is stochastic, we may assume (after possibly relabelling the agents and types) that there exists a profile θ^0 such that $\phi_1(\theta_2^0, \theta_3^0)$ and $\phi_2(\theta_1^0, \theta_3^0)$ are interior.

Let Θ_2° denote the set of types θ_2 for which $\phi_1(\theta_2, \theta_3^0)$ is interior. Let $\Theta_2^\partial = \Theta_2 \setminus \Theta_2^\circ$. Similarly, let Θ_1° denote the set of types θ_1 such that $\phi_2(\theta_1, \theta_3^0)$ is interior, and let $\Theta_1^\partial = \Theta_1 \setminus \Theta_1^\circ$. Notice that Θ_1° and Θ_2° are non-empty as, by assumption, agents 1 and 2 are enjoying interior winning probabilities at θ^0.

We consider two cases.

Case 1. Let $\Theta_1^\partial \neq \emptyset$ and $\Theta_2^\partial \neq \emptyset$.

We establish two auxiliary claims.

Claim 3.2. *If $\theta_1 \in \Theta_1^\partial$, then $\phi_2(\theta_1, \theta_3^0) = 0$. Similarly, if $\theta_2 \in \Theta_2^\partial$, then $\phi_1(\theta_2, \theta_3^0) = 0$. If $(\theta_1, \theta_2) \in (\Theta_1^\circ \times \Theta_2^\partial) \cup (\Theta_1^\partial \times \Theta_1^\circ)$, then $\phi_3(\theta_1, \theta_2)$ is interior.*

Proof of Claim 3.2. Consider the first part of the claim. Let $\theta_1 \in \Theta_1^\partial$. Recalling that Θ_1° is non-empty, let us find a type $\theta_2 \in \Theta_1^\circ$. By definition, $\phi_1(\theta_2, \theta_3^0)$ is interior. By definition of Θ_1^∂, we also know that $\phi_2(\theta_1, \theta_3^0)$ must either equal 0 or 1. But it cannot equal 1 since $\phi_2(\theta_1, \theta_3^0)$ and $\phi_1(\theta_2, \theta_3^0)$ both appear in the feasibility constraint at the profile $(\theta_1, \theta_2, \theta_3^0)$, and since $\phi_1(\theta_2, \theta_3^0)$ is interior. Thus $\phi_2(\theta_1, \theta_3^0) = 0$, as desired.

A similar argument establishes the second claim.

As for the third claim, let $(\theta_1, \theta_2) \in \Theta_1^\circ \times \Theta_2^\partial$. The previous two paragraphs imply that at the profile $(\theta_1, \theta_2, \theta_3^0)$ the winning probability of agent 1 is 0. Moreover, by definition of Θ_1°, the winning probabiltiy of agent 2 is interior. Thus agent 3's winning probability at this profile must be interior, meaning $\phi_3(\theta_1, \theta_2)$ is interior. A similar argument shows that $\phi_3(\theta_1, \theta_2)$ is interior whenever (θ_1, θ_2) is in $\Theta_1^\partial \times \Theta_1^\circ$. □

The second auxiliary result is:

Claim 3.3. *Let $\theta_3 \in \Theta_3$. If $\theta_2 \in \Theta_2^\circ$, then $\phi_1(\theta_2, \theta_3)$ is interior. Similarly, if $\theta_1 \in \Theta_1^\circ$, then $\phi_2(\theta_1, \theta_3)$ is interior.*

Proof of Claim 3.3. We will prove the first part of the claim, the second being similar. Thus let $\theta_2 \in \Theta_2^\circ$. By assumption of Case 1, we may find $\theta_1^\partial \in \Theta_1^\partial$ and $\theta_2^\partial \in \Theta_2^\partial$. We make two auxiliary observations.

First, consider the profile $(\theta_1^\partial, \theta_2^\partial, \theta_3^0)$. According to Claim 3.2, both agent 1's and agent 2's winning probabilities at this profile equal 0. Thus $\phi_3(\theta_1^\partial, \theta_2^\partial) = 1$. But $\phi_3(\theta_1^\partial, \theta_2^\partial)$ and $\phi_2(\theta_1^\partial, \theta_3)$ both appear in the feasibility constraint at the profile $(\theta_1^\partial, \theta_2^\partial, \theta_3)$. Hence $\phi_2(\theta_1^\partial, \theta_3) = 0$.

Second, since $\theta_1^\partial \in \Theta_1^\partial$ and $\theta_2 \in \Theta_2^\circ$, we infer from Claim 3.2 that $\phi_3(\theta_1^\partial, \theta_2)$ is interior.

The previous two observations imply that at the profile $(\theta_1^\partial, \theta_2, \theta_3)$ agent 2's winning probability is 0 and that agent 3's winning probability is interior. Hence $\phi_1(\theta_2, \theta_3)$ is interior, as promised. □

We are ready to define the sets R and B. We assign the following colors (recall the terminology introduced in the paragraph before Claim 3.1):

- red to all nodes of agent 1 with coordinates in $\Theta_2^\circ \times \Theta_3$,
- blue to all nodes of agent 3 with coordinates in $\Theta_1^\partial \times \Theta_2^\circ$,
- blue to all nodes of agent 2 with coordinates in $\Theta_1^\circ \times \Theta_3$,
- red to all nodes of agent 3 with coordinates in $\Theta_1^\circ \times \Theta_2^\partial$.

According to Claim 3.2 and Claim 3.3, all of these nodes are interior. Moreover, all profiles are now either two-colored or uncolored: The profiles in $\Theta_1^\partial \times \Theta_2^\circ \times \Theta_3$ are two-colored via red nodes of agent 1 and blue nodes of agent 3; the profiles in $\Theta_1^\circ \times \Theta_2^\circ \times \Theta_3$ are two-colored via red nodes of agent 1 and blue nodes of agent 2; the profiles in $\Theta_1^\circ \times \Theta_2^\partial \times \Theta_3$ are two-colored via blue nodes of agent 2 and red nodes of 3; and the profiles in $\Theta_1^\partial \times \Theta_2^\partial \times \Theta_3$ are uncolored. ▲

Case 2. Suppose at least one of the sets Θ_1^∂ and Θ_2^∂ is empty. In what follows, we assume that Θ_2^∂ is empty, the other case being analogous (switch the roles of agents 1 and 2).

The assumption that Θ_2^∂ is empty means that $\phi_1(\theta_2, \theta_3^0)$ is interior for all θ_2. Let Θ_1^* be the set of types θ_1 such that for all $\theta_2 \in \Theta_2$ the probability $\phi_3(\theta_1, \theta_2)$ is interior. Notice that at this point Θ_1^* may or may not be empty; we will make a case distinction further below.

We first claim that if $\theta_1 \in (\Theta_1 \setminus \Theta_1^*)$, then $\phi_2(\theta_1, \theta_3^0)$ is interior. Towards a contradiction, suppose this were false for some $\theta_1 \in (\Theta_1 \setminus \Theta_1^*)$. This means that we can find a type $\theta_2 \in \Theta_2$ such that $\phi_2(\theta_1, \theta_3^0)$ and $\phi_3(\theta_1, \theta_2)$ both fail to be interior. Recall from the previous paragraph that $\phi_1(\theta_2, \theta_3^0)$ is interior for all θ_2. Hence at the profile $(\theta_1, \theta_2, \theta_3^0)$ only agent 1 is enjoying an interior winning probability; this is impossible.

Before proceeding further, let us assign the following colors:

- red to all nodes of agent 1 with coordinates in $\Theta_2 \times \{\theta_3^0\}$. These nodes are all interior since Θ_2^∂ is empty.
- blue to all nodes of agent 2 with coordinates in $(\Theta_1 \setminus \Theta_1^*) \times \{\theta_3^0\}$. The previous paragraph implies that these nodes are all interior.
- blue to all nodes of agent 3 with coordinates in $\Theta_1^* \times \Theta_2$. These nodes are all interior by definition of Θ_1^*.

Observe that all profiles in $\Theta_1 \times \Theta_2 \times \{\theta_3^0\}$ are now either two-colored or uncolored.

If Θ_1^* is empty, then the colors assigned above already define sets R and B with the desired properties, completing the proof. Thus suppose Θ_1^* is non-empty.

Let $\theta_3 \in \Theta_3 \setminus \{\theta_3^0\}$ be arbitrary. The fact that we have already assigned blue to the nodes of agent 3 with coordinates $\Theta_1^* \times \Theta_2$ requires us to assign some colors to the nodes of agents 1 or 2 whose 3'rd coordinate is θ_3. In this step, we will not color any further nodes of agent 3. We make a case distinction.

(1) Suppose that for all θ_1 in Θ_1^* the probability $\phi_2(\theta_1, \theta_3)$ is interior. We assign red to all nodes of agent 2 with coordinates in $\Theta_1^* \times \{\theta_3\}$. This yields a coloring of the profiles in $\Theta_1 \times \Theta_2 \times \{\theta_3^0\}$ with the desired properties: The profiles in $\Theta_1^* \times \Theta_2 \times \{\theta_3\}$ are two-colored via red nodes of agent 2 and blue nodes of 3; the profiles in $(\Theta_1 \setminus \Theta_1^*) \times \Theta_2 \times \{\theta_3\}$ are uncolored.

(2) Suppose there exists $\tilde{\theta}_1 \in \Theta_1^*$ such that $\phi_2(\theta_1, \theta_3)$ is interior. Given that $\phi_3(\tilde{\theta}_1, \theta_2)$ is interior for all $\theta_2 \in \Theta_2$ (recall the definition of Θ_1^*), it must be the case that, for all $\theta_2 \in \Theta_2$, the probability $\phi_1(\theta_2, \theta_3)$ is interior.

We next claim that $\phi_2(\theta_1, \theta_3)$ is interior for all $\theta_1 \in (\Theta_1 \setminus \Theta_1^*)$. Suppose this were false for some $\theta_1 \in (\Theta_1 \setminus \Theta_1^*)$. The previous paragraph tells us that $\phi_1(\theta_2, \theta_3)$ is interior for all θ_2. Thus, if $\phi_2(\theta_1, \theta_3)$ fails to be interior, then $\phi_3(\theta_1, \theta_2)$ would have to be interior for all $\theta_2 \in \Theta_2$; this is a contradiction since $\theta_1 \in (\Theta_1 \setminus \Theta_1^*)$.

We now assign red to all nodes of agent 1 with coordinates in $\Theta_2 \times \{\theta_3\}$, and assign blue to all nodes of agent 2 with coordinates in $(\Theta_1 \setminus \Theta_1^*) \times \{\theta_3\}$. The previous two paragraphs imply that all of these nodes are interior. Moreover the profiles in $\Theta_1^* \times \Theta_2 \times \{\theta_3\}$ are two-colored via red nodes of agent 1 and blue nodes of agent 3, and the profiles in $(\Theta_1 \setminus \Theta_1^*) \times \Theta_2 \times \{\theta_3\}$ are two-colored via red nodes of agent 1 and blue nodes of agent 2.

If we apply this case distinction separately to all θ_3 in $\Theta_3 \setminus \{\theta_3^0\}$, this completes the construction of R and B in Case 2. ▲

Case 1 and Case 2 together complete the proof of Claim 3.1. □

□

3.A.1.2 Approximate optimality of jury mechanisms

In this part of the appendix, we prove Theorem 3.2. To distinguish a random variable from its realization, we denote the former using a tilde $\sim$. Given a set N of agents, we denote the profile of their types by θ_N, and the set of these profiles by Θ_N. For example, given $i \in N$, $\omega_i \in \Omega_i$, and $\theta_{N\setminus\{i\}} \in \Theta_{N\setminus\{i\}}$, we write $\mu\left(\tilde{\omega}_i = \omega_i, \tilde{\theta}_{N\setminus\{i\}} = \theta_{N\setminus\{i\}}\right)$ to mean the probability of the event that i's value is ω_i and the types of the other agents in N are $\theta_{N\setminus\{i\}}$.

Assumption 3.2. For all $n \in \mathbb{N}$, there exists $m \in \mathbb{N}$ with the following property: Denoting $N = \{1, \dots, n\}$ and $N' = \{n+1, \dots, n+m\}$, there is a function $g\colon \Theta_{N'} \times \Theta_N \to \mathbb{R}_+$ with the following two properties:

(1) For all $i \in N$, all $\omega_i \in \Omega_i$ and $\theta_{N\setminus\{i\}} \in \Theta_{N\setminus\{i\}}$ we have

$$\begin{aligned} &\mu\big(\tilde{\omega}_i = \omega_i, \tilde{\theta}_{N\setminus\{i\}} = \theta_{N\setminus\{i\}}\big) \\ &= \sum_{\theta_{N'} \in \Theta_{N'}} \sum_{\theta_i \in \Theta_i} g(\theta_{N'}, \theta_{N\setminus\{i\}}, \theta_i) \mu\big(\tilde{\omega}_i = \omega_i, \tilde{\theta}_{N'} = \theta_{N'}\big). \end{aligned} \tag{3.A.1}$$

(2) For all $\theta_{N'} \in \Theta_{N'}$ we have

$$\sum_{\theta_N \in \Theta_N} g(\theta_{N'}, \theta_N) = 1. \tag{3.A.2}$$

Lemma 3.3. *Assumption 3.1 implies Assumption 3.2.*

Proof of Lemma 3.3. Let $m = n$. Let $N = \{1, \dots, n\}$ and $N' = \{n+1, \dots, 2n\}$, and let $\xi\colon N \to N'$ be a bijection. It is straightforward to verify that the function g defined as follows has the desired properties: For all $(\theta_N, \theta_{N'})$, let $g(\theta_N, \theta_{N'}) = 1$ if for all $i \in N$ the types of i and $\xi(i)$ agree; else, let $g(\theta_N, \theta_{N'}) = 0$. □

Proof of Theorem 3.2. The second part of the claim is immediate from the first. For the first part, let ϕ be an arbitrary DIC mechanism with n agents. Let $N = \{1, \dots, n\}$. For this choice of N, we invoke Lemma 3.3 to find m and g as in Assumption 3.2. Let $N' = \{n+1, \dots, n+m\}$. We define our candidate jury mechanism as follows: For all $i \in N$, let $\psi_i\colon \Theta_{N'} \to \mathbb{R}^n$ be defined by

$$\forall_{\theta_{N'} \in \Theta_{N'}}, \quad \psi_i(\theta_{N'}) = \sum_{\theta_N \in \Theta_N} g(\theta_{N'}, \theta_N) \phi_i^*(\theta_{N\setminus\{i\}}).$$

For all $i \in N'$, let $\psi_i = 0$. Let $\psi = (\psi_1, \dots, \psi_m)$.

Notice that ψ only depends on the reports of agents in N'. Since N' is disjoint from N, we can show that ψ is a jury mechanism in the setting with $n+m$ agents by showing that ψ maps to probability distributions over N. It is clear that ϕ is non-negative (as g and ψ^* are non-negative). To verify that ψ almost surely allocates to an agent in N, we observe that for all profiles $\theta_{N'}$ we have the following (the first equality is by definition of ψ; the second is from the fact that ϕ^* is a well-defined mechanism when the set of agents is N; the third is from (3.A.2)):

$$\sum_{i \in N} \psi_i(\theta_{N'}) = \sum_{i \in N} \sum_{\theta_N \in \Theta_N} g(\theta_{N'}, \theta_N) \phi_i^*(\theta_{N\setminus\{i\}}) = \sum_{\theta_N \in \Theta_N} g(\theta_{N'}, \theta_N) = 1,$$

as desired. We complete the proof by verifying that ϕ and ψ lead to the same expected value. We write the expected value from ϕ as follows (the first equality follows from (3.A.1); the remaining equalities obtain by rearranging):

$$\begin{aligned}
&\sum_{i\in N}\sum_{\theta_{N\setminus\{i\}}}\sum_{\omega_i}\omega_i\mu\big(\tilde{\omega}_i = \omega_i, \tilde{\theta}_{N-i} = \theta_{N\setminus\{i\}}\big)\,\phi_i^*(\theta_{N\setminus\{i\}})\\
=&\sum_{i\in N}\sum_{\theta_{N\setminus\{i\}}}\sum_{\omega_i}\omega_i\sum_{\theta_{N'}}\sum_{\theta_i} g(\theta_{N'},\theta_{N\setminus\{i\}},\theta_i)\mu\big(\tilde{\omega}_i = \omega_i, \tilde{\theta}_{N'} = \theta_{N'}\big)\,\phi_i^*(\theta_{N\setminus\{i\}})\\
=&\sum_{i\in N}\sum_{\omega_i}\sum_{\theta_{N'}}\omega_i\mu\big(\tilde{\omega}_i = \omega_i, \tilde{\theta}_{N'} = \theta_{N'}\big)\sum_{\theta_{N\setminus\{i\}}}\sum_{\theta_i} g(\theta_{N'},\theta_{N\setminus\{i\}},\theta_i)\phi_i^*(\theta_{N\setminus\{i\}})\\
=&\sum_{i\in N}\sum_{\omega_i}\sum_{\theta_{N'}}\omega_i\mu\big(\tilde{\omega}_i = \omega_i, \tilde{\theta}_{N'} = \theta_{N'}\big)\psi_i(\theta_{N'}).
\end{aligned}$$

This last expression is precisely the expected value from ψ. □

3.A.2 Random allocations

3.A.2.1 Proof of Proposition 3.1

Proof of Proposition 3.1. To keep calculations readable, it will be convient to adopt the following notation: When a DIC mechanism ϕ is given, we denote

$$\begin{aligned}
&\phi_1(\theta^a) = p^{a|b}, \quad \phi_3(\theta^c) = p^{b|c}, \quad \phi_2(\theta^c) = p^{c|d}, \quad \phi_3(\theta^e) = p^{d|e},\\
&\phi_1(\theta^e) = p^{e|f}, \quad \phi_3(\theta^f) = p^{f|g}, \quad \phi_2(\theta^a) = p^{g|a}.
\end{aligned}$$

The probabilities in the previous display do not fully describe the mechanism, but these are the only ones needed to evaluate the mechanism. For a given value of ρ, we denote the expected value from ϕ by $V_\rho(\phi)$. Direct computation shows

$$V_\rho(\phi) = p^{a|b} + p^{b|c} + p^{c|d} + 2p^{d|e} + p^{e|f} + p^{f|g} + p^{g|a} - \rho\left(p^{b|c} + 2p^{d|e} + p^{f|g}\right). \tag{3.A.3}$$

In particular, $V_\rho(\phi^*) = 4 - 2\rho$.

We first show that ϕ^* is uniquely optimal if $\rho \in (0, \frac{1}{2})$. The following auxiliary claim is central.

Claim 3.4. *Let ϕ be a DIC mechanism distinct from ϕ^*. We have $V_{\frac{1}{2}}(\phi) \le V_{\frac{1}{2}}(\phi^*)$. Further, there exists $\rho_\phi \in (0, \frac{1}{2})$ such that $\rho \in (0, \rho_\phi)$ implies $V_\rho(\phi) < V_\rho(\phi^*)$.*

Proof of Claim 3.4. Inspection of Figure 3.1 shows that ϕ must satisfy the following system of inequalities:

$$\begin{aligned}
&p^{a|b} + p^{g|a} \le 1, \quad p^{a|b} + p^{b|c} \le 1, \quad p^{c|d} + p^{b|c} \le 1, \quad p^{c|d} + p^{d|e} \le 1,\\
&p^{e|f} + p^{d|e} \le 1, \quad p^{e|f} + p^{f|g} \le 1, \quad p^{g|a} + p^{f|g} \le 1.
\end{aligned} \tag{3.A.4}$$

Turning to the first part of the claim, we have to show $V_{\frac{1}{2}}(\phi) \leq V_{\frac{1}{2}}(\phi^*)$. Direct computation shows $V_{\frac{1}{2}}(\phi^*) = 3$. Using (3.A.4), we can bound $V_{\frac{1}{2}}(\phi)$ as follows.

$$\begin{aligned}
V_{\frac{1}{2}}(\phi) &= p^{a|b} + p^{b|c} + p^{c|d} + 2p^{d|e} + p^{e|f} + p^{f|g} + p^{g|a} - \frac{1}{2}\left(p^{b|c} + 2p^{d|e} + p^{f|g}\right) \\
&= p^{a|b} + \frac{p^{b|c}}{2} + p^{c|d} + p^{d|e} + p^{e|f} + \frac{p^{f|g}}{2} + p^{g|a} \\
&= \underbrace{p^{a|b} + p^{g|a}}_{\leq 1} + \underbrace{\frac{p^{b|c} + p^{c|d}}{2}}_{\leq \frac{1}{2}} + \underbrace{\frac{p^{c|d} + p^{d|e}}{2}}_{\leq \frac{1}{2}} + \underbrace{\frac{p^{d|e} + p^{e|f}}{2}}_{\leq \frac{1}{2}} + \underbrace{\frac{p^{e|f} + p^{f|g}}{2}}_{\leq \frac{1}{2}} \\
&\leq 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} \\
&= 3.
\end{aligned}$$

Hence $V_{\frac{1}{2}}(\phi) \leq V_{\frac{1}{2}}(\phi^*)$, as promised.

Now consider the second part of the claim. We show the contrapositive: If there exists a sequence $\{\rho_k\}_{k\in\mathbb{N}}$ in $(0, \frac{1}{2})$ that converges to 0 and such that $V_{\rho_k}(\phi) \geq V_{\rho_k}(\phi^*)$ holds for all k, then $\phi = \phi^*$. Let $\{\rho_k\}_{k\in\mathbb{N}}$ be such a sequence. For all ρ_k, the system (3.A.4) implies the following upper bound on $V_{\rho_k}(\phi)$:

$$\begin{aligned}
V_{\rho_k}(\phi) &= \underbrace{p^{a|b} + p^{b|c}}_{\leq 1} + \underbrace{p^{c|d} + p^{d|e}}_{\leq 1} + \underbrace{p^{d|e} + p^{e|f}}_{\leq 1} + \underbrace{p^{f|g} + p^{g|a}}_{\leq 1} \\
&\quad - \rho_k\left(p^{b|c} + 2p^{d|e} + p^{f|g}\right) \\
&\leq 4 - \rho_k\left(p^{b|c} + 2p^{d|e} + p^{f|g}\right).
\end{aligned} \tag{3.A.5}$$

Since $V_{\rho_k}(\phi) \geq V_{\rho_k}(\phi^*) = 4 - 2\rho_k$ and $\rho_k > 0$, we find

$$p^{b|c} + 2p^{d|e} + p^{f|g} \ \leq \ 2. \tag{3.A.6}$$

Further, since $V_{\rho_k}(\phi) \geq 4 - 2\rho_k$ holds for all k, taking limits implies $V_0(\phi) \geq 4$. Together with the bound in (3.A.5) we get $V_0(\phi) = 4$; that is,

$$V_0(\phi) \ = \ p^{a|b} + p^{b|c} + p^{c|d} + p^{d|e} + p^{d|e} + p^{e|f} + p^{f|g} + p^{g|a} \ = \ 4 \tag{3.A.7}$$

Hence (3.A.4) and (3.A.7) imply

$$p^{a|b} + p^{b|c} \ = \ p^{c|d} + p^{d|e} \ = \ p^{d|e} + p^{e|f} \ = \ p^{f|g} + p^{g|a} \ = \ 1. \tag{3.A.8}$$

We now bound $V_0(\phi)$ a second time (the equality is by direct computation; the inequality follows from (3.A.4)):

$$V_0(\phi) \ = \ p^{a|b} + p^{g|a} + p^{b|c} + p^{c|d} + 2p^{d|e} + p^{e|f} + p^{f|g} \ \leq \ 3 + 2p^{d|e}. \tag{3.A.9}$$

Hence $V_0(\phi) = 4$ implies $p^{d|e} \geq \frac{1}{2}$. We next claim $p^{d|e} = \frac{1}{2}$. Towards a contradiction, suppose not, meaning $p^{d|e} > \frac{1}{2}$. Hence (3.A.8) implies $p^{c|d} = p^{e|f} < \frac{1}{2}$. Now, we also know from (3.A.6) and (3.A.7) that

$$p^{a|b} + p^{c|d} + p^{e|f} + p^{g|a} \ \geq \ 2$$

holds. However, in light of (3.A.4) we have $p^{a|b} + p^{g|a} \leq 1$, and hence the previous display requires $p^{c|d} + p^{e|f} \geq 1$. This contradicts $p^{c|d} = p^{e|f} < \frac{1}{2}$. Thus $p^{d|e} = \frac{1}{2}$.

Let us now return to the bound derived in (3.A.9). In view of $p^{d|e} = \frac{1}{2}$ and (3.A.4), we can infer from (3.A.9) that $p^{a|b} + p^{g|a} = p^{b|c} + p^{c|d} = p^{e|f} + p^{f|g} = 2p^{d|e} = 1$ holds. Together with (3.A.8), we find

$$p^{a|b} \ = \ 1 - p^{b|c} \ = \ p^{c|d} \ = \ 1 - p^{d|e} \ = \ p^{e|f} \ = \ 1 - p^{f|g} \ = \ p^{g|a}. \tag{3.A.10}$$

We already know that $p^{d|e} = \frac{1}{2}$ holds. Hence all probabilities (3.A.10) must equal $\frac{1}{2}$. This shows that ϕ agrees with ϕ^* at all profiles in $\Theta^* = \{\theta^a, \theta^b, \theta^c, \theta^d, \theta^e, \theta^f, \theta^g\}$. By inspecting $\Theta \setminus \Theta^*$, it is now easy to verify that ϕ and ϕ^* also agree on $\Theta \setminus \Theta^*$. □

We next use Claim 3.4 to show that ϕ^* is uniquely optimal if $\rho \in (0, \frac{1}{2})$. Let ϕ be an arbitrary DIC mechanisms distinct from ϕ^*. Inspection of (3.A.3) shows that the difference $V_\rho(\phi) - V_\rho(\phi^*)$ is an affine function of ρ. That is, there exist reals a_ϕ and b_ϕ such that $V_\rho(\phi) - V_\rho(\phi^*) = a_\phi + b_\phi \rho$ holds for all $\rho \in [0, \frac{1}{2}]$. Let $\rho_\phi \in (0, \frac{1}{2})$ be as in the conclusion of Claim 3.4. If $\rho \in (0, \rho_\phi)$, the choice of ρ_ϕ implies $V_\rho(\phi) < V_\rho(\phi^*)$, and so we are done. Hence in what follows we assume $\rho \in [\rho_\phi, \frac{1}{2})$. We distinguish two cases.

(1) If $b_\phi \leq 0$, then

$$V_\rho(\phi) - V_\rho(\phi^*) \ = \ a_\phi + b_\phi \rho \ \leq \ a_\phi + b_\phi \frac{\rho_\phi}{2} \ = \ V_{\frac{\rho_\phi}{2}}(\phi) - V_{\frac{\rho_\phi}{2}}(\phi^*).$$

Now $\frac{\rho_\phi}{2} \in (0, \rho_\phi)$ and the choice of ρ_ϕ imply $V_{\frac{\rho_\phi}{2}}(\phi) - V_{\frac{\rho_\phi}{2}}(\phi^*) < 0$, and we are done.

(2) If $b_\phi > 0$, then

$$V_\rho(\phi) - V_\rho(\phi^*) \ = \ a_\phi + b_\phi \rho \ < \ a_\phi + b_\phi \frac{1}{2} \ = \ V_{\frac{1}{2}}(\phi) - V_{\frac{1}{2}}(\phi^*).$$

Now Claim 3.4 implies $V_{\frac{1}{2}}(\phi) - V_{\frac{1}{2}}(\phi^*) \leq 0$, and we are done.

Hence all $\rho \in (0, \frac{1}{2})$ and all DIC mechanisms ϕ distinct from ϕ^* satisfy $V_\rho(\phi) < V_\rho(\phi^*)$.

It remains to show that ϕ^* is not uniquely optimal if $\rho \in \{0, \frac{1}{2}\}$, and that ϕ^* is not optimal if $\rho \notin [0, \frac{1}{2}]$. To that end, recall the constant mechanism and the jury mechanism described in the paragraphs after Proposition 3.1. By direct computation one can show that the constant mechanism or the jury mechanism, respectively, generate the same expected value as ϕ^* if $\rho = 0$ or $\rho = \frac{1}{2}$, respectively. Thus ϕ^* is not uniquely optimal if $\rho \in \{0, \frac{1}{2}\}$. Since ϕ^* is uniquely optimal on $(0, \frac{1}{2})$, and since the expected value is affine in ρ, we conclude that ϕ^* is not optimal if $\rho \notin [0, \frac{1}{2}]$. □

3.A.2.2 Proof of Theorem 3.3

Lemma 3.4. *If for all agents i we have $|\Theta_i| \leq 2$, then all extreme points of the set of DIC mechanisms are deterministic.*

For the proof, recall the following definitions for a given (simple undirected) graph G with node set V and edge set E. Given a node v, the set of edges which are incident to v is denoted $E(v)$. A *perfect matching* is a function $\psi: E \to \{0,1\}$ such that all $v \in V$ satisfy $\sum_{e \in E(v)} \psi(e) = 1$. The *perfect matching polytope* is the set $\{\psi: E \to [0,1]: \forall_{v \in V}, \sum_{e \in E(v)} \psi(e) = 1\}$.

Proof of Lemma 3.4. Let us relabel types such that we have $\Theta_i \subseteq \{0,1\}$ for all i. First, suppose we have $\Theta_i = \{0,1\}$ for all i.

For all DIC mechanisms ϕ, all agents i and all profiles θ, we may drop i's report from i's winning probability, writing $\phi_i(\theta_{-i})$ instead of $\phi_i(\theta)$. Under this convention, we claim that the set of DIC mechanisms is the perfect matching polytope of the graph G that has node set $\{0,1\}^n$ and where two nodes are adjacent if and only if they differ in exactly one coordinate. (This graph is known as the *n-hypercube*.) Indeed, each node of the graph is a type profile θ, and each edge may be identified with a pair of the form (i, θ_{-i}). The set of edges incident to θ is the set $\{(i, \theta_{-i})\}_{i=1}^n$. Hence the constraint $\sum_{e \in E(v)} \psi(e) = 1$ is exactly the constraint that the object be allocated to one of the agents.

Now, the graph G described in the previous paragraph is bi-partite (partition the type profiles (that is, the nodes of G) according to whether the profile has an odd or even number of entries equal to 0). It follows from Theorem 11.4 of Korte and Vygen (2018) that all extreme points of the perfect matching polytope are perfect matchings. All perfect matchings represent deterministic DIC mechanisms. Hence all extreme points of the set of DIC mechanisms are deterministic.

The claim for the general case, where we have $\Theta_i \subseteq \{0,1\}$ for all i, follows from the previous paragraph by viewing a DIC mechanism on Θ as a mechanism on $\{0,1\}^n$ that ignores the reports of those whose type spaces are singletons. □

Lemma 3.5. *If $|\{i \in \{1,\ldots,n\}: |\Theta_i| \geq 2\}| \leq 2$, then all extreme points of the set of DIC mechanisms are deterministic.*

Proof of Lemma 3.5. We may assume $n \geq 3$, as otherwise the claim follows from Lemma 3.4. We will prove the claim for the case where $|\{i \in \{1,\ldots,n\}: |\Theta_i| \geq 2\}| = 2$, the other cases being simpler. After possibly relabelling the agents, suppose we have $|\Theta_1| \geq 2$ and $|\Theta_2| \geq 2$. Let ϕ be a stochastic DIC mechanism. Notice that at all profiles θ where either agent 1 or agent 2 but not both is enjoying an interior winning probability, there must be an agent in $\{3,\ldots,n\}$ who is also enjoying an interior winning probability; let i_θ denote one such agent. For a number $\varepsilon > 0$ to be chosen later, consider $f: \Theta \to \{-\varepsilon, 0, \varepsilon\}^n$ defined for all θ as follows:

(1) If $\phi_1(\theta) \in (0,1)$ and $\phi_2(\theta) \in (0,1)$, let $f_1(\theta) = \varepsilon$, let $f_2(\theta) = -\varepsilon$, and let $f_i(\theta) = 0$ for all $i \notin \{1,2\}$.

(2) If $\phi_1(\theta) \in (0,1)$ and $\phi_2(\theta) \notin (0,1)$, let $f_1(\theta) = \varepsilon$, let $f_{i_\theta}(\theta) = -\varepsilon$, and let $f_i(\theta) = 0$ for all $i \notin \{1, i_\theta\}$.

(3) If $\phi_1(\theta) \notin (0,1)$ and $\phi_2(\theta) \in (0,1)$, let $f_2(\theta) = -\varepsilon$, let $f_{i_\theta}(\theta) = \varepsilon$, and let $f_i(\theta) = 0$ for all $i \notin \{2, i_\theta\}$.

Using that, for all θ, agent i_θ has a singleton type space, it is easy to see that $\phi + f$ and $\phi - f$ are two DIC mechanisms distinct from ϕ whenever ε is sufficiently small. Thus ϕ is not an extreme point. □

Proof of Theorem 3.3. Lemma 3.2, Lemma 3.4, and Lemma 3.5 imply that all extreme points are deterministic if one of the conditions (1) to (3) holds. Now let conditions (1) to (3) all fail. We know from Section 3.5.2 that a stochastic extreme point exists in the hypothetical situation where $n = 4$ and the set of type profiles is $\hat{\Theta} = \{\ell, r\} \times \{u, d\} \times \{f, c, b\} \times \{0\}$. Since (1) to (3) all fail, we can relabel the agents and types such that agents 1 to 4 have these sets as subsets of their respective sets of types. Let ϕ^* denote the stochastic extreme point Section 3.5.2. Using ϕ^*, it is straightforward to define a stochastic extreme point for the actual set of type profiles with n agents. To see this in detail, let us agree to the following notation: when $i \in \{1,2,3\}$, then $\hat{\Theta}_{-i}$ means the sets of type profiles of agents $\{1,2,3,4\} \setminus \{i\}$ that belong to $\hat{\Theta}$. Now consider $\psi^* : \Theta \to \mathbb{R}^n$ defined as follows: For all $i \in \{1, \dots, n\} \setminus \{1,2,3,4\}$, let $\psi_i^* = 0$; for all $i \in \{1,2,3\}$ and all $\theta \in \Theta$, let $\psi_i^*(\theta) = \phi_i^*(\theta_1, \theta_2, \theta_3, \theta_4)$ if $(\theta_j)_{j \in \{1,2,3,4\} \setminus \{i\}} \in \hat{\Theta}_{-i}$, and let $\psi_i^*(\theta) = 0$ if $(\theta_j)_{j \in \{1,2,3,4\} \setminus \{i\}} \notin \hat{\Theta}_{-i}$; let $\psi_4^* = 1 - \sum_{i=1}^{3} \psi_i^*$. A moment's thought reveals that ψ^* is a well-defined DIC mechanism. To see that it is a stochastic extreme point, consider an arbitrary DIC mechanism ψ that appears in a convex combination that equals ψ^*. We know from Section 3.5.2 that ψ must agree with ψ^* whenever the types of agents 1 to 4 are in $\hat{\Theta}$. From here it is easy to see that ψ must agree with ψ^* at all other profiles, too. □

3.A.3 Anonymous juries

3.A.3.1 Proof of Theorem 3.4

Proof of Theorem 3.4. Let ϕ be DIC and anonymous.

The following notation is useful. Let T denote the common type space. Let T^{n-1} with generic element θ^{n-1} denote the $(n-1)$-fold Cartesian product of T. We will frequently consider profiles obtained from a profile θ^{n-1} in T^{n-1} by replacing one entry of θ^{n-1}. For instance, we write (t, θ^{n-1}_{-j}) to denote the profile obtained by replacing the j'th entry of θ^{n-1} by t.

By DIC, for all i, we may drop i's type from i's winning probability. Thus we write $\phi_i(\theta^{n-1})$ for i's winning probability when the types of the others are $\theta^{n-1} \in T^{n-1}$. Anonymity implies that $\phi_i(\theta^{n-1})$ is invariant to permutations of θ^{n-1}.

We use the following auxiliary claim.

Claim 3.5. *Let $i \in \{1, \dots, n\}$, $t \in T$, $t' \in T$, and $\theta^{n-1} \in T^{n-1}$. Then*

$$\sum_{j=1}^{n-1} \Big(\phi_i(t, \theta_{-j}^{n-1}) - \phi_i(t', \theta_{-j}^{n-1}) \Big) = 0. \tag{3.A.11}$$

Proof of Claim 3.5. Let us arbitrarily label θ^{n-1} as $(\theta_j)_{j \in N \setminus \{i\}}$. Let us also fix an arbitrary type $\theta_i \in T$.

In an intermediate step, let j be distinct from i. For clarity, we spell out winning probabilities as follows: $\phi_i(r_i = t, r_j = t', r_{-ij} = \theta_{-ij})$ means i's winning probability when i reports t, j reports t', and all remaining agents report θ_{-ij}. A permutation of i's and j's reports does not change the winning probabilities of the agents other than i and j. Since the object is allocated with probability one, we have

$$\begin{aligned} &\phi_i(r_i = t, r_j = t', r_{-ij} = \theta_{-ij}) + \phi_j(r_i = t, r_j = t', r_{-ij} = \theta_{-ij}) \\ =&\phi_i(r_i = t', r_j = t, r_{-ij} = \theta_{-ij}) + \phi_j(r_i = t', r_j = t, r_{-ij} = \theta_{-ij}). \end{aligned}$$

By rearranging the previous display, and by DIC, we obtain

$$\begin{aligned} &\phi_i(r_i = t, r_j = t', r_{-ij} = \theta_{-ij}) - \phi_i(r_i = t', r_j = t, r_{-ij} = \theta_{-ij}) \\ =&\phi_j(r_i = t', r_j = \theta_j, r_{-ij} = \theta_{-ij}) - \phi_j(r_i = t, r_j = \theta_j, r_{-ij} = \theta_{-ij}). \end{aligned} \tag{3.A.12}$$

Now consider summing (3.A.12) over all $j \in \{1, \dots, n\} \setminus \{i\}$. This summation yields

$$\sum_{j\,:\, j \neq i} \Big(\phi_i(r_i = t, r_j = t', r_{-ij} = \theta_{-ij}) - \phi_i(r_i = t', r_j = t, r_{-ij} = \theta_{-ij}) \Big) \tag{3.A.13}$$

$$= \sum_{j\,:\, j \neq i} \Big(\phi_j(r_i = t', r_j = \theta_j, r_{-ij} = \theta_{-ij}) - \phi_j(r_i = t, r_j = \theta_j, r_{-ij} = \theta_{-ij}) \Big). \tag{3.A.14}$$

In (3.A.14), the profiles considered are all of the form $(r_i = t', r_{-i} = \theta_{-i})$ and $(r_i = t, r_{-i} = \theta_{-i})$, respectively. Note that by DIC we have $\phi_i(r_i = t', r_{-i} = \theta_{-i}) - \phi_i(r_i = t, r_{-i} = \theta_{-i}) = 0$. Hence (3.A.14) equals

$$\sum_{j=1}^{n} \Big(\phi_j(r_i = t', r_{-i} = \theta_{-i}) - \phi_j(r_i = t, r_{-i} = \theta_{-i}) \Big).$$

Since the object is always allocated, the term in the previous display equals 0. Hence the sum in (3.A.13) equals

$$\sum_{j:\, j\neq i}\Big(\phi_i(r_i = \theta_i, r_j = t', r_{-ij} = \theta_{-ij}) - \phi_i(r_i = \theta_i, r_j = t, r_{-ij} = \theta_{-ij})\Big) = 0.$$

We now revert to our usual notation. By DIC, we may drop i's report from ϕ_i. Since ϕ_i is permutation-invariant with respect to $N \setminus \{i\}$, we may also write

$$\begin{aligned}\phi_i(r_i = \theta_i, r_j = t', r_{-ij} = \theta_{-ij}) &= \phi_i(t', \theta^{n-1}_{-j}) \qquad \text{and}\\ \phi_i(r_i = \theta_i, r_j = t, r_{-ij} = \theta_{-ij}) &= \phi_i(t, \theta^{n-1}_{-j}).\end{aligned}$$

Thus we obtain the desired equality $\sum_{j=1}^{n-1}\Big(\phi_i(t', \theta^{n-1}_{-j}) - \phi_i(t, \theta^{n-1}_{-j})\Big) = 0.$ □

In what follows, let i be an arbitrary agent. We show i's winning probability is constant in the reports of others. To that end, let us fix an arbitrary type $t^* \in T$. For all $k \in \{0, \ldots, n-1\}$, let T^{n-1}_k denote the subset of profiles in T^{n-1} where exactly k-many entries are distinct from t^*. Let p_i denote i's winning probability when all other agents report t^*. We will show via induction over k that i's winning probability is equal to p_i whenever the others report a profile in T^{n-1}_k. This completes the proof since $T^{n-1} = \cup_{k=0}^{n-1} T^{n-1}_k$ holds.

Base case $k = 0$. Immediate from the definitions of p_i and T^{n-1}_0.

Induction step. Let $k \geq 1$. Let all $\hat{\theta}^{n-1} \in \cup_{\ell=0}^{k-1} T^{n-1}_\ell$ satisfy $\phi_i(\hat{\theta}^{n-1}) = p_i$. Letting $\theta^{n-1} \in T^{n-1}_k$ be arbitrary, we show $\phi_i(\theta^{n-1}) = p_i$.

By anonymity, we may assume that exactly the first k entries of θ^{n-1} are distinct from t^*. That is, there exist types $t_1, \ldots, t_k$ all distinct from t^* such that $\theta^{n-1} = (t_1, \ldots, t_k, t^*, \ldots, t^*)$.

Let $\tilde{\theta}^{n-1} = (t_1, \ldots, t_{k-1}, t^*, \ldots, t^*)$. This profile is obtained from θ^{n-1} by replacing t_k by t^*. We now invoke Claim 3.5 to infer

$$\sum_{j=1}^{n-1} \phi_i(t_k, \tilde{\theta}^{n-1}_{-j}) = \sum_{j=1}^{n-1} \phi_i(t^*, \tilde{\theta}^{n-1}_{-j}). \tag{3.A.15}$$

Consider the profiles appearing in the sum on the left of (3.A.15) as j varies from 1 to $n-1$.

(1) Let $j \leq k-1$. Since exactly the first $k-1$ entries of $\tilde{\theta}$ are distinct from t^*, it follows that $(t_k, \tilde{\theta}^{n-1}_{-j})$ is another profile where exactly $k-1$ entries differ from t^*. Hence the induction hypothesis implies $\phi_i(t_k, \tilde{\theta}^{n-1}_{-j}) = p_i$.

(2) Let $j > k-1$. In the profile $(t_k, \tilde{\theta}^{n-1}_{-j})$, the first $k-1$ entries are $t_1, \ldots, t_{k-1}$, the j'th entry is t_k, and all remaining entries are t^*. Hence $(t_k, \tilde{\theta}^{n-1}_{-j})$ is a permutation of θ^{n-1}. Anonymity implies $\phi_i(t_k, \tilde{\theta}^{n-1}_{-j}) = \phi_i(\theta^{n-1})$.

Hence the sum on the left of (3.A.15) equals $\sum_{j=1}^{n-1} \phi_i(t, \tilde{\theta}_{-j}^{n-1}) = (k-1)p_i + (n-k)\phi_i(\theta^{n-1})$

Now consider the sum on the right of (3.A.15). For all j, a moment's thought reveals that the profile $(t^*, \tilde{\theta}_{-j}^{n-1})$ contains at most $(k-1)$-many entries different from t^*. By the induction hypothesis, therefore, the sum on the right of (3.A.15) equals $(n-1)p_i$.

The previous two paragraphs and (3.A.15) imply $(k-1)p_i + (n-k)\phi_i(\theta^{n-1}) = (n-1)p_i$. Equivalently, $(n-k)(\phi_i(\theta^{n-1}) - p_i) = 0$. Since $k \leq n-1$, we find $\phi_i(\theta^{n-1}) = p_i$, as promised. □

3.A.3.2 Proof of Theorem 3.5

Proof of Theorem 3.5. We omit the straightforward verification that a jury mechanism with an anonymous jury is partially anonymous.

For the converse, let ϕ be deterministic, partially anonymous, and DIC. Let N denote the set of agents, and let T denote the common type space. For this proof, we write $\phi(\theta)$ to mean the agent who wins at profile θ; this makes sense since ϕ is deterministic.

Let I_i denote the set of agents that influence agent i's winning probability. For all $j \in N$, let $A_j = \{i \in N : j \in I_i\}$ be the set of agents that are influenced by j. Let $I = \{i \in N : A_i \neq \emptyset\}$. We may assume that ϕ is non-constant, meaning $I \neq \emptyset$, as otherwise the proof is trivial.

Given two agents i and j, let $D_{i-j} = A_i \setminus A_j$, and $D_{j-i} = A_j \setminus A_i$, and $C_{ij} = A_j \cap A_i$, and $N_{ij} = N \setminus (A_i \cup A_j)$. Note that, by DIC, the set C_{ij} contains neither i nor j. Hence partial anonymity implies that for all $k \in C_{ij}$ the winning probability of k is invariant with respect to permutations of i and j.

When i, j, and k are given, we write $(t, t', t'', \theta_{-ijk})$ to mean the profile where i, j, and k, respectively, report t, t', and t'', respectively, and all others report θ_{-ijk}.

Claim 3.6. *Let i and j be distinct. Let $\theta_{-ij} \in \Theta_{-ij}$. If there exists $\theta_i, \theta_j \in T$ such that $\phi(\theta_i, \theta_j, \theta_{-ij}) \in D_{i-j}$, then all $\theta_i', \theta_j' \in T$ satisfy $\phi(\theta_i', \theta_j', \theta_{-ij}) \in D_{i-j}$.*

Proof of Claim 3.6. We drop the fixed type profile θ_{-ij} of the others from the notation. To show $\phi(\theta_i', \theta_j') \in D_{i-j}$, it suffices to show $\phi(\theta_i', \theta_j) \in D_{i-j}$ since if the latter is true then definition of D_{i-j} implies $\phi(\theta_i', \theta_j') = \phi(\theta_i', \theta_j)$.

We first claim $\phi(\theta_j, \theta_i) \in D_{i-j}$. If $\phi(\theta_j, \theta_i) \in N_{ij}$, then $\phi(\theta_j, \theta_i) = \phi(\theta_i, \theta_j)$, and we have a contradiction to $\phi(\theta_i, \theta_j) \in D_{i-j}$. If $\phi(\theta_j, \theta_i) \in C_{ij}$, then partial anonymity implies $\phi(\theta_i, \theta_j) \in C_{ij}$, and we have another contradiction to $\phi(\theta_i, \theta_j) \in D_{i-j}$. If $\phi(\theta_j, \theta_i) \in D_{j-i}$, then $\phi(\theta_j, \theta_i) = \phi(\theta_i, \theta_i) \in D_{j-i}$. However, from $\phi(\theta_i, \theta_j) \in D_{i-j}$ we know $\phi(\theta_i, \theta_j) = \phi(\theta_i, \theta_i) \in D_{i-j}$; contradiction. Thus $\phi(\theta_j, \theta_i) \in D_{i-j}$.

We next claim $\phi(\theta_i', \theta_j) \in (D_{i-j} \cup C_{ij})$. Towards a contradiction, suppose not. Then $\phi(\theta_i', \theta_j) \in (D_{j-i} \cup N_{ij})$, and hence $\phi(\theta_i', \theta_j) = \phi(\theta_i, \theta_j) \notin D_{i-j}$. This contradicts the assumption $\phi(\theta_i, \theta_j) \in D_{i-j}$.

In view of the previous paragraph, we can complete the proof by showing $\phi(\theta_i', \theta_j) \notin C_{ij}$. Towards a contradiction, let $\phi(\theta_i', \theta_j) \in C_{ij}$. Partial anonymity implies $\phi(\theta_j, \theta_i') \in C_{ij}$. We have shown earlier that $\phi(\theta_j, \theta_i) \in D_{i-j}$ holds. Hence $\phi(\theta_j, \theta_i') \in D_{i-j}$, and this contradicts $\phi(\theta_j, \theta_i') \in C_{ij}$. Thus $\phi(\theta_i', \theta_j) \notin C_{ij}$ and the proof is complete. □

Claim 3.7. *Let i, j, k be distinct. Let $\theta_k \in T$ and $\theta_{-ijk} \in \Theta_{-ijk}$ be such that all $\theta_i', \theta_j' \in T$ satisfy $\phi(\theta_i', \theta_j', \theta_k, \theta_{-ijk}) \in (C_{ij} \cup N_{ij})$. Then, all $\theta_i', \theta_j', \theta_k' \in T$ satisfy $\phi(\theta_i', \theta_j', \theta_k', \theta_{-ijk}) \in (C_{ij} \cup N_{ij})$.*

Proof of Claim 3.7. Towards a contradiction, suppose $\phi(\theta_i', \theta_j', \theta_k', \theta_{-ijk}) \in (D_{i-j} \cup D_{j-i})$. Suppose $\phi(\theta_i', \theta_j', \theta_k', \theta_{-ijk}) \in D_{i-j}$, the other case being similar. The inclusions $\phi(\theta_i', \theta_j', \theta_k, \theta_{-ijk}) \in (C_{ij} \cup N_{ij})$ and $\phi(\theta_i', \theta_j', \theta_k', \theta_{-ijk}) \in D_{i-j}$ together imply $\phi(\theta_i', \theta_j', \theta_k', \theta_{-ijk}) \in A_k$. Hence $\phi(\theta_i', \theta_j', \theta_k', \theta_{-ijk}) \in D_{k-j}$. We now invoke Claim 3.6 to infer $\phi(\theta_i', \theta_j', \theta_k, \theta_{-ijk}) \in D_{k-j}$. Since we also have $\phi(\theta_i', \theta_j', \theta_k, \theta_{-ijk}) \in (C_{ij} \cup N_{ij})$, we infer $\phi(\theta_i', \theta_j', \theta_k, \theta_{-ijk}) \in N_{ij}$. In particular, we have $\phi(\theta_i', \theta_j', \theta_k, \theta_{-ijk}) \notin A_i$. Hence $\phi(\theta_i', \theta_j', \theta_k, \theta_{-ijk}) \in D_{k-i}$. We now invoke Claim 3.6 to infer $\phi(\theta_i', \theta_j', \theta_k', \theta_{-ijk}) \in D_{k-i}$. In particular, we have $\phi(\theta_i', \theta_j', \theta_k', \theta_{-ijk}) \notin A_i$. This contradicts the assumption $\phi(\theta_i', \theta_j', \theta_k', \theta_{-ijk}) \in D_{i-j}$. □

Claim 3.8. *If $C_{ij} \neq \emptyset$, then $D_{i-j} \cup D_{j-i} = \emptyset$.*

Proof of Claim 3.8. Let $k \in C_{ij}$. We may find a profile θ such that $\phi(\theta) = k$ as else k's winning probability is constantly 0 (which would contradict $k \in C_{ij}$). Denoting by θ_{-ij} the types of agents other than i and j at θ, we appeal to Claim 3.6 to infer that all $\theta_i', \theta_j' \in T$ satisfy $\phi(\theta_i', \theta_j', \theta_{-ij}) \in (C_{ij} \cup N_{ij})$. Repeatedly applying Claim 3.7 implies that all profiles θ' satisfy $\phi(\theta') \in (C_{ij} \cup N_{ij})$. It follows that all agents in $D_{i-j} \cup D_{j-i}$ enjoy a winning probability that is constantly equal to 0. Recalling the definitions $D_{i-j} = A_i \setminus A_j$, and $D_{j-i} = A_j \setminus A_i$, it follows that $D_{i-j} \cup D_{j-i}$ is empty. □

Recall the definition $I = \{i \in N : A_i \neq \emptyset\}$. Consider the binary relation $\sim$ on I defined as follows: Given i and j in I, we let $i \sim j$ if and only if $C_{ij} \neq \emptyset$.

Claim 3.9. *The relation $\sim$ is an equivalence relation. For all $i, j \in I$, if $i \sim j$, then $i \notin A_j$ and $A_i = A_j$.*

Proof of Claim 3.9. It is clear that $\sim$ is symmetric. As for reflexivity, note that $i \in I$ implies $A_i = C_{ii} \neq \emptyset$. Turning to transitivity, suppose $i \sim j$ and $j \sim k$. Hence $C_{ij} \neq \emptyset$ and $C_{jk} \neq \emptyset$. Let $\ell \in C_{jk}$. Claim 3.8 and $C_{ij} \neq \emptyset$ together imply $D_{j-i} = \emptyset$. Hence $\ell \in C_{jk}$ implies $\ell \in C_{ij}$. Hence $\ell \in C_{jk} \cap C_{ij}$, implying $\ell \in C_{ik}$. Hence $i \sim k$.

As for the second part of the claim, let $i \sim j$. Thus $C_{ij} \neq \emptyset$. Claim 3.8 implies $D_{j-i} = D_{i-j} = \emptyset$. This immediately implies $A_i = A_j$. Together with DIC, we also infer $i \notin A_j$. □

Claim 3.9 implies that we may partition I into finitely-many non-empty $\sim$-equivalence classes. (Recall that I is non-empty.) We now claim that there is exactly one $\sim$-equivalence class. Towards a contradiction, suppose not. In view of Claim 3.9, this means that there are distinct i and j such that $A_i \cap A_j = \emptyset$ and $A_i \neq \emptyset \neq A_j$. Let J_i and J_j, respectively, denote the equivalence classes containing i and j, respectively. Let $k \in A_i$ and $\ell \in A_j$. Claim 3.9 implies $k \notin J_i$ and $\ell \notin J_j$ and $k \neq \ell$. Since $k \in A_i$ and ϕ is deterministic, there is a type profile θ such that $\phi(\theta) = k$; there must be another type profile θ' such that $\phi(\theta') = \ell$. However, the definition of equivalence classes implies that k's winning probability depends only on the types of agents in J_i, and that ℓ's winning probability depends only on the types of agents in J_j. Hence there is a type profile where both k and ℓ are winning with probability 1 (such a type profile is obtained by changing at the profile θ the types of agents in J_j to their respective types at θ', and keeping all other types fixed). Contradiction.

Now, Claim 3.9 implies that the members of the unique $\sim$-equivalence class do not influence one another, and that they influence the same set of others. By partial anonymity, it follows ϕ that is a deterministic jury mechanism with an anonymous jury. □

3.A.3.3 Proof of Proposition 3.2

We first give an informal sketch of the proof. The idea is to "symmetrize" the stochastic extreme point ϕ^* from Section 3.5.2.

In Section 3.5.2, there are four agents, the set of type profiles of agents 1 to 3 is a $2 \times 2 \times 3$ set $\hat{\Theta}$, and agent 4 has a singleton type space. Let us view allocating to agent 4 as disposing the object. Let us relabel the types of agents 1 to 3 so that they are all distinct. Across agents 1 to 3 we thus have a set T of seven distinct types. The 3-fold Cartesian product T^3 of T with itself contains six permutations of $\hat{\Theta}$ (one for each permutation of $\{1, 2, 3\}$). In Figure 3.A.1, the common type space is labelled $T = \{1, \dots, 7\}$, and the six permutations of $\hat{\Theta}$ are depicted via six symbols (square, circle, etc.).

We can associate to each permutation of $\hat{\Theta}$ a permutation of the mechanism ϕ^*. The idea is to extend these permutations to a DIC mechanism with disposal on all of T^3. The difficulty is to verify that the resulting mechanism is well-defined. To see the issue, reconsider Figure 3.A.1. For each of the six subsets, imagine rays emanating from the subset and travelling parallel to the axes. Along the ray, exactly one agent's type changes. Hence DIC requires that this agent's winning probability remain constant along the ray. The rays emanating from distinct subset intersect, and we have verify that the sum of the associated winning probabilities does not exceed 1. We use two observations. The first is that, at most two such rays intersect simultaneously; this is a consequence of the fact that the types in $\hat{\Theta}$ are distinct across agents. The second is that the winning probabilities associated with the rays are at most $\frac{1}{2}$; this is a consequence of the construction of ϕ^* in Section 3.5.2.

Proof of Proposition 3.2. We first prove part (2) of the claim, assuming for a moment that part (1) is true. Let $\psi^*\colon T^3 \to [0,1]^3$ be a mechanism with disposal for three agents that meets the conclusion of part (1). We view ψ^* as a mechanism (without disposal) with four agents that ignores the report of agent 4, and where agent 4's winning probability equals the probability that ψ^* does not allocate the object to the first three agents. Using the assumed properties of ψ^*, we obtain a mechanism without disposal that is DIC, partially anonymous, and an extreme point of the set of DIC mechanisms without disposal for four agents.

It remains to prove part (1) of the claim. That is, we show that if $n = 3$, then there is an anonymous DIC mechanism with disposal that is an extreme point of the set of all DIC mechanisms with disposal.

Let us relabel the common type space as $T = \{1,2,3,4,5,6,7\}$. Let $T^3 = \times_{i=1}^{3} T$ denote the 3-fold Cartesian product of T. Let $T_1 = \{1,2\}$, $T_2 = \{3,4\}$ and $T_3 = \{5,6,7\}$ and $\hat{\Theta} = T_1 \times T_2 \times T_3$. In Section 3.5.2, we constructed a stochastic DIC mechanism ϕ^* without disposal in a setting with 4 agents, where the types of agents 1, 2, and 3, respectively, are $\{\ell, r\}$, $\{u,d\}$, $\{f,c,b\}$, respectively, and where agent 4's type is degenerate. By relabeling types, we view ϕ^* as a mechanism with disposal with 3 agents on the set of type profiles $\hat{\Theta}$, and where allocating to agent 4 is identified with disposing the object. The arguments from Section 3.5.2 show that, if $n = 3$ and the set of type profiles is $\hat{\Theta}$, then ϕ^* is an extreme point of the set of DIC mechanisms with disposal.

For later reference, we note that, at all type profiles $\theta \in \hat{\Theta}$ and all $i \in \{1,2,3\}$, agent i's winning probability at θ under ϕ^* is either 0 or 1/2.

Our candidate mechanism will be denoted ψ^*. Let Ξ denote the set of permutations of $\{1,2,3\}$. Let $\Theta^* = \{\xi(\theta)\colon \theta \in \hat{\Theta}, \xi \in \Xi\}$ denote the set of type profiles obtained by permuting a type profile in $\hat{\Theta}$; see Figure 3.A.1. Fixing an arbitrary type profile in $\hat{\Theta}$, the types of the agents at this type profile are all distinct. Consequently, for all θ^* in Θ^* there is a unique profile θ in $\hat{\Theta}$ and ξ in Ξ such that $\theta^* = \xi(\theta)$.

For later reference, we also note that at an arbitrary type profile in Θ^*, the types of distinct agents must belong to distinct elements of the partition $\{T_1, T_2, T_3\}$.

We define ψ^* as follows: For all θ^* in Θ^*, we find the unique $(\theta, \xi) \in T \times \Xi$ such that $\theta^* = \xi(\theta)$, and then let

$$\left(\psi_i^*(\theta^*)\right)_{i=1}^{n} = \left(\phi_{\xi(i)}^*(\xi(\theta))\right)_{i=1}^{n}. \tag{3.A.16}$$

For the remaining profiles, we proceed as follows: For all agents i and profiles θ, if θ differs from at least one profile θ^* in Θ^* in agent i's type and no other agent's type, then i's winning probability at θ equals i's winning probability at θ^* (which makes sense since the latter probability has already been defined in (3.A.16)); else, if no such profile θ^* in Θ^* exists, then agent i's winning probability is set equal to 0.

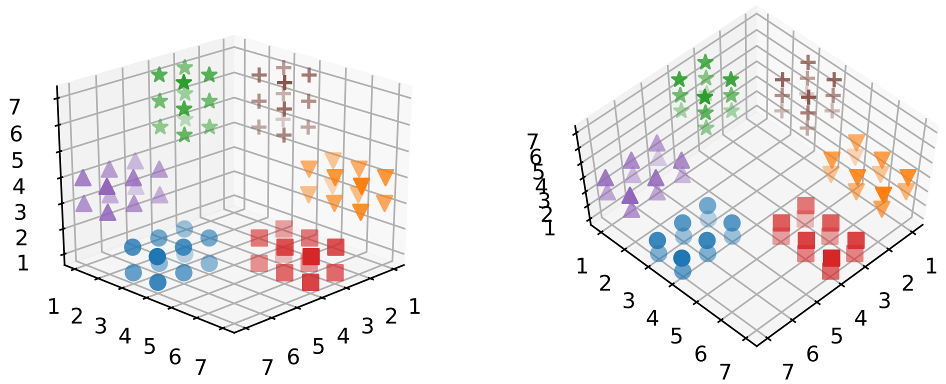


Figure 3.A.1. The set Θ^* viewed from two different angles. Each agent is associated with a distinct axis. Each symbol (square, circle, upward-pointing triangle, etc.) identifies a particular permutation of $\{1, 2, 3\}$. For instance, the upward-pointing triangles are obtained from the downward-pointing triangles by permuting the two agents on the horizontal axes.

To complete the argument, we have to show that ψ^* is well-defined, DIC, stochastic, anonymous, and an extreme point of the set of DIC mechanisms with disposal. Assuming for a moment that the mechanism is well-defined, it is clear that the mechanism is stochastic, and one can easily verify from the definition that it is DIC and anonymous. To show that it is an extreme point of the set of DIC mechanisms, we proceed via the arguments from Section 3.5.2. Indeed, we know from Section 3.5.2 that all DIC mechanisms ψ with disposal that appear in a candidate convex combination must agree with ψ^* on $\hat{\Theta}$, and hence on Θ^*. It is then straightforward to verify that such a mechanism ψ must also agree with ψ^* on $\Theta \setminus \Theta^*$.

It remains to show that ψ^* is well-defined. Towards a contradiction, suppose there is a profile $\theta = (\theta_1, \theta_2, \theta_3)$ in Θ such that $\sum_{i=1}^3 \psi_i^*(\theta) > 1$. By construction, all $i \in \{1, 2, 3\}$ satisfy $\psi_i^* \in \{0, \frac{1}{2}\}$. Hence all three agents enjoy non-zero winning probabilities at θ. By definition of ψ^*, we can infer the following: Since agent 1's winning probability at θ is non-zero, there exists t_1 such that $(t_1, \theta_2, \theta_3) \in \Theta^*$. Similarly, there are t_2 and t_3 such that $(\theta_1, t_2, \theta_3) \in \Theta^*$ and $(\theta_1, \theta_2, t_3) \in \Theta^*$. Recall that $\{T_1, T_2, T_3\}$ is a partition of T. Hence, for all agents i, there is a unique integer $\xi(i)$ in $\{1, 2, 3\}$ such that $\theta_i \in T_{\xi(i)}$. We now recall that if a profile is in Θ^*, then the types of distinct agents belong to distinct elements of the partition $\{T_1, T_2, T_3\}$. Hence we infer from $(t_1, \theta_2, \theta_3) \in \Theta^*$ that $\xi(2) \neq \xi(3)$ holds. Similarly, from $(\theta_1, t_2, \theta_3) \in \Theta^*$ and $(\theta_1, \theta_2, t_3) \in \Theta^*$ we infer $\xi(1) \neq \xi(2)$ and $\xi(1) \neq \xi(3)$. Taken together, we infer $\theta \in \Theta^*$. Hence the vector of winning probabilities at θ is a permutation of the vector of winning probabilities at a profile θ' in $\hat{\Theta}$. At the profile θ', the winning probabilities under ψ^* agree with ϕ^*. Thus there is a profile where the winning probabilities under ϕ^* sum to a number strictly greater than 1. This contradicts the fact that ϕ^* is a well-defined mechanism on $\hat{\Theta}$. □

Appendix 3.B Supplementary Material: Disposal

In this part of the appendix, we relax the requirement that the object always be allocated. An intepretation is that the mechanism designer can dispose or privately consume the object. Accordingly, we refer to such mechanisms as mechanisms with disposal. We discuss how this affects our results from the main text (Section 3.B.1). Further, we show how the existence of stochastic extreme points of the set of DIC mechanisms with disposal can be related to a certain graph (Section 3.B.2).

Beginning with the definitions, a *mechanism with disposal* is a function $\phi : \Theta \to [0,1]^n$ satisfying

$$\forall_{\theta \in \Theta}, \quad \sum_{i=1}^{n} \phi_i(\theta) \leq 1.$$

A mechanism from the main text will be referred to as a mechanism with no disposal. If there is no risk of confusion, we will drop the qualifiers "with disposal" or "with no disposal".

A mechanism with disposal is DIC if and only if for arbitrary i the winning probability ϕ_i is constant in i's report. We will sometimes drop i's report θ_i from $\phi_i(\theta_i, \theta_{-i})$.

A jury mechanism with disposal is defined as in the basic model: For all i, if agent i influences the allocation, then i never wins the object.

We normalize the value from not allocating the object to 0.

A mechanism with n agents and disposal can be viewed as a mechanism with no disposal and with $n+1$ agents where agent $n+1$ has a singleton type space; the value from allocating to $n+1$ is always 0. Likewise, if there are other agents with singleton type spaces, we can always renormalize values and view allocating to one of these agents as disposing the object. In what follows, whenever considering mechanisms with disposal, let us thus simplify by assuming that no agent has a singleton type space; that is, for all agents i we have $|\Theta_i| \geq 2$.

3.B.1 Results from the main text

Here we discuss how our results change when the mechanism can dispose the object.

To begin with, we have the following analogue of Theorem 3.3.

Theorem 3.6. *Fix n and $\Theta_1, \ldots, \Theta_n$. For all agents i, let $|\Theta_i| \geq 2$. All extreme points of the set of DIC mechanisms with disposal are deterministic if and only if at least one of the following is true:*

(1) We have $n \leq 2$.

(2) For all agents i we have $|\Theta_i| = 2$.

Proof of Theorem 3.6. As discussed above, a DIC mechanism with n agents and disposal is a DIC mechanism with $n+1$ agents and no disposal. The claim follows from Theorem 3.3. □

Further below, we provide an alternative proof of Theorem 3.6 that does not invoke Theorem 3.3 but relies on graph-theoretic results. We emphasize that Theorem 3.6 does not imply Theorem 3.3. Namely, we cannot conclude from Theorem 3.6 that if $n = 3$ all extreme points of the set of DIC mechanisms with no disposal are deterministic.

It follows from Theorem 3.6 that Theorem 3.1 (jury mechanisms with 3 agents) carries over to mechanisms with disposal in the sense that all mechanisms with disposal and 2 agents are convex combinations of deterministic jury mechanisms with disposal. Note that, according to Theorem 3.6, this result does not extend to $n = 3$. With $n = 2$, a jury mechanism with disposal admits a single juror whose report determines whether or not the object is disposed or allocated to the other agent.

Proposition 3.1 (on the suboptimality of deterministic DIC mechanisms) analogizes straightforwardly to mechanisms with disposal. Indeed, note that in our proof of Proposition 3.1 agent 4 was simply a dummy agent with value normalized to 0.

Theorem 3.2 (approximate optimality of jury mechanisms under Assumption 3.1 and large n) extends to mechanisms with disposal in a straightforward way, with no changes to the proof.

We already showed via Proposition 3.2 that Theorem 3.4 does not extend to mechanisms with disposal. In fact, the non-constant mechanism constructed in the proof of Proposition 3.2 actually satisfies an even stronger notion of anonymity. Namely, whenever one permutes the type profiles, the vector of winning probabilities is permuted in the same manner.

We next turn to partial anonymity for mechanisms with disposal. In particular, we show that Theorem 3.5 extends under a slight strengthening of partial anonymity. Given a mechanism ϕ, let $\phi_0 = 1 - \sum_{i=1}^{n} \phi_i$ denote the probability that the object is not allocated.

Definition 3.4. Let ϕ be a mechanism with disposal. Let $N = \{1, \dots, n\}$ and $N_0 = N \cup \{0\}$.

(1) Given distinct $i \in N$ and $k \in N_0$, agent i *influences* k if ϕ_k is non-constant in i's report.
(2) The mechanism is *partially $*$-anonymous* if for all $i \in N, j \in N$, and $k \in N_0$ that are all distinct and are such that i and j influence k, agents i and j are exchangeable for k.

In words, partial anonymity is strengthened by demanding that the disposal probability ϕ_0 is permutation-invariant with respect to those agents who influence ϕ_0.

It follows from Theorem 3.5 that a deterministic partially $*$-anonymous DIC mechanism with disposal is a deterministic jury mechanism with an anonymous jury. To see this, let us view disposing the object as allocating to agent 0. Now, agent 0 does not have the same type space as the other agents. Since this was a maintained assumption of Section 3.6, we cannot yet appeal to Theorem 3.5. But, we can simply view the mechanism as a mechanism where agent 0's type space is same as the type spaces of the others, and where agent 0's report is always ignored. By now appealing to Theorem 3.5, the claim follows.

3.B.2 Stochastic extreme points and perfect graphs

In this section, we relate the existence of stochastic extreme points with disposal to a graph-theoretic property called perfection.

3.B.2.1 Preliminaries

We first recall several definitions for a simple undirected graph G with nodes V and edges E.

An *induced cycle of length k* is a subset $\{v_1, \ldots, v_k\}$ of V such that, denoting $v_{k+1} = v_1$, two nodes v_ℓ and $v_{\ell'}$ in the subset are adjacent if and only if $|\ell - \ell'| = 1$.

The *line graph* of G is the graph that has as node set the edge set of G; two nodes of the line graph are adjacent if and only if the two associated edges of G share a node in G.

A *clique* of G is a set of nodes such that every pair in the set are adjacent. A clique is *maximal* if it is not a strict subset of another clique. A *stable set* of G is a subset of nodes of which no two are adjacent. The *incidence vector* of a subset of nodes $\hat{V}$ is the function $x\colon V \to \{0,1\}$ that equals one on $\hat{V}$ and equals zero otherwise. Let $S(G)$ denote the set of incidence vectors belonging to some stable set of G.

The upcoming result uses another property of graphs called *perfection*. For our purposes, it will be enough to know the following facts, all of which may be found in Korte and Vygen (2018).

Lemma 3.6. *All bi-partite graphs and line graphs of bi-partite graphs are perfect. If a graph admits an induced cycle of odd length greater than five, then it is not perfect.*

Our interest in perfect graphs is due to the following theorem of Chvátal (1975, Theorem 3.1); one may also find it in Korte and Vygen (2018, Theorem 16.21).

Theorem 3.7. *A graph G with node set V and edge set E is perfect if and only if the convex hull* $\operatorname{conv} S(G)$ *is equal to the set*

$$\left\{x\colon V \to [0,1] \colon \textit{all maximal cliques } X \textit{ of } G \textit{ satisfy} \sum_{v \in X} x(v) \le 1\right\}. \tag{3.B.1}$$

The set $\operatorname{conv} S(G)$ is the *stable set polytope* of G. The set in (3.B.1) is the *clique-constrained stable set polytope* of G.

3.B.2.2 The feasibility graph

We next define a graph G such that the set of deterministic DIC mechanisms with disposal corresponds to $S(G)$, and such that the set of all DIC mechanisms with disposal coincides with the clique-constrained stable set polytope of G. In view of Theorem 3.7, the question of whether all extreme points are deterministic thus reduces to checking whether G is a perfect graph.

Consider the following graph G with node set V and edge set E. Let

$$V = \cup_{i=1}^{n} \left(\{i\} \times \Theta_{-i}\right),$$

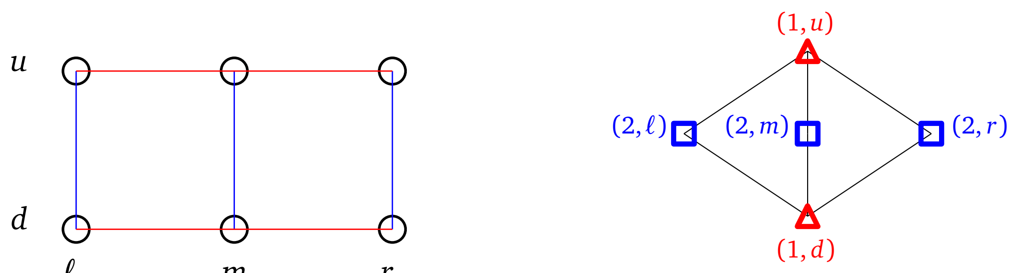
and let two nodes (i, θ_{-i}) and (j, θ'_{-j}) be adjacent if and only if $i \neq j$ and there is a type profile $\hat{\theta}$ satisfying $\hat{\theta}_{-i} = \theta_{-i}$ and $\hat{\theta}_{-j} = \theta'_{-j}$. We refer to G as the *feasibility graph*.

Informally, a node (i, θ_{-i}) is the index for agent i's winning probability when the type profile of the others is θ_{-i}. Two nodes are adjacent if and only if there is a profile $\hat{\theta}$ such that the associated winning probabilities simultaneously appear in the feasibility constraint

$$\sum_{i=1}^{n} \phi_i(\hat{\theta}_{-i}) \leq 1 \tag{3.B.2}$$

of the profile $\hat{\theta}$.

Figure 3.B.1 shows the feasibility graph in an example with two agents; Figure 3.B.2 shows it in an example with three agents.



(a) The set of type profiles Θ. Circles represent type profiles.

(b) The graph G. Red triangles represent nodes of G that are associated with agent 1. Blue squares represent nodes associated with agent 2.

Figure 3.B.1. There are two agents with types $\Theta_1 = \{\ell, m, r\}$ and $\Theta_2 = \{u, d\}$.

Given a node $v = (i, \theta_{-i})$ of G, let us write $\phi(v) = \phi_i(\theta_{-i})$. Note that a clique in the feasibility graph is a subset of nodes of V such that the winning probabilities associated with these nodes all appear in the same feasibility constraint (3.B.2). It follows that there is a one-to-one mapping between maximal cliques of G and type profiles. For a DIC mechanism with disposal, the feasibility constraint (3.B.2) may

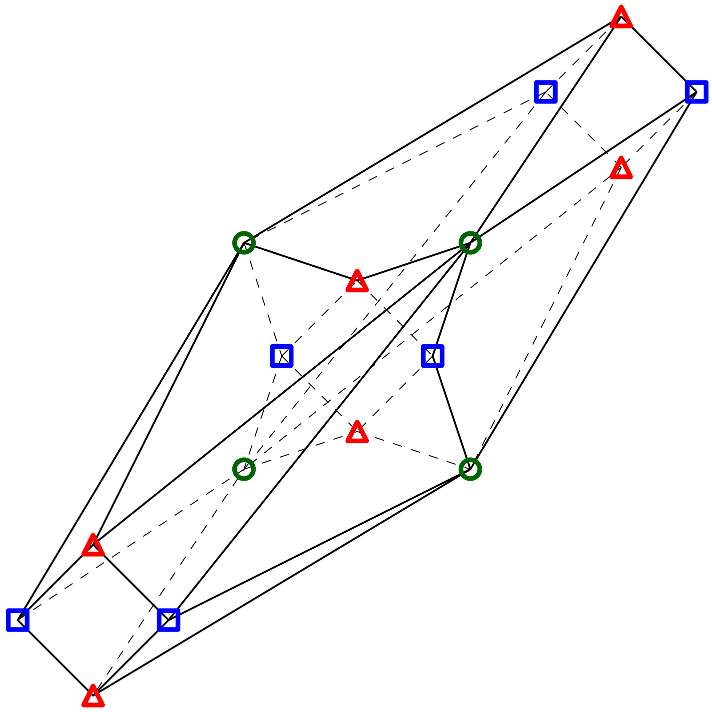


Figure 3.B.2. The feasibility graph G in an example with three agents. Agents 1 and 2 each have two possible types. The nodes of G associated with agents 1 and 2, respectively, are depicted by red triangles and blue squares, respectively. Agent 3 has three possible types; the associated nodes are depicted by green circles. One may view this as the graph G associated with the four-agent environment of Section 3.5.2, except that all nodes of the dummy agent 4 are omitted.

thus be equivalently stated as follows: For all maximal cliques of X of G, we have $\sum_{v \in X} \phi(v) \leq 1$. Thus the set of DIC mechanisms with disposal coincides with the set (3.B.1). One may similarly verify that the set of deterministic DIC mechanisms with disposal coincides with $S(G)$. In view of Theorem 3.7, we deduce:

Lemma 3.7. *All extreme points of the set of DIC mechanisms with disposal are deterministic if and only if G is perfect.*

This leads us to the following alternative proof of Theorem 3.6.

Alternative proof of Theorem 3.6. Let $n = 2$. Observe that the node set of G may be partitioned into the sets $\{1\} \times \Theta_2$ and $\{2\} \times \Theta_1$. By definition, two nodes (i, θ_{-i}) and (j, θ_{-j}) are adjacent only if $i \neq j$. Thus G is bi-partite. Since every bi-partite graph is perfect (Lemma 3.6), the claim follows from Theorem 3.7.

Suppose $|\Theta_i| = 2$ holds for all i. We may relabel the types so that $\Theta_i = \{0, 1\}$ holds for all i. In this case G is the line graph of a bi-partite graph; namely the bi-partite graph with node set $\{0, 1\}^n$ and where two nodes are adjacent if and only if they differ in exactly one entry. The line graph of a bi-partite graph is perfect (Lemma 3.6), and so the claim again follows from Theorem 3.7.

Lastly, suppose $n \geq 3$ and $|\Theta_i| > 2$ for at least one i. We will show that G admits an odd induced cycle of length seven. In view of Lemma 3.6 and Theorem 3.7, this proves that there exists a stochastic extreme point. Let us relabel the agents and types such that the type spaces contain the following subsets of types:

$$\tilde{\Theta}_1 = \{\ell, r\} \quad \text{and} \quad \tilde{\Theta}_2 = \{u, d\} \quad \text{and} \quad \tilde{\Theta}_3 = \{f, c, b\}$$

all hold. Let θ_{-123} be an arbitrary type profile of agents other than 1, 2 and 3 (assuming such agents exist). One may verify that the following is an induced cycle of length seven:

$$\begin{aligned}\big(2,(\ell,c,\theta_{-123})\big) &\longleftrightarrow \big(1,(d,c,\theta_{-123})\big)\\ &\longleftrightarrow \big(3,(r,d,\theta_{-123})\big)\\ &\longleftrightarrow \big(2,(r,b,\theta_{-123})\big)\\ &\longleftrightarrow \big(3,(r,u,\theta_{-123})\big)\\ &\longleftrightarrow \big(1,(u,f,\theta_{-123})\big)\\ &\longleftrightarrow \big(3,(\ell,u,\theta_{-123})\big)\\ \longleftrightarrow \big(2,(\ell,c,\theta_{-123})\big).\end{aligned}$$

$\square$

The proof in the main text for the existence of a stochastic extreme point is slightly more elaborate than the one given above since in the former we explicitly spell out the extreme point. (The proof in the main text uses one of the agents as a dummy, and therefore also works for mechanisms with disposal.) In our view, the advantage of the more elaborate argument is that it facilitates the construction of environments where all deterministic DIC mechanisms fail to be optimal. This lets us give an interpretation as to why it may be optimal to use a lottery. That said, it is clear how the induced cycle defined in the proof of Theorem 3.6 relates to the construction from the main text. The nodes of the cycle correspond to the bold edges of the hyperrectangle in Figure 3.1.

Appendix 3.C Supplementary Material: Additional Results

3.C.1 All extreme points are candidates for optimality

For the following lemma, observe that the set of DIC mechanisms depends only on the number of agents and their type spaces.

Lemma 3.8. *Let $n \in \mathbb{N}$. Let $\Theta_1, \dots, \Theta_n$ be finite sets, and let $\Theta = \times_{i=1}^n \Theta_i$. If ϕ is an extreme point of the set of DIC mechanisms when there are n agents and the set of type profiles is Θ, then there exists a set Ω of value profiles and a distribution μ over $\Omega \times \Theta$ such that in the environment (n, Ω, Θ, μ) the mechanism ϕ is the unique optimal DIC mechanism.*

Proof of Lemma 3.8. The set of DIC mechanisms is a polytope in Euclidean space (being the set of solutions to a finite system of linear inequalities). Hence all its extreme points are exposed (Aliprantis and Border, 2006, Corollary 7.90). Hence there is a function $p\colon \{1,\dots,n\} \times \Theta \to \mathbb{R}$ such that for all DIC mechanisms ψ different from ϕ we have $\sum_{i,\theta} p_i(\theta)(\phi_i(\theta) - \psi_i(\theta)) > 0$. By suitably choosing Ω and μ, the function p represents the objective function of our model. For example, one possible choice of Ω and μ is as follows: Let the marginal of μ on Θ be uniform; for all agents i, let Ω_i be the image of p_i; for all θ, conditional on the type profile realizing as θ, let the value of allocating to agent i be $|\Theta| p_i(\theta)$. □

3.C.2 Implementation with deterministic outcome functions

An indirect mechanism specifies a tuple $M = (M_1, \dots, M_n)$ of finite message sets, and an outcome function $g\colon \times_i M_i \to \Delta\{0, \dots, n\}$. (Given a finite set X, we denote by ΔX the set of distributions over X.) The outcome function is *deterministic* if for all m the distribution $g(m)$ is degenerate. A *strategy* of agent i in (M, g) is a function $\sigma_i\colon \Theta \to \Delta M_i$; let Σ_i denote the set of strategies of agent i in (M, g). A DIC mechanism ϕ is *implementable (in dominant strategies)* via (M, g) if there is a dominant-strategy equilibrium $(\sigma_1, \dots, \sigma_n)$ of (M, g) such that all profiles θ satisfy $\phi(\theta) = \sum_m g(m) \prod_i \sigma_i(m_i|\theta_i)$.

Lemma 3.9. *If a stochastic DIC mechanism ϕ is implementable via an indirect mechanism with a deterministic outcome function, then ϕ is not an extreme point of the set of DIC mechanisms.*

Proof of Lemma 3.9. Towards a contradiction, suppose ϕ is an extreme point. As in the proof of Lemma 3.8, we may find $p\colon \{1, \dots, n\} \times \Theta \to \mathbb{R}$ such that all DIC mechanisms ψ distinct from ϕ satisfy $\sum_{i,\theta} p_i(\theta)(\phi_i(\theta) - \psi_i(\theta)) > 0$. However, since ϕ is implementable via an indirect mechanism with a deterministic outcome function, Proposition 1 of Jarman and Meisner (2017) implies that there is a deterministic DIC mechanism ψ such that

$$\forall_{\theta \in \Theta}, \quad \sum_i p_i(\theta)(\phi_i(\theta) - \psi_i(\theta)) \;\le\; 0.$$

Hence $\sum_{i,\theta} p_i(\theta)(\phi_i(\theta) - \psi_i(\theta)) \le 0$. Since ϕ is stochastic, we have $\psi \ne \phi$; contradiction. □

3.C.3 Total unimodularity

This section of the appendix discusses another potential approach for showing that all extreme points are deterministic. Our aim is to explain why this approach does not help us for the proof of Theorem 3.3 in the difficult case with three agents.

For a function $\phi : \Theta \to [0,1]^n$ to be a DIC mechanism, the function should satisfy the following:

$$\begin{aligned} \forall_{i,\theta}, \quad & 1 \geq \phi_i(\theta) \\ \forall_{i,\theta_i,\theta_i',\theta_{-i}}, \quad & 0 \geq \phi_i(\theta_i,\theta_{-i}) - \phi_i(\theta_i',\theta_{-i}) \geq 0 \\ \forall_{\theta}, \quad & 1 \geq \sum_i \phi_i(\theta) \geq 1 \end{aligned} \tag{3.C.1}$$

For a suitable matrix A and vector b, the set of DIC mechanisms is then the polytope $\{\phi : A\phi \geq b, \phi \geq 0\}$. Here, the matrix A has one row for every constraint in (3.C.1) (after splitting the constraints into one-sided inequalities). Each column of A identifies a pair of the form (i, θ).

A matrix or a vector is *integral* if its entries are all in $\mathbb{Z}$. A polytope is *integral* if all its extreme points are integral. In this language, all extreme points of the set of DIC mechanisms are deterministic if and only if the polytope $\{\phi : A\phi \geq b, \phi \geq 0\}$ is integral.

Recall that a matrix is *totally unimodular* if all its square submatrices have a determinant equal to -1, 0, or 1. A submatrix of a totally unimodular matrix is itself totally unimodular.

Our interest in total unimodularity is due the Hoffman-Kruskal theorem (Korte and Vygen, 2018, Theorem 5.21).

Theorem 3.8. *An integral matrix A is totally unimodular if and only if for all integral vectors b all extreme points of the set $\{\phi : A\phi \geq b, \phi \geq 0\}$ are integral.*

Thus a sufficient condition for all extreme points of the set of DIC mechanisms to be deterministic is that the constraint matrix A be totally unimodular. Unfortunately:

Lemma 3.10. *For all agents i, let $|\Theta_i| \geq 2$. Let $n = 3$. If there exists i such that $|\Theta_i| \geq 3$, then A is not totally unimodular.*

Proof of Lemma 3.10. Towards a contradiction, suppose A is totally unimodular. Consider the constraint matrix $\tilde{A}$ and vector $\tilde{b}$ that define the set of DIC mechanisms with disposal (where such mechanisms are defined in Section 3.B). That is, ϕ is a DIC mechanism with disposal if and only if $\tilde{A}\phi \geq \tilde{b}$ and $\phi \geq 0$. Notice that $\tilde{A}$ is obtained from A by dropping all rows corresponding to constraints of the form $\sum_i \phi_i(\theta) \geq 1$; the vector $\tilde{b}$ is obtained from b by dropping the corresponding entries. In particular, the matrix $\tilde{A}$ is a submatrix of A. Since A is totally unimodular, we conclude that $\tilde{A}$ is totally unimodular. We infer from Theorem 3.8 that all extreme points of the set of DIC mechanism with disposal are deterministic. Since $n = 3$, since all agents have at least binary types, and since at least one agent has non-binary types, we have a contradiction to Theorem 3.6. □

We can give an alternative proof of Lemma 3.10 that does not require Theorem 3.6. Consider the following characterization of total unimodularity due to Ghouila-Houri (1962) (Korte and Vygen, 2018, Theorem 5.25).

Theorem 3.9. *A matrix A with entries in $\{-1,0,1\}$ is totally unimodular if and only if all subsets C of columns of A satisfy the following: There exists a partition of C into subsets C^+ and C^- such that for all rows r of A we have*

$$\left(\sum_{c\in C^+} A(r,c) - \sum_{c\in C^-} A(r,c)\right) \in \{-1,0,1\}. \tag{3.C.2}$$

Alternative proof of Lemma 3.10. Let us relabel the agents and types such that the type spaces contain the following subsets:

$$\tilde{\Theta}_1 = \{\ell, r\} \quad \text{and} \quad \tilde{\Theta}_2 = \{u,d\} \quad \text{and} \quad \tilde{\Theta}_3 = \{f,c,b\}$$

Fixing an arbitrary type profile θ_{-123} of agents other than 1, 2, and 3, let us define the type profiles $\{\theta^a, \theta^b, \theta^c, \theta^e, \theta^f, \theta^g\}$ as in (3.2) in Section 3.5.2. That is, let

$$\begin{aligned} \theta^a &= (\ell,d,c,\theta_{-123}), \quad & \theta^b &= (r,d,c,\theta_{-123}), \quad & \theta^c &= (r,d,b,\theta_{-123}),\\ & & \theta^d &= (r,u,b,\theta_{-123}), \quad & \theta^e &= (r,u,f,\theta_{-123}),\\ & & \theta^f &= (\ell,u,f,\theta_{-123}), \quad & \theta^g &= (\ell,u,c,\theta_{-123}). \end{aligned}$$

Recall that each column of A corresponds to an entry of the form (i,θ). We will argue that the following set C of columns does not admit a partition in the sense of Theorem 3.9.

$$\begin{aligned} C = \{&(1,\theta^a),(1,\theta^b),(3,\theta^b),(3,\theta^c),\\ &(2,\theta^c),(2,\theta^d),(3,\theta^d),(3,\theta^e),\\ &(1,\theta^e),(1,\theta^f),(3,\theta^f),(3,\theta^g),\\ &(2,\theta^g),(2,\theta^a)\} \end{aligned}$$

Towards a contradiction, suppose C does admit a partition into sets C^+ and C^- in the sense of Theorem 3.9. Let us assume $(1,\theta^a)\in C^+$, the other case being similar. Note that θ^a and θ^b differ only in the type of agent 1. Consider the row of A corresponding to the DIC constraint for agent 1 at these type profiles. By referring to (3.C.2) for this row, we deduce $(1,\theta^b)\in C^+$. Next, via a similar argument, the constraint that the object is allocated at θ^b requires $(3,\theta^b)\in C^-$. Continuing in this manner, it is easy to see that $(1,\theta^a)$ must be in C^-. Since $(1,\theta^a)$ is assumed to be in C^+, we have a contradiction to the assumption that C^+ and C^- are a partition of C. □

3.C.4 Maximum weight perfect hypergraph matching

In this section, we explain that the problem of finding an optimal deterministic DIC mechanism corresponds to finding a maximum weight perfect matching on a certain hypergraph.

The hypergraph has as vertices the set of type profiles. Its hyperedges are those type profiles along which the type of exactly one agent i varies across Θ_i. That is, a set of type profiles e is a hyperedge if and only if there exist $i \in \{1, \ldots, n\}$ and $\theta_{-i} \in \Theta_{-i}$ such that $e = \{(\theta_i, \theta_{-i}) : \theta_i \in \Theta_i\}$. We index this hyperedge by (i, θ_{-i}). The weight attached to hyperedge (i, θ_{-i}) is $\mathbb{E}_{\omega_i}[\omega_i | \theta_{-i}]$.

In a matching of this hypergraph, including edge (i, θ_{-i}) in the matching corresponds to allocating to agent i at all type profiles incident to (i, θ_{-i}); this respects DIC for agent i. In a perfect matching, each type profile is covered by some edge; this respects the requirement that the object is always allocated.

If we relax the requirement that the object is always allocated (Section 3.B), we instead consider the larger set of all matchings on the hypergraph. Such a matching can also be interpreted as a stable set of the feasibility graph introduced in Section 3.B.2.2.

References

Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A. Olken, and Julia Tobias. 2012. "Targeting the poor: evidence from a field experiment in Indonesia." *American Economic Review* 102 (4): 1206–40. [98]

Aliprantis, Charalambos D., and Kim C. Border. 2006. *Infinite Dimensional Analysis : A Hitchhikers Guide.* Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-29587-9. [86, 122]

Alon, Noga, Felix Fischer, Ariel Procaccia, and Moshe Tennenholtz. 2011. "Sum of us: Strategyproof selection from the selectors." In *Proceedings of the 13th Conference on Theoretical Aspects of Rationality and Knowledge,* 101–10. [80, 82, 84, 98]

Aziz, Haris, Omer Lev, Nicholas Mattei, Jeffrey Rosenschein, and Toby Walsh. 2016. "Strategyproof peer selection: Mechanisms, analyses, and experiments." In *Proceedings of the AAAI Conference on Artificial Intelligence,* vol. 30. 1. [82]

Aziz, Haris, Omer Lev, Nicholas Mattei, Jeffrey S Rosenschein, and Toby Walsh. 2019. "Strategyproof peer selection using randomization, partitioning, and apportionment." *Artificial Intelligence* 275: 295–309. [82]

Baumann, Leonie. 2018. "Self-ratings and peer review." https://sites.google.com/site/leoniebaumann/. [83]

Ben-Porath, Elchanan, Eddie Dekel, and Barton L Lipman. 2014. "Optimal allocation with costly verification." *American Economic Review* 104 (12): 3779–813. [80, 82]

Ben-Porath, Elchanan, Eddie Dekel, and Barton L Lipman. 2019. "Mechanisms with evidence: Commitment and robustness." *Econometrica* 87 (2): 529–66. [80]

Bhaskar, Dhruva, and Evan Sadler. 2019. "Resource Allocation with Positive Externalities." [83]

Bjelde, Antje, Felix Fischer, and Max Klimm. 2017. "Impartial selection and the power of up to two choices." *ACM Transactions on Economics and Computation (TEAC)* 5 (4): 1–20. [82]

Bloch, Francis, Bhaskar Dutta, and Marcin Dziubiński. 2022. "Selecting a Winner with Impartial Referees." https://www.sites.google.com/site/francisbloch1/papers. [83]

Bloch, Francis, and Matthew Olckers. 2021. "Friend-based ranking in practice." In *AEA Papers and Proceedings,* 111: 567–71. [83]

Bloch, Francis, and Matthew Olckers. 2022. "Friend-based ranking." *American Economic Journal: Microeconomics* 14 (2): 176–214. [83]

Bousquet, Nicolas, Sergey Norin, and Adrian Vetta. 2014. "A near-optimal mechanism for impartial selection." In *International Conference on Web and Internet Economics,* 133–46. Springer. [82]

Budish, Eric, Yeon-Koo Che, Fuhito Kojima, and Paul Milgrom. 2013. "Designing random allocation mechanisms: Theory and applications." *American Economic Review* 103 (2): 585–623. [83]

Caragiannis, Ioannis, George Christodoulou, and Nicos Protopapas. 2019. "Impartial selection with additive approximation guarantees." In *International Symposium on Algorithmic Game Theory,* 269–83. Springer. [82]

Caragiannis, Ioannis, George Christodoulou, and Nicos Protopapas. 2021. "Impartial selection with prior information." *arXiv preprint arXiv:2102.09002.* [82]

Chakravarty, Surajeet, and Todd R Kaplan. 2013. "Optimal allocation without transfer payments." *Games and Economic Behavior* 77 (1): 1–20. [83]

Chen, Yi-Chun, Wei He, Jiangtao Li, and Yeneng Sun. 2019. "Equivalence of Stochastic and Deterministic Mechanisms." *Econometrica* 87 (4): 1367–90. [80, 83, 89, 93]

Chvátal, Vašek. 1975. "On certain polytopes associated with graphs." *Journal of Combinatorial Theory, Series B* 18 (2): 138–54. [83, 93, 118]

Condorelli, Daniele. 2012. “What money can’t buy: Efficient mechanism design with costly types.” *Games and Economic Behavior* 75 (2): 613–24. [83]

Crémer, Jacques, and Richard P McLean. 1985. “Optimal selling strategies under uncertainty for a discriminating monopolist when demands are interdependent.” *Econometrica: Journal of the Econometric Society,* 345–61. [83]

Crémer, Jacques, and Richard P. McLean. 1988. “Full extraction of the surplus in Bayesian and dominant strategy auctions.” *Econometrica: Journal of the Econometric Society,* 1247–57. [83]

de Clippel, Geoffroy, Kfir Eliaz, Daniel Fershtman, and Kareen Rozen. 2021. “On selecting the right agent.” *Theoretical Economics* 16 (2): 381–402. [98]

de Clippel, Geoffroy, Hervé Moulin, and Nicolaus Tideman. 2008. “Impartial division of a dollar.” *Journal of Economic Theory* 139 (1): 176–91. [82]

Edelman, Paul H, and Attila Por. 2021. “A new axiomatic approach to the impartial nomination problem.” *Games and Economic Behavior* 130: 443–51. [82]

Epitropou, Markos, and Rakesh Vohra. 2019. “Optimal On-Line Allocation Rules with Verification.” In *International Symposium on Algorithmic Game Theory,* 3–17. Springer. [82]

Erlanson, Albin, and Andreas Kleiner. 2019. “A note on optimal allocation with costly verification.” *Journal of Mathematical Economics* 84: 56–62. [82]

Fischer, Felix, and Max Klimm. 2015. “Optimal impartial selection.” *SIAM Journal on Computing* 44 (5): 1263–85. [82]

Ghouila-Houri, Alain. 1962. “Caractérisation des matrices totalement unimodulaires.” *Comptes Redus Hebdomadaires des Séances de l’Académie des Sciences (Paris)* 254: 1192–94. [124]

Goldlücke, Susanne, and Thomas Tröger. 2020. “The multiple-volunteers principle.” *Available at SSRN 3753985.* [83]

Guo, Yingni, and Johannes Hörner. 2021. “Dynamic allocation without money.” https://hal.archives-ouvertes.fr/hal-03187506. [83, 98]

Holzman, Ron, and Hervé Moulin. 2013. “Impartial nominations for a prize.” *Econometrica* 81 (1): 173–96. Publisher: Wiley Online Library. [80–82, 84, 85, 96–98]

Jarman, Felix, and Vincent Meisner. 2017. “Deterministic mechanisms, the revelation principle, and ex-post constraints.” *Economics Letters* 161: 96–98. [83, 122]

Kato, Miki, and Shinji Ohseto. 2002. “Toward general impossibility theorems in pure exchange economies.” *Social Choice and Welfare* 19 (3): 659–64. [85]

Kattwinkel, Deniz. 2019. “Allocation with Correlated Information: Too good to be true,” mimeo. [83]

Kattwinkel, Deniz, and Jan Knoepfle. 2021. “Costless Information and Costly Verification: A Case for Transparency.” https://dx.doi.org/10.2139/ssrn.3426817. [83]

Kattwinkel, Deniz, Axel Niemeyer, Justus Preusser, and Alexander Winter. 2022. “Mechanisms without transfers for fully biased agents.” https://arxiv.org/abs/2205.10910. [83]

Korte, Bernhard, and Jens Vygen. 2018. *Combinatorial Optimization: Theory and Algorithms.* 6th ed. 2018. Algorithms and Combinatorics; 21. Berlin, Heidelberg: Springer Berlin Heidelberg. [89, 107, 118, 123, 124]

Lev, Omer, Nicholas Mattei, Paolo Turrini, and Stanislav Zhydkov. 2021. “Peer Selection with Noisy Assessments.” *arXiv preprint arXiv:2107.10121.* [82]

Li, Yunan. 2020. “Mechanism design with costly verification and limited punishments.” *Journal of Economic Theory* 186: 105000. [82, 83]

Lipnowski, Elliot, and Joao Ramos. 2020. “Repeated delegation.” *Journal of Economic Theory* 188: 105040. [98]

Mackenzie, Andrew. 2015. “Symmetry and impartial lotteries.” *Games and Economic Behavior* 94: 15–28. Accessed January 14, 2022. https://doi.org/10.1016/j.geb.2015.08.007. [81, 82, 96, 97]

Mackenzie, Andrew. 2020. “An axiomatic analysis of the papal conclave.” *Economic Theory* 69 (3): 713–43. [82, 96, 97]

Mattei, Nicholas, Paolo Turrini, and Stanislav Zhydkov. 2020. “PeerNomination: Relaxing exactness for increased accuracy in peer selection.” *arXiv preprint arXiv:2004.14939.* [82]

McAfee, R. Preston, and Philip J. Reny. 1992. “Correlated information and mecanism design.” *Econometrica: Journal of the Econometric Society,* 395–421. [83]

Mylovanov, Tymofiy, and Andriy Zapechelnyuk. 2017. “Optimal allocation with ex post verification and limited penalties.” *American Economic Review* 107 (9): 2666–94. [83]

Niemeyer, Axel, and Justus Preusser. 2022. “Simple Allocation with Correlated Types.” Working Paper. [79]

Olckers, Matthew, and Toby Walsh. 2022. “Manipulation and Peer Mechanisms: A Survey.” *arXiv preprint arXiv:2210.01984.* [82]

Pycia, Marek, and M Utku Ünver. 2015. “Decomposing random mechanisms.” *Journal of Mathematical Economics* 61: 21–33. [80, 83, 89]

Rivera Mora, Ernesto. 2022. “Deterministic Mechanism Design.” [83, 89]

Tamura, Shohei. 2016. “Characterizing minimal impartial rules for awarding prizes.” *Games and Economic Behavior* 95: 41–46. Publisher: Elsevier. [82]

Tamura, Shohei, and Shinji Ohseto. 2014. “Impartial nomination correspondences.” *Social Choice and Welfare* 43 (1): 47–54. [82]

Chapter 4

Mechanisms without Transfers for Fully Biased Agents*

4.1 Introduction

A principal has to decide between two options. Which one she prefers depends on the private information of two agents. One agent always prefers the first option; the other always prefers the second. Transfers are infeasible. The principal designs and commits to a mechanism: a mapping from reported information profiles to a – potentially randomized – decision. One prominent example of a setting without transfers is the allocation of a fixed amount of money:

Example 4.1 (Budget allocation). Upper management has endowed a division manager with a fixed budget. She can divide these funds between her two departments L, R. Her objective is to maximize expected returns. Department heads $i = \ell, r$ hold private information θ_i about the future marginal returns y_L, y_R and want to maximize their department's budget. Formally, $(\theta_\ell, \theta_r, y_L, y_R)$ follows some joint distribution and conditional on the private information the manager's marginal return from allocating \$1 to L is $v(\theta_\ell, \theta_r) = E[y_L - y_R | \theta_\ell, \theta_r]$.

To the best of our knowledge this is the first paper that characterizes all implementable mechanisms without transfers under arbitrary correlation. We find a connection between our mechanism design setting and a zero-sum game. Incentive compatibility of a mechanism given a type distribution corresponds to this distribution being a correlated equilibrium in the game induced by the mechanism.

Crémer and McLean's results (Crémer and McLean, 1985, 1988) for the corresponding setting with transfers suggest that the principal should be able to exploit correlation to induce truthful reporting. We define a preorder on type distributions and find that correlation has the opposite effect in our setting: it restricts the

* This chapter is based on Kattwinkel et al. (2022).

set of implementable mechanisms. Under their full-rank condition the set of implementable mechanisms collapses completely and the principal can never do better than choosing her ex-ante preferred option. We give necessary and sufficient conditions for the existence of a "profitable" mechanism that allows the principal to do better. When she is ex-ante indifferent the existence of a profitable mechanism is equivalent to a non-additive payoff structure. When she is not ex-ante indifferent a key insight is that choosing a mechanism corresponds to introducing endogenous correlation. Existence of a profitable mechanism depends on the value of a related optimal transport problem in which the principal chooses this endogenous correlation structure. Incentive constraints translate into an equal marginals condition and an orthogonality constraint between the exogenously given type distribution and the endogenously chosen one.

One application of our results is the problem of allocating a single nondisposable good between two agents. In Section 4.6, we extend our setting and study the problem of allocating a (potentially disposable) good among n agents under independence. When the good has to be allocated, we find that a profitable mechanism exists if and only if a generalized version of the additivity condition is violated. Under free disposal, a profitable mechanism exists if and only if there is an agent such that the principal's value from allocating to that agent depends on the types of other agents.

More broadly, our results convey that there is large class of settings without transfers where the principal can profit from designing a mechanism that elicits the agent's information despite their opposed interests. This scope for communication does not rely on any correlation of the agents' information but instead on types interdependence in the principal's preferences.

4.2 Model

There is a principal, two agents $i = \ell, r$ and a decision: L or R. Agent ℓ always prefers L; agent r always prefers R. Agents enjoy utility 1 if their favored decision is taken and 0 otherwise.[1] Each agent has a private *type* $\theta_i \in \Theta_i$ ($|\Theta_i| < \infty$) and the type profile $\theta = (\theta_\ell, \theta_r)$ is drawn from a commonly known distribution $\pi(\theta_\ell, \theta_r)$ with positive[2] marginals π_ℓ, π_r. Let Π be the set of joint type distributions with positive marginals and let $\Pi(\pi_\ell, \pi_r)$ be the set of joint type distributions with marginals π_i. The independent type distribution with marginals π_i is denoted by $\pi_\ell \pi_r$.

1. All of our results would apply unchanged if agents receive utility $\bar{u}_i(\theta_i)$ from their preferred decision and utility $\underline{u}_i(\theta_i)$ from their less preferred decision, where $\bar{u}_i > \underline{u}_i$.

2. This is without loss. Note that we do not assume full support.

The principal designs and commits to a mechanism. By the revelation principle she can restrict attention to direct, incentive-compatible[3] mechanisms $x\colon \Theta = \Theta_\ell \times \Theta_r \to [0,1]$, where $x(\theta_\ell, \theta_r)$ denotes the probability that L is chosen if agent ℓ reports θ_ℓ and agent r reports θ_r. From now on we refer to direct mechanisms simply as mechanisms.

If the realized type profile is θ then the principal receives a payoff of $v_L(\theta)$ from L and of $v_R(\theta)$ from R, her (without loss) ex-ante preferred option. We normalize $v_R = 0$ so that $E_\pi[v_L(\boldsymbol{\theta})] \le 0$.[4] From now on we denote $v_L = v$.

The principal's problem then reads:

$$\begin{array}{llll} \max\limits_{0 \le x(\theta) \le 1} & E_\pi[v(\boldsymbol{\theta})x(\boldsymbol{\theta})] & & \\ \text{s.t.} & E_\pi[x(\theta_\ell, \boldsymbol{\theta}_r)|\theta_\ell] \ge E_\pi[x(\theta'_\ell, \boldsymbol{\theta}_r)|\theta_\ell] & \forall \theta_\ell, \theta'_\ell & (IC_\ell) \\ & E_\pi[x(\boldsymbol{\theta}_\ell, \theta_r)|\theta_r] \le E_\pi[x(\boldsymbol{\theta}_\ell, \theta'_r)|\theta_\ell] & \forall \theta_r, \theta'_r & (IC_r) \end{array}$$

Given $\pi \in \Pi$, let the set of IC mechanisms be $\mathscr{X}(\pi)$. A mechanism is said to be *profitable* if it is IC and yields the principal a strictly greater payoff than choosing her ex-ante preferred option R without consulting the agents. Given our normalization of the principal's payoff, an IC mechanism x is profitable if and only if $E_\pi[v(\boldsymbol{\theta})x(\boldsymbol{\theta})] > 0$.

4.3 Implementation

In this section we characterize the set of IC mechanisms given a type-distribution. The proof is based on the observation that incentive-compatibility can be phrased in terms of the correlated equilibria of an auxiliary two-player zero-sum game.

Let $\pi \in \Pi$ and let x be an IC mechanism. The IC conditions read

$$\sum_{\theta_r} \pi(\theta_r|\theta_\ell)x(\theta_\ell, \theta_r) \ge \sum_{\theta_r} \pi(\theta_r|\theta_\ell)x(\theta'_\ell, \theta_r) \quad \forall\, \theta_\ell, \theta'_\ell \qquad (IC'_\ell)$$

$$\sum_{\theta_\ell} \pi(\theta_\ell|\theta_r)x(\theta_\ell, \theta_r) \le \sum_{\theta_\ell} \pi(\theta_\ell|\theta_r)x(\theta_\ell, \theta'_r) \quad \forall\, \theta_r, \theta'_r. \qquad (IC'_r)$$

Consider now the auxiliary two-player zero-sum game G in which the Maximizer chooses θ_ℓ, the Minimizer chooses θ_r and the objective (i.e. the Maximizer's payoff if θ_ℓ and θ_r is chosen) is $x(\theta_\ell, \theta_r)$. In this game we can interpret π as a *correlated strategy* under which the Maximizer's payoff is $\sum_{\theta_\ell}\sum_{\theta_r} \pi(\theta_\ell, \theta_r)x(\theta_\ell, \theta_r)$. With this interpretation the IC conditions become *obedience conditions* and π becomes a *correlated equilibrium* of G:

3. More precisely: Bayesian IC. In this setting, the only ex-post IC mechanisms are constant mechanisms.

4. Bold symbols denote random variables.

Lemma 4.1. *A mechanism x is IC under some type distribution $\pi \in \Pi$ if and only if π is a correlated equilibrium of the auxiliary two-player zero-sum game in which the Maximizer chooses $\theta_\ell \in \Theta_\ell$, the Minimizer chooses $\theta_r \in \Theta_r$ and the Maximizer's payoff is $x(\theta_\ell, \theta_r)$.*

Note that under the mechanism design interpretation π is an exogenous part of the environment while x is endogenous. In the auxiliary game the roles are exactly flipped: x is an exogenous while π is endogenous.

Proposition 4.1. *Let $\pi \in \Pi$ and let x be some mechanism. The following are equivalent.*

(i) x is IC under π.
(ii) For each type θ_i of each agent i, $E_\pi[x(\theta_i', \boldsymbol{\theta}_{-i})|\theta_i] = E_\pi[x(\boldsymbol{\theta})]$, for any report θ_i'.

Moreover, if x is IC then $E_\pi[x(\boldsymbol{\theta})] = \bar{x}$, where

$$\bar{x} = \max_{\sigma_\ell \in \Delta\Theta_\ell} \min_{\sigma_r \in \Delta\Theta_r} \sum_{\theta_\ell} \sum_{\theta_r} \sigma_\ell(\theta_\ell)\sigma_r(\theta_r)x(\theta_\ell, \theta_r)$$

is the maximin value of the auxiliary game.

Proposition 4.1 says that a mechanism is IC if and only if each type of each agent is indifferent between every possible report and each type's expectations of x are given by the distribution-independent constant $\bar{x}$.

Proof. Any mechanism that satisfies (ii) is clearly IC. To show the converse let $\pi \in \Pi$ and let x be an IC mechanism. By Lemma 4.1, π is a correlated equilibrium of the auxiliary game G in which the Maximizer chooses θ_ℓ, the Minimizer chooses θ_r and the Maximizer's payoff from such an action profile is $x(\theta_\ell, \theta_r)$.

Suppose now that the Minimizer obeys while the Maximizer ignores his recommendation under π and instead plays the mixed strategy π_ℓ. As this cannot be profitable to him we get

$$\sum_{\theta_\ell} \sum_{\theta_r} \pi(\theta_\ell, \theta_r)x(\theta_\ell, \theta_r) \ \geq \ \sum_{\theta_\ell} \sum_{\theta_r} \pi(\theta_\ell, \theta_r) \sum_{\theta_\ell'} \pi_\ell(\theta_\ell')x(\theta_\ell', \theta_r).$$

But the last term is simply $\sum_{\theta_\ell} \sum_{\theta_r} \pi_\ell(\theta_\ell)\pi_r(\theta_r)x(\theta_\ell, \theta_r)$ and so we obtain

$$\sum_{\theta_\ell} \sum_{\theta_r} \pi(\theta_\ell, \theta_r)x(\theta_\ell, \theta_r) \ \geq \ \sum_{\theta_\ell} \sum_{\theta_r} \pi_\ell(\theta_\ell)\pi_r(\theta_r)x(\theta_\ell, \theta_r). \tag{4.1}$$

The symmetric argument for the Minimizer implies that the opposite inequality to (4.1) must also hold. We conclude that (4.1) must hold with equality. Finally, since π_ℓ has full support if there were some pair $\theta_\ell, \theta_\ell'$ for which the inequality in the obedience constraint (IC_ℓ') were strict then (4.1) could not hold with equality. Hence (IC_ℓ') and (IC_r') must always bind. Thus, if player i is recommended some action θ_i

then he is indifferent between all actions and his interim expectation of x is $\bar{x}_i(\theta_i) = E_\pi[x(\boldsymbol{\theta})|\theta_i]$. We will now show that the interim expectations $\bar{x}_i(\theta_i)$ are actually all the same.

Let $\sigma = (\sigma_\ell, \sigma_r)$ be a Nash equilibrium of G and let $\bar{x} = \sum_{\theta_\ell} \sigma_\ell(\theta_\ell) \sum_{\theta_r} \sigma_r(\theta_r) x(\theta_\ell, \theta_r)$ be the Maximizer's expected payoff under σ. Then for any θ_r it holds that

$$\begin{aligned}\bar{x} &\geq \sum_{\theta_\ell} \pi(\theta_\ell|\theta_r) \sum_{\tilde{\theta}_r} \sigma_r(\tilde{\theta}_r) x(\theta_\ell, \tilde{\theta}_r) \;=\; \sum_{\tilde{\theta}_r} \sigma_r(\tilde{\theta}_r) \sum_{\theta_\ell} \pi(\theta_\ell|\theta_r) x(\theta_\ell, \tilde{\theta}_r) \\ &\geq \sum_{\theta_\ell} \pi(\theta_\ell|\theta_r) x(\theta_\ell, \theta_r) \;=\; \bar{x}_r(\theta_r),\end{aligned}$$

where the first inequality holds since the mixed strategy $\pi(\cdot|\theta_r)$ is not a profitable deviation from σ_ℓ; the second inequality follows from type θ_r's IC constraint for $\tilde{\theta}_r$ and the last equality is by definition. Combining this inequality with the corresponding inequality for the other player we thus have that

$$\bar{x}_r(\theta_r) \;\leq\; \bar{x} \;\leq\; \bar{x}_\ell(\theta_\ell) \quad \forall \theta_\ell, \theta_r.$$

Since the terms on the left and the right hand side of the above inequalities are equal in expectation and all θ_ℓ and θ_r occur with positive probability both inequalities above must always bind. That is to say, for each agent i, $E_\pi[x(\theta_i', \boldsymbol{\theta}_{-i})|\theta_i] = \bar{x}$ for any θ_i and θ_i'. Finally, note that $\bar{x} = \max_{\sigma_\ell \in \Delta\Theta_\ell} \min_{\sigma_r \in \Delta\Theta_r} \sum_{\theta_\ell} \sum_{\theta_r} \sigma_\ell(\theta_\ell)\sigma_r(\theta_r)x(\theta_\ell, \theta_r)$ holds since (σ_ℓ, σ_r) is a Nash equilibrium of the zero sum game G. □

4.4 Comparative Statics for Implementation

In this section, we study how the set of implementable mechanisms depends on the type distribution. We define a preorder on distributions and derive a monotone comparative statics result for the correspondence $\pi \mapsto \mathcal{X}(\pi)$. We conclude that correlation has a restrictive effect.

Definition 4.1. Let $\tau^0, \tau^1, \ldots, \tau^k \in \Delta\Theta_{-i}$ be any beliefs over types of agent $-i$. Then $\{\tau^1, \ldots, \tau^k\}$ is said to *span* τ^0 if there exist coefficients $\alpha_j \in \mathbb{R}$ such that

$$\tau^0(\theta_{-i}) \;=\; \sum_{j=1}^{k} \tau^j(\theta_{-i})\alpha_j \quad \forall \theta_{-i}.$$

Given joint type distributions $\pi, \tilde{\pi} \in \Pi$, π is said to span $\tilde{\pi}$ if for all θ_i, $\{\pi(\cdot|\tilde{\theta}_i) : \tilde{\theta}_i \in \Theta_i\}$ spans $\tilde{\pi}(\cdot|\theta_i)$, $i = \ell, r$.

Hence π spans $\tilde{\pi}$ if each interim belief an agent can hold under $\tilde{\pi}$ is a linear combination of some interim beliefs that he can hold under π. Spanning is reflexive and transitive but not anti-symmetric and therefore defines a preorder.

Example 4.2. Let $\pi \in \Pi$ with marginals π_i. Then π spans the independent type distribution $\tilde{\pi} = \pi_\ell \pi_r$ because

$$\pi_i(\theta_i) = \sum_{\theta_{-i}} \pi(\theta_i|\theta_{-i})\pi_{-i}(\theta_{-i}) \quad \forall \theta_i \forall i.$$

Example 4.3. A joint distribution $\pi \in \Pi$ spans every other joint distribution $\tilde{\pi} \in \Pi$ if and only if the matrix $(\pi(\theta_\ell, \theta_r)) \in \mathbb{R}^{\Theta_\ell \times \Theta_r}$ has full column-rank and full row-rank. This is exactly the condition introduced by Crémer and McLean (see Assumption 4 in their 1985 paper and Theorem 1 in their 1988 paper).

Our first application of the spanning relation shows that the set of IC mechanisms cannot shrink when passing from π to some other type distribution $\tilde{\pi}$ that is spanned by π.

Proposition 4.2. *Let $\pi, \tilde{\pi} \in \Pi$ be type distributions. If π spans $\tilde{\pi}$ then*

$$\mathscr{X}(\pi) \subset \mathscr{X}(\tilde{\pi}).$$

Proof. By Proposition 4.1 a mechanism x is IC under π if and only if

$$\sum_{\theta_{-i}} \pi(\theta_{-i}|\theta_i)(x(\theta_i', \theta_{-i}) - \bar{x}) = 0 \quad \forall \theta_i, \theta_i', i = \ell, r.$$

Now let x be IC under π and consider some $\tilde{\pi} \in \Pi$ spanned by π. By definition, there exist coefficients $\alpha_i(\theta_i, \tilde{\theta}_i)$ such that

$$\tilde{\pi}(\theta_{-i}|\theta_i) = \sum_{\tilde{\theta}_i} \pi(\theta_{-i}|\tilde{\theta}_i)\alpha_i(\theta_i, \tilde{\theta}_i) \quad \forall \theta_i, \theta_{-i}, i = \ell, r.$$

But then x must also be IC under $\tilde{\pi}$ because for all θ_i, θ_i':

$$\sum_{\theta_{-i}} \tilde{\pi}(\theta_{-i}|\theta_i)(x(\theta_i', \theta_{-i}) - \bar{x}) = \sum_{\tilde{\theta}_i} \alpha_i(\theta_i, \tilde{\theta}_i) \sum_{\theta_{-i}} \pi(\theta_{-i}|\tilde{\theta}_i)(x(\theta_i', \theta_{-i}) - \bar{x}) = 0.$$

□

The proof (see Section 4.A) of the next result is another application of the spanning relation.

Proposition 4.3. *Let $\pi \in \Pi$ with marginals π_i and let x be some mechanism. Then:*

(1) If x is IC under π then x is also IC under the independent type distribution $\tilde{\pi} = \pi_\ell \pi_r$.
(2) If the matrix $(\pi(\theta_\ell, \theta_r)) \in \mathbb{R}^{\Theta_\ell \times \Theta_r}$ has full rank then only constant mechanisms are IC.

The maximal elements of the spanning preorder are exactly the full-rank distributions and its minimal elements are exactly the independent distributions.

Crémer and McLean (1985) show in a setting with transfers that full rank correlation makes it possible to implement any allocation rule while paying zero information rents.[5] We show that under the same full-rank condition only mechanisms that ignore the agents' reports are IC. Absent full rank correlation, any mechanism that is IC under correlation must also be IC when types are independent. This shows that the spirit of Crémer and McLean's results is inverted in our setting. The next example illustrates this difference.

Example 4.4. Assume $\Theta_\ell = \Theta_r = \{-1, 1\}$, $\pi_\ell = \pi_r = \frac{1}{2}$ and $v(\theta) = \theta_\ell \theta_r$. Both options yield the principal an ex-ante expected payoff of 0 while the first best mechanism x^* would choose L iff $\theta_\ell = \theta_r$ and yield $E[v(\boldsymbol{\theta})x^*(\boldsymbol{\theta})] = \frac{1}{2} > 0$. If types are independent then x^* is actually IC because from each agent's perspective, any report will lead to the same probability of L. Now assume instead that types are correlated and that π is given by

$$\begin{array}{c|cc|} & -1 & 1 \\ \hline -1 & 0.25-\varepsilon & 0.25+\varepsilon \\ 1 & 0.25+\varepsilon & 0.25-\varepsilon \\ \hline \end{array}$$

where $0 < \varepsilon \leq \frac{1}{4}$ is arbitrary. Then x^* is not IC anymore: For example, type 1 of agent ℓ would infer from his type that the other agent's type is probably -1 and would therefore claim to be type -1 instead of being truthful. Since the distribution matrix has full rank, Proposition 4.3 shows that the only remaining IC mechanisms are constant.

This example also illustrates how more correlated distributions make implementation harder because agents become more informed about each other's types.

4.5 Profitable Mechanisms

In this section we investigate when the principal can design a profitable mechanism. We attack this question from two different angles. Our first characterization is in terms of the principal's objective and applies when the principal is ex-ante indifferent between the two options. The second characterization is in terms of a related optimal transport problem. It also yields an explicit characterization of incentive-compatible mechanisms under independence.

5. The full-rank condition is often seen as generic. In many applications, though, it is not satisfied even when types are correlated. For example, assume that there exists a finite underlying state of the world $\omega \in \{1, \ldots k\}$ such that θ_ℓ and θ_r are independent given ω. That is, $\pi(\theta_\ell, \theta_r | \omega) = \pi_\ell(\theta_\ell|\omega)\pi_r(\theta_r|\omega) \quad \forall \theta_\ell, \theta_r, \omega$. Then

$$\pi(\theta_\ell, \theta_r) = \sum_{\omega=1}^{k} \pi_\ell(\theta_\ell|\omega)\pi_r(\theta_r|\omega)\Pr(\omega).$$

and so each column $\pi(\cdot, \theta_r)$ of the matrix $(\pi(\theta_\ell, \theta_r))_{\theta_\ell, \theta_r}$ is a linear combination of the k vectors $\pi_\ell(\cdot|\omega)$, $\omega = 1, \ldots, k$ (with coefficients $\alpha_{\theta_r}(\omega) = \pi_r(\theta_r|\omega)\Pr(\omega)$). Hence $\mathrm{rank}(\pi) \leq k$.

4.5.1 The role of the objective

Definition 4.2. The principal's objective is said to be additive if there exist functions $v_i : \Theta_i \to \mathbb{R}$ such that

$$v(\theta_\ell, \theta_r) \;=\; v_\ell(\theta_\ell) + v_r(\theta_r) \quad \forall \theta_\ell, \theta_r.$$

Given $\pi \in \Pi$ the objective is said to be π-additive if there exist coefficients $\lambda_\ell(\theta_\ell, \tilde{\theta}_\ell)$, $\lambda_r(\theta_r, \tilde{\theta}_r) \in \mathbb{R}$ such that

$$v(\theta_\ell, \theta_r)\pi(\theta_\ell, \theta_r) \;=\; \sum_{\tilde{\theta}_\ell} \lambda_\ell(\theta_\ell, \tilde{\theta}_\ell)\pi(\theta_r|\tilde{\theta}_\ell) + \sum_{\tilde{\theta}_r} \lambda_r(\theta_r, \tilde{\theta}_r)\pi(\theta_\ell|\tilde{\theta}_r) \quad \forall \theta_\ell, \theta_r. \tag{4.2}$$

Additivity is a special case of π-additivity (take $\lambda_i(\theta_i, \tilde{\theta}_i) = v_i(\theta_i)\pi_i(\tilde{\theta}_i)\mathbf{1}_{(\tilde{\theta}_i=\theta_i)}$) and it is easily seen that the two concepts coincide when $\pi = \pi_\ell \pi_r$. To interpret π-additivity let the type distribution by $\pi \in \Pi$ and consider some mechanism x. When v is π-additive we then get from (4.2) that

$$E_\pi[v(\boldsymbol{\theta})x(\boldsymbol{\theta})] \;=\; \sum_{\theta_\ell, \tilde{\theta}_\ell} \lambda_\ell(\theta_\ell, \tilde{\theta}_\ell) E_\pi[x(\theta_\ell, \boldsymbol{\theta}_r)|\tilde{\theta}_\ell] + \sum_{\theta_r, \tilde{\theta}_r} \lambda_r(\theta_\ell, \tilde{\theta}_r) E_\pi[x(\boldsymbol{\theta}_\ell, \theta_r)|\tilde{\theta}_r]$$

so that $E_\pi[v(\boldsymbol{\theta})x(\boldsymbol{\theta})]$ is a linear combination of the potential expected payoffs $E_\pi[x(\theta_i, \boldsymbol{\theta}_{-i})|\tilde{\theta}_i]$ of types $\tilde{\theta}_i$ from any (mis-)report θ_i. If x is IC then $E_\pi[v(\boldsymbol{\theta})x(\boldsymbol{\theta})]$ is the principal's expected payoff from x and the "misreporting expectations" must all coincide with the maximin value $\bar{x}$. Hence replacing x by the constant mechanism $\tilde{x} \equiv \bar{x}$ does not change the principal's payoff and x cannot be profitable. A necessary condition for the existence of a profitable mechanism is thus that the principal's objective is not π-additive. If the principal is ex-ante indifferent between L and R then this condition is also sufficient

Proposition 4.4. *Let types be distributed according to $\pi \in \Pi$. A profitable mechanism can exist only if the principal's objective is not π-additive. If $E_\pi[v(\boldsymbol{\theta})] = 0$ then a profitable mechanism exists if and only if the principal's objective is not π-additive. In particular, if types are independent then a profitable mechanism exists if and only if the principal's objective is not additive.*

The proof (see Section 4.A) works by projecting $v\pi$ on the linear subspace U of functions that can be expressed in the form of the right hand side of (4.2). Given ex-ante indifference the principal's expected payoff in an IC mechanism depends only on the part of $v\pi$ that is orthogonal to U. We construct a mechanism that yields a strictly positive payoff whenever this projection residual is nonzero.

Example 4.5 (continued from Example 4.1)**.** Consider again the budget allocation problem. Recall that

$$v(\theta_\ell, \theta_r) \;=\; E[y_L - y_R|\theta_\ell, \theta_r] \;=\; E[y_L|\theta_\ell, \theta_r] - E[y_R|\theta_\ell, \theta_r].$$

Let $h_i(\theta_\ell, \theta_r) = E[y_i|\theta_i, \theta_{-i}]$. If $h_i(\theta_\ell, \theta_r)$ depends only on θ_i then Proposition 4.4 implies that there does not exist a profitable mechanism. Hence a necessary condition for the existence of a profitable mechanism is that at least one department head has information that is relevant to the future return of the other department. Now assume that types are iid and that $h_\ell = h_r = h$. Then $E[y_\ell] = E[h(\theta_\ell, \theta_r)] = E[h(\theta_r, \theta_\ell)] = E[y_r]$ so that the principal is ex-ante indifferent. Then a profitable mechanism exists if, and only if $h(\theta_\ell, \theta_r) - h(\theta_r, \theta_\ell)$ is not additive.

4.5.2 The role of correlation

Correlation between agent-types affects the principal through two distinct channels. Firstly, correlation affects the set of mechanisms in which agents find it optimal to be truthful (see Section 4.4). Secondly, fixing a mechanism and assuming that agents are truthful, correlation can increase or decrease the principal's expected payoff by concentrating more mass on specific type profiles. In this section we show that the principal's problem can be viewed as a problem of choosing an "optimal correlation structure".

We start by reinterpreting incentive-compatibility. The proof is in the appendix.

Lemma 4.2. *Let the type distribution be $\pi \in \Pi$. A mechanism x is IC if and only if*

(1) agents are ex-ante indifferent between reports: for all i, θ_i', and θ_i'',

$$E_\pi[x(\theta_i', \boldsymbol{\theta}_{-i})] = E_\pi[x(\theta_i'', \boldsymbol{\theta}_{-i})];$$

(2) their type realizations are uninformative: for all i, θ_i', and θ_i'',

$$E_\pi[x(\theta_i', \boldsymbol{\theta}_{-i})|\theta_i] = E_\pi[x(\theta_i', \boldsymbol{\theta}_{-i})].$$

Ex-ante indifference is equivalent to IC under the independent type distribution $\pi_\ell \pi_r$. Uninformativeness implies that agents cannot gain any payoff-relevant information from their type about their opponent's type. Note that this is automatically satisfied if types are independent. Correlation therefore restricts the set of IC mechanisms by making the agents more informed which adds additional incentive-constraints. From this perspective, IC under correlation lies mid-way between IC under independence and IC under full information.

Lemma 4.2 allows us to derive a necessary and sufficient criterion for the existence of a profitable mechanism. We need the following definition.

Definition 4.3. Two joint type distributions $\pi, \tilde{\pi} \in \Pi(\pi_\ell, \pi_r)$ with the same marginals $\pi_i > 0$ are said to be orthogonal if

$$\mathrm{Cov}(\pi(\theta_i|\boldsymbol{\theta}_{-i}), \tilde{\pi}(\theta_i'|\boldsymbol{\theta}_{-i})) \;=\; 0 \quad \forall \theta_i, \theta_i' \,\forall i.$$

Hence π and $\tilde{\pi}$ are orthogonal if the random variables $\pi(\theta_i|\boldsymbol{\theta}_{-i})$ and $\tilde{\pi}(\theta_i'|\boldsymbol{\theta}_{-i})$ are uncorrelated for all θ_i, θ_i', $i = \ell, r$. Note that

$$\mathrm{Cov}(\pi(\theta_i|\boldsymbol{\theta}_{-i}), \tilde{\pi}(\theta_i'|\boldsymbol{\theta}_{-i})) = \sum_{\theta_{-i}}[\pi(\theta_i|\theta_{-i}) - \pi_i(\theta_i)][\tilde{\pi}(\theta_i'|\theta_{-i}) - \pi_i(\theta_i')]\pi_{-i}(\theta_{-i})$$

and $\pi(\theta_i|\theta_{-i}) - \pi_i(\theta_i)$ is the update of type θ_{-i} about the probability of type θ_i under π. Clearly, if one of π or $\tilde{\pi}$ is the independent type distribution $\pi_l\pi_r$ then orthogonality is automatically satisfied. Otherwise the condition says that for all θ_i, θ_i', the vector $\pi(\theta_i|\cdot) - \pi_i(\theta_i) \in \mathbb{R}^{\Theta_{-i}}$ of possible belief updates of agent $-i$ about the probability of type θ_i under π must be orthogonal to the vector of updates $\tilde{\pi}(\theta_i'|\cdot) - \pi_i(\theta_i) \in \mathbb{R}^{\Theta_{-i}}$ about the probability of θ_i' under $\tilde{\pi}$ under the inner product $\langle a, b\rangle = \sum_{\theta_{-i}} a(\theta_{-i})b(\theta_{-i})$ on $\mathbb{R}^{\Theta_{-i}}$.

The next result shows that the problem of finding a profitable mechanism is intricately related to the choice of an "optimal correlation strucuture": A profitable mechanism exists if and only if it is possible to find some alternative correlation structure that is orthogonal to the exogenously given one and such that — under the alternative correlation structure (and with a suitably transformed objective) — L becomes the principal's ex-ante strictly preferred option. This can be phrased as a constrained optimal transport problem.[6]

Proposition 4.5. *Let the type distribution be $\pi \in \Pi$ and denote its marginals by π_i. Let $\hat{v} = v\pi/\pi_\ell\pi_r$. A profitable mechanism exists if and only if*

$$\left(\max_{\tilde{\pi}\in\Pi(\pi_\ell,\pi_r)} E_{\tilde{\pi}}[\hat{v}(\boldsymbol{\theta})] \text{ s.t. } \tilde{\pi} \text{ is orthogonal to } \pi\right) > 0. \tag{4.3}$$

In particular, if types are independent then a profitable mechanism exists if and only if

$$\max_{\tilde{\pi}\in\Pi(\pi_\ell,\pi_r)} E_{\tilde{\pi}}[v(\boldsymbol{\theta})] > 0.$$

To explain how the constrained optimal transport problem in Proposition 4.5 is related to the principal's problem let x be some mechanism. Together with π, x induces a density $g(\theta) = \pi(\theta)x(\theta) = \frac{\pi(\theta)}{\pi_\ell(\theta_\ell\pi_r(\theta_r)}f(\theta)$ of a measure on Θ whose "correlation structure" depends on an exogenous part $\frac{\pi}{\pi_\ell\pi_r}$ and an endogenous part f. Instead of in terms of mechanisms, the principal's problem can also be phrased in terms of f. Requiring ex-ante indifference for the agents then translates into requiring that f should be a nonnegative multiple of some probability distribution $\tilde{\pi} \in \Pi$ with the same marginals as π: $f = q\tilde{\pi}$ ($q \in [0, 1]$). Uninformativeness translates into $\tilde{\pi}$ being orthogonal to π. Under this reparametrization the principal's objective becomes $qE_{\tilde{\pi}}[\hat{v}(\theta)]$ and the (upper) feasibility constraint on the mechanism becomes

6. For an in-depth treatment of optimal transport see Villani (2009).

a correlation constraint: $q\frac{\tilde{\pi}}{\pi_\ell \pi_r} \leq 1$. If there is a profitable $\tilde{\pi}$ then the principal therefore faces a tradeoff between up-scaling her objective (by increasing q) and the ability to concentrate more mass on type profiles with a positive objective value (by decreasing q). The mere existence of a profitable $\tilde{\pi}$ does not depend on the correlation constraint, however, and after dropping this constraint and dividing everything by q we arrive at the formulation in Proposition 4.5.

The proof of Proposition 4.5 (see Section 4.A) yields another characterization of incentive compatible mechanisms when types are independent.

Corollary 4.1. *If types are independent then a mechanism x is IC if and only if there exist nonnegative coefficients $\{\gamma_j\}_{j=1}^k$ ($k \geq 0$) and extreme points*[7] *π^j of $\Pi(\pi_\ell, \pi_r)$ such that*

$$x = \sum_{j=1}^{k} \frac{\pi^j}{\pi_\ell \pi_r} \gamma_j.$$

Consider an example where both agents have the same number of types (without loss $\Theta_\ell = \Theta_r$) and where marginals are uniform. Together with the Birkhoff-von Neumann Theorem the characterization then implies that a mechanism is IC if and only if it can be decomposed into mechanisms where, up to relabeling of the types, the principal chooses L if and only if both agents make the same report. This illustrates how incentive-compatibility is fundamentally based on the inability (and unwillingness) of the agents to coordinate.

Example 4.6. Assume $\Theta_\ell = \Theta_r = \{1, \dots, m\}$. A mechanism x is said to be a match-your-opponent mechanism if there exists a matching[8] $m : \Theta_\ell \to \Theta_r$ such that

$$x(\theta_\ell, \theta_r) = \begin{cases} 1, & \text{if } \theta_r = m(\theta_\ell) \\ 0, & \text{otherwise} \end{cases}$$

Assume that types are independent with $\pi_i = \frac{1}{N}$. Using the Birkhoff-von Neumann Theorem and Corollary 4.1, a mechanism x is IC if and only if there exist match-your-opponent mechanisms x^j and nonnegative coefficients γ_j such that

$$x = \sum_{j=1}^{k} x^j \gamma_j.$$

7. Recall that an element of a convex set is an extreme point of the set if is is not the midpoint of a line-segment connecting two distinct points in the set. For a characterization of the extreme points of $\Pi(\pi_\ell, \pi_r)$ see Brualdi (2006)[Theorem 8.1.2.].

8. A matching is a bijective function.

Thus, a profitable mechanism exists if and only if there exists a profitable match-your-opponent mechanism. If the principal's objective is supermodular it follows[9] that a profitable mechanism exists if and only if

$$\sum_{t=1}^{m} v(t,t) \ > \ 0.$$

Indeed, as long as types are independent and agents are symmetric (i.e. $\pi_\ell(t) = \pi_r(t)$, $t = 1, \ldots, m$), it can be shown that a profitable mechanism exists if and only if $\sum_{t=1}^{m} \pi_\ell(t) v(t,t) > 0$.[10]

4.6 Allocation with More Than Two Agents and Disposal

One application of our results is the problem of allocating a single nondisposable good between two agents. In this section, we extend our setting and study the problem of allocating a (potentially disposable) good among n agents $i = 1, \ldots, n$.

Agents are again expected utility maximizers and enjoy utility 1 from receiving the good and 0 otherwise.[11] Every agent has a private type $\theta_i \in \Theta_i$. The set of type profiles $\Theta = \Pi_i \Theta_i$ is finite. Throughout this section we assume that types are independent; the joint type distribution is denoted by $\pi(\theta_1, \ldots, \theta_n) = \pi_1(\theta_1) \ldots \pi_n(\theta_n)$[12].

The principal's value from allocating the good to agent i can depend on the types $\theta = (\theta_1, \ldots, \theta_n)$ of all agents and is denoted by $v_i(\theta) \in \mathbb{R}$. A (direct) mechanism specifies for each agent i and every profile θ the probability of allocating the good to this agent when the report profile is θ.

We distinguish between the case where the principal can commit to dispose the good from the case where she is forced to allocate to one of the agents.[13] We normalize the principal's utility from disposing the good to 0. If the principal must allocate the good the feasibility constraint reads: $\sum_{i=1}^{n} x_i(\theta) = 1$; under free disposal it reads: $\sum_{i=1}^{n} x_i(\theta) \le 1$. In either case, the principal's problem is to find a feasible, incentive compatible mechanism that maximizes $E[\sum_{i=1}^{n} v_i(\boldsymbol{\theta}) x_i(\boldsymbol{\theta})]$. As before, we will be interested in whether the principal can do better than choosing her ex-ante preferred option.

We first characterize the set of incentive compatible mechanisms. Whether or not disposal is possible, a mechanism is incentive compatible if and only if each agent's interim probability of obtaining the good does not depend on his report:

9. See Hardy, Littlewood, and Pólya (1952), Becker (1973), and Vince (1990).

10. See Hoffman (1963).

11. All results apply unchanged if agents receive utility $\bar{u}_i(\theta_i) > 0$ from getting the good and 0 otherwise.

12. As before, we assume without loss of generality that $\pi_i > 0$, $i = 1, \ldots, n$.

13. An alternative interpretation of disposal is that the principal allocates the good to herself.

Lemma 4.3. *Assume there are n agents with independent types and let x be a mechanism (with or without disposal). Then x is incentive compatible if and only if*

$$E[x_i(\theta_i, \boldsymbol{\theta}_{-i})] = E[x_i(\boldsymbol{\theta})] \quad \forall i \, \forall \theta_i.$$

Let $\bar{v} = \max_i E[v_i(\boldsymbol{\theta})]$ be the principal's expected payoff from allocating to her ex-ante preferred agent.

An incentive compatible mechanism is profitable if it yields the principal strictly more than choosing her ex-ante preferred option (ignoring type reports). Formally, when there is free disposal, an incentive compatible mechanism is profitable if $\sum_i E[v_i(\boldsymbol{\theta})x_i(\boldsymbol{\theta})] > \max\{0, \bar{v}\}$. Without disposal it is profitable if $\sum_i E[v_i(\boldsymbol{\theta})x_i(\boldsymbol{\theta})] > \bar{v}$.

The following proposition generalizes the scope of Proposition 4.4 to the n-agent case under independence.

Proposition 4.6. *Assume there are n agents with independent types and that the principal has to allocate to some agent. If the principal is unbiased then a profitable mechanism exists if and only if there do not exist functions $u_1(\theta_1), \ldots, u_n(\theta_n)$ such that*

$$v_i(\theta) - v_j(\theta) = u_i(\theta_i) - u_j(\theta_j) \quad \forall i, j \, \forall \theta. \tag{4.4}$$

In the proof (see Section 4.A) we show that a violation of (4.4) remains a necessary condition for the existence of a profitable mechanism when the principal is not unbiased. Proposition 4.6 also allows us to state a necessary and sufficient condition for the existence of a profitable mechanism when the principal is allowed to discard the good:

Proposition 4.7. *Assume there are n agents with independent types and that the principal does not have to allocate the good to the agents. If the principal is unbiased and $\bar{v} = 0$ then a profitable mechanism exists if and only there is an agent j such that $v_j(\theta_j, \theta_{-j})$ is not constant in θ_{-j}.*

4.7 Related Literature

Our main setting can be interpreted as an allocation problem without disposal. It therefore relates to Myerson (1981) who characterizes the set of IC-mechanisms with transfers under independence. Crémer and McLean (1985, 1988) show that with transfers, full rank correlation makes any allocation rule implementable.

Börgers and Postl (2009) study a setting without transfers and two agents with opposed interests. Their setting has a third option that acts as a compromise and types are iid. They consider utilitarian welfare and study second-best rules using numerical tools. Since utilitarian welfare is additive, our results underline the importance of the compromise option for their results. Kim (2017) considers a related setting with at least three ex-ante symmetric alternatives and several agents with iid private values whose interests are not necessarily opposed.

Feng and Wu (2019) ask in a setting without transfers with a perfect conflict of interests not between the agents but between the agents and the principal if the later can do better than choosing her ex-ante preferred option. Goldlücke and Tröger (2020) study "threshold mechanisms" with binary message spaces to assign an unpleasant task without transfers in a setting with symmetric agents with iid types.

The proof of Proposition 4.1 connects implementation of mechanisms with the properties of correlated equilibrium in zero sum games (Aumann, 1974; Rosenthal, 1974; Aumann, 1987).

Our comparative statics result for the set of IC mechanisms with respect to the spanning preorder relates to Blackwell's comparison of experiments (see Blackwell, 1951, 1953; also Börgers, Hernando-Veciana, and Krähmer, 2013) and Bergemann and Morris' comparison of information structures in games (Bergemann and Morris, 2016). A difference is that they compare signals which are informative about a payoff-relevant state while in our setting the signals themselves are payoff-relevant.

Since the set of DIC mechanisms in our setting coincides with the set of constant mechanisms, the existence of non-constant IC mechanisms is related to BIC-DIC equivalence (Manelli and Vincent, 2010; Gershkov et al., 2013). In our setting, IC under correlation can be viewed as a mid-way point between IC under independence and DIC.

Appendix 4.A Proofs

4.A.1 Proof of Lemma 4.2

Proof. Let $\pi \in \Pi$ and let x be IC. By Proposition 4.1, agents must be ex-ante indifferent between reports and their type realizations must be uninformative. Conversely, suppose that x satisfies the assumptions of Lemma 4.2. Ex-ante indifference combined with the law of iterated expectations implies that $E_\pi[x(\theta_i', \boldsymbol{\theta}_{-i})] = E_\pi[x(\boldsymbol{\theta})]\ \forall i, \theta_i'$. Hence for any i and θ_i, θ_i':

$$\begin{aligned} E_\pi[x(\theta_i', \boldsymbol{\theta}_{-i})|\theta_i] &= E_\pi[x(\theta_i', \boldsymbol{\theta}_{-i})] \\ &= E_\pi[x(\boldsymbol{\theta})], \end{aligned}$$

where the first equality follows from uninformativeness. □

4.A.2 Proof of Proposition 4.3

Proof. The first assertion is an immediate consequence of Proposition 4.2 and Example 4.2. To see why the second assertion holds note that if $\tau^0, \tau^1, \ldots, \tau^k \in \Delta\Theta_{-i}$

and $y(\theta_{-i}) : \Theta_{-i} \to \mathbb{R}$ is any function such that

$$\sum_{\theta_{-i}} \tau^j(\theta_{-i}) y(\theta_{-i}) = 0 \quad \forall j = 1, \ldots, k.$$

then also $\sum_{\theta_{-i}} \tau^0(\theta_{-i}) y(\theta_{-i}) = 0$ if τ^0 is spanned by $\{\tau^1, \ldots, \tau^k\}$.

Now assume without loss that $|\Theta_\ell| \geq |\Theta_r|$. By the full rank-condition the vectors $(\pi(\cdot|\theta_\ell))_{\theta_\ell}$ contain a basis of $\mathbb{R}^{\Theta_r}$. In particular, for any θ_r, they span the belief $\mathbb{1}_{\theta_r}$ which puts mass 1 on θ_r. But then ℓ's IC constraints must be satisfied under that belief (consider the function $y_{\theta'_\ell}(\theta_r) = x(\theta'_\ell, \theta_r) - \bar{x}$). But that means that

$$x(\theta'_\ell, \theta_r) = \bar{x} \quad \forall \theta'_\ell.$$

Since θ_r was arbitrary it follows that x must be constant.

Next we will show that $\pi \in \Pi$ is maximal iff it has full rank and minimal iff it is an independent type distribution. We need to show that (i) π has full rank iff it spans every $\tilde{\pi}$ that spans it and (ii) π is an independent type distribution iff for every $\tilde{\pi}$ spanned by π it is also the case that $\tilde{\pi}$ spans π. Throughout, assume without loss of generality that $|\Theta_\ell| \geq |\Theta_r|$.

First note that for $\pi, \tilde{\pi} \in \Pi$, π spans $\tilde{\pi}$ if and only if the row space and the column space of $\tilde{\pi}$ are contained in the row space and the column space, respectively, of π.

Assume that $\pi \in \Pi$ has full rank and that $\tilde{\pi} \in \Pi$ spans π. Denote the column and row spaces of π and $\tilde{\pi}$ by V, W and $\tilde{V}, \tilde{W}$, respectively. Since $\tilde{\pi}$ spans π it holds that $V \subset \tilde{V}$ and $W \subset \tilde{W}$. Since π has full rank, V and W both have dimension $|\Theta_r|$. $\tilde{\pi}$ is an $|\Theta_\ell| \times |\Theta_r|$ matrix and so its column and row spaces cannot have a dimension larger than $|\Theta_r|$. Hence we must have $V = \tilde{V}$ and $W = \tilde{W}$. But that implies that π also spans $\tilde{\pi}$. Thus π is maximal. Conversely, assume that π is maximal. Let $\tilde{\pi} \in \Pi$ be some full rank distribution that spans π. Since π is maximal, π must then also span $\tilde{\pi}$. Hence the row space (and also the column space) of π must have dimension $|\Theta_r|$. This means that π has full rank.

Now let π be an independent distribution. Let $\tilde{\pi}$ be spanned by π. Then, for all θ_i, $\tilde{\pi}(\cdot|\theta_i)$ is a linear combination of the vectors $\pi(\cdot|\tilde{\theta}_i)$ ($\tilde{\theta}_i \in \Theta_i$). But since types are independent under π, the latter vectors all coincide with $\pi_i(\cdot)$. Hence $\tilde{\pi}(\cdot|\theta_i) = \pi_i(\cdot)$ for all θ_i, $i = \ell, r$. Thus $\tilde{\pi} = \pi$; in particular $\tilde{\pi}$ spans π. Hence π is a minimal element. Now let $\tilde{\pi}$ be a minimal element. Let $\tilde{\pi}$ be the independent distribution with the same marginals as π. Then π spans $\tilde{\pi}$ and since π is minimal, $\tilde{\pi}$ must also span π. But $\tilde{\pi}$ has rank one and so π must also have rank one. But that means that π must be an independent type distribution (thus $\pi = \tilde{\pi}$). $\square$

4.A.3 Proof of Proposition 4.4

Proof. First consider a general $\pi \in \Pi$. Define $w(\theta) = v(\theta)\pi(\theta)$. If $E_\pi[v(\boldsymbol{\theta})] = 0$ and agents are truthful then the principal's payoff from some mechanism x is

$$\begin{aligned}\sum_\theta v(\theta)\pi(\theta)x(\theta) &= E_\pi[v(\boldsymbol{\theta})]E_\pi[x(\boldsymbol{\theta})] + \sum_\theta v(\theta)\pi(\theta)(x(\theta) - E_\pi[x(\boldsymbol{\theta})]) \\ &= \sum_\theta w(\theta)(x(\theta) - E_\pi[x(\boldsymbol{\theta})]).\end{aligned}$$

By Proposition 4.1, x is IC if and only if

$$\sum_{\theta_{-i}} \pi(\theta_{-i}|\theta_i)(x(\theta_i', \theta_{-i}) - E_\pi[x(\boldsymbol{\theta})]) = 0 \quad \forall \theta_i, \theta_i' \,\forall i. \tag{3}$$

First assume that there exist coefficients $\lambda_\ell(\theta_\ell, \tilde{\theta}_\ell)$, $\lambda_r(\theta_r, \tilde{\theta}_r) \in \mathbb{R}$ such that

$$w(\theta) = \sum_{\tilde{\theta}_\ell} \lambda_\ell(\theta_\ell, \tilde{\theta}_\ell)\pi(\theta_r|\tilde{\theta}_\ell) + \sum_{\tilde{\theta}_r} \lambda_r(\theta_r, \tilde{\theta}_r)\pi(\theta_\ell|\tilde{\theta}_r) \quad \forall \theta_\ell, \theta_r. \tag{4}$$

Then by (3), any IC mechanism satisfies $\sum_\theta w(\theta)(x(\theta) - E_\pi[x(\boldsymbol{\theta})]) = 0$ and so there is no profitable mechanism.

Now assume instead that there exist no coefficients λ_ℓ and λ_r such that w satisfies (4). Let U be the set of all $u \in \mathbb{R}^{\Theta_\ell \times \Theta_r}$ for which there exist coefficients $\lambda_\ell(\theta_\ell, \tilde{\theta}_\ell)$, $\lambda_r(\theta_r, \tilde{\theta}_r) \in \mathbb{R}$ such that

$$u(\theta) = \sum_{\tilde{\theta}_\ell} \lambda_\ell(\theta_\ell, \tilde{\theta}_\ell)\pi(\theta_r|\tilde{\theta}_\ell) + \sum_{\tilde{\theta}_r} \lambda_r(\theta_r, \tilde{\theta}_r)\pi(\theta_\ell|\tilde{\theta}_r) \quad \forall \theta_\ell, \theta_r.$$

Note that U is a linear subspace of $\mathbb{R}^{\Theta_\ell \times \Theta_r}$ (Indeed, $0 \in U$ and U is closed under addition and multiplication by scalars). Let $\hat{u}$ be the orthogonal projection of w onto U and let $\hat{w}$ be the orthogonal projection of w onto the orthogonal complement of U so that $w = \hat{w} + \hat{u}$. By assumption $w \notin U$, and so $\hat{w} \neq 0$.

As before, given any IC mechanism x it holds that $\sum_\theta \hat{u}(\theta)(x(\theta) - E_\pi[x(\boldsymbol{\theta})]) = 0$ and so the principal's payoff from any IC mechanism x is

$$\sum_\theta w(\theta)(x(\theta) - E_\pi[x(\boldsymbol{\theta})]) = \sum_\theta \hat{w}(\theta)(x(\theta) - E_\pi[x(\boldsymbol{\theta})]).$$

We will now construct a profitable mechanism, i.e. an IC mechanism for which the latter expression is positive. Define

$$\hat{x}(\theta) = \varepsilon(\hat{w}(\theta) - \min \hat{w})$$

where $\varepsilon > 0$ is sufficiently small such that $\hat{x} \leq 1$. First note that $\hat{x}$ is IC. Indeed, for any $\theta'_\ell, \theta''_\ell$,

$$\begin{aligned}
\sum_{\theta_r} \pi(\theta_r|\theta''_\ell)\hat{x}(\theta'_\ell, \theta_r) &= \varepsilon \sum_{\theta_r} \pi(\theta_r|\theta''_\ell)\hat{w}(\theta'_\ell, \theta_r) - \varepsilon \min \hat{w} \\
&= \varepsilon \sum_{\theta_\ell} \sum_{\theta_r} \underbrace{\left(\sum_{\tilde{\theta}_\ell} \mathbb{1}_{\theta_\ell = \theta'_\ell, \tilde{\theta}_\ell = \theta''_\ell} \pi(\theta_r|\tilde{\theta}_\ell) \right)}_{\in U} \hat{w}(\theta_\ell, \theta_r) - \varepsilon \min \hat{w} \\
&= -\varepsilon \min \hat{w},
\end{aligned}$$

where the last equality follows because $\hat{w}$ lies in the orthogonal complement of U. Hence $\hat{x}$ is IC for agent ℓ. The proof that $\hat{x}$ is IC for agent r is symmetric.

We now show that $\hat{x}$ is actually a profitable mechanism. First note that the above incentive-compatibility calculation implies that $E_\pi[\hat{x}(\boldsymbol{\theta})|\theta_\ell] = -\varepsilon \min \hat{w}$ and in particular $E_\pi[x(\boldsymbol{\theta})] = -\varepsilon \min \hat{w}$. Thus the principal's payoff from $\hat{x}$ is

$$\begin{aligned}
\sum_{\theta} \hat{w}(\theta)(x(\theta) - E_\pi[x(\boldsymbol{\theta})]) &= \varepsilon \sum_{\theta} \hat{w}(\theta)\hat{w}(\theta) \\
&> 0.
\end{aligned}$$

Hence $\hat{x}$ is a profitable mechanism.

Finally, assume that types are independent. Note that

$$v(\theta_\ell, \theta_r)\pi_\ell(\theta_\ell)\pi_r(\theta_r) = \sum_{\tilde{\theta}_\ell} \lambda_\ell(\theta_\ell, \tilde{\theta}_\ell)\pi(\theta_r) + \sum_{\tilde{\theta}_r} \lambda_r(\theta_r, \tilde{\theta}_r)\pi(\theta_\ell) \quad \forall \theta_\ell, \theta_r$$

if and only if

$$v(\theta_\ell, \theta_r) = \underbrace{\sum_{\tilde{\theta}_\ell} \tilde{\lambda}_\ell(\theta_\ell, \tilde{\theta}_\ell)}_{v_\ell(\theta_\ell)} + \underbrace{\sum_{\tilde{\theta}_r} \tilde{\lambda}_r(\theta_r, \tilde{\theta}_r)}_{v_r(\theta_r)},$$

where $\tilde{\lambda}_i(\theta_i, \tilde{\theta}_i) = \frac{\lambda_i(\theta_i, \tilde{\theta}_i)}{\pi_i(\theta_i)}$ and so the earlier condition reduces to additivity. □

4.A.4 Proof of Proposition 4.5

Proof. By Lemma 4.2, the principal's problem can be written as

$$\begin{aligned}
\max \quad & \sum_{\theta} \hat{v}(\theta)\pi_\ell(\theta_\ell)\pi_r(\theta_r)x(\theta) \\
\text{s.t.} \quad & \sum_{\theta_{-i}} \pi_{-i}(\theta_{-i})x(\theta'_i, \theta_{-i}) = \sum_{\theta} \pi_\ell(\theta_\ell)\pi_r(\theta_r)x(\theta_\ell, \theta_r) && \forall \theta'_i \, \forall i && (I_i) \\
& \sum_{\theta_{-i}} \pi(\theta_{-i}|\theta_i)x(\theta'_i, \theta_{-i}) = \sum_{\theta_{-i}} \pi_{-i}(\theta_{-i})x(\theta'_i, \theta_{-i}) && \forall \theta_i, \theta'_i \, \forall i && (U_i) \\
& 0 \leq x(\theta) \leq 1 && \forall \theta. && (F)
\end{aligned}$$

Here, (I_i) are the ex-ante indifference constraints (or, equivalently, the IC constraints under the independent type distribution $\pi_\ell \pi_r$) and (U_i) are the uninformativeness constraints. Now define

$$f(\theta_\ell, \theta_r) = \pi_\ell(\theta_\ell)\pi_r(\theta_r)x(\theta_\ell, \theta_r).$$

Using this substitution the principal's objective becomes $\sum_\theta v(\theta)f(\theta)$, the ex-ante indifference constraints become

$$\sum_{\theta_{-i}} f(\theta_i', \theta_r) = \pi_i(\theta_i') \sum_\theta f(\theta) \qquad \forall \theta_i' \, \forall i \tag{4.A.2}$$

and the uninformativeness constraints can be written as

$$\sum_{\theta_{-i}} (\pi(\theta_i|\theta_{-i}) - \pi_i(\theta_i)) f(\theta_i', \theta_{-i}) = 0 \qquad \forall \theta_i, \theta_i' \, \forall i \tag{4.A.3}$$

while the feasibility constraints become

$$0 \le f(\theta_\ell, \theta_r) \le \pi_\ell(\theta_\ell)\pi_r(\theta_r) \qquad \forall \theta_\ell, \theta_r.$$

Equation 4.A.2 says that the marginals of f are proportional to π_i. Since f is nonzero, it is thus a nonnegative multiple of some joint probability distribution $\tilde{\pi}$ with marginals π_i. Hence the principal's problem can be written as

$$\begin{aligned}
&\max_{q \in [0,1]} \max_{\tilde{\pi} \in \Pi(\pi_l, \pi_r)} && q \sum_\theta \hat{v}(\theta)\tilde{\pi}(\theta) && \\
&\text{s.t.} && \sum_{\theta_{-i}} (\pi(\theta_i|\theta_{-i}) - \pi_i(\theta_i))\tilde{\pi}(\theta_i', \theta_{-i}) = 0 && \forall \theta_i, \theta_i' \, \forall i \\
& && q\tilde{\pi}(\theta_\ell, \theta_r) \le \pi_l(\theta_\ell)\pi_r(\theta_r) && \forall \theta_\ell, \theta_r
\end{aligned}$$

where $\Pi(\pi_l, \pi_r)$ is the set of joint type distributions with marginals π_i. A profitable mechanism therefore exists if and only if the latter problem has a positive optimal value.

Since any $\tilde{\pi} \in \Pi(\pi_\ell, \pi_r)$ can be made to satisfy the constraint $q\tilde{\pi}(\theta_\ell, \theta_r) \le \pi_l(\theta_\ell)\pi_r(\theta_r)$ after appropriate scaling, the problem's optimal value is positive if and only if the value of the relaxed problem in which that constraint is left out is positive. Finally, note that

$$\begin{aligned}
&\sum_{\theta_{-i}} (\pi(\theta_i|\theta_{-i}) - \pi_i(\theta_i))\tilde{\pi}(\theta_i', \theta_{-i}) \\
=&\sum_{\theta_{-i}} (\pi(\theta_i|\theta_{-i}) - \pi_i(\theta_i))\tilde{\pi}(\theta_i'|\theta_{-i})\pi_{-i}(\theta_{-i}) \\
=&\sum_{\theta_{-i}} (\pi(\theta_i|\theta_{-i}) - \pi_i(\theta_i))(\tilde{\pi}(\theta_i'|\theta_{-i}) - \pi_i(\theta_i'))\pi_{-i}(\theta_{-i})
\end{aligned}$$

because $\sum_{\theta_{-i}}(\pi(\theta_i|\theta_{-i}) - \pi_i(\theta_i))\pi_i(\theta_i')\pi_{-i}(\theta_{-i}) = 0$. Therefore a profitable $\tilde{\pi}$ exists if and only if the optimal value of the following problem is positive:

$$\begin{aligned} \max_{\tilde{\pi}\in\Pi(\pi_l,\pi_r)} \quad & \sum_{\theta}\hat{v}(\theta)\tilde{\pi}(\theta) \\ \text{s.t.} \quad & \sum_{\theta_{-i}}(\pi(\theta_i|\theta_{-i}) - \pi_i(\theta_i))(\tilde{\pi}(\theta_i'|\theta_{-i}) - \pi_i(\theta_i'))\pi_{-i}(\theta_{-i}) \;=\; 0 \quad \forall\theta_i,\theta_i'\,\forall i. \end{aligned}$$

This concludes the proof of Proposition 4.5. □

4.A.5 Proof of Corollary 4.1

Proof. The proof of Proposition 4.5 shows that under independence, a mechanism x is IC if and only if there exists some $q \in [0,1]$ and $\tilde{\pi} \in \Pi(\pi_l,\pi_r)$ such that $\pi_l\pi_r x = q\tilde{\pi}$. The set $\Pi(\pi_l,\pi_r)$ is a polytope (known as the transportation polytope), hence by the Weyl-Minkowski Theorem it is the convex hull of its finitely many extreme points. This implies the claim. □

4.A.6 Proof of Lemma 4.3

Proof. Let $x = (x_1,\ldots,x_n)$ be an incentive compatible mechanism. Let i be an agent and let $\theta_i,\theta_i' \in \Theta_i$. In order for type θ_i to be truthful, it must hold that $E[x_i(\theta_i,\boldsymbol{\theta}_{-i})] \geq E[x_i(\theta_i',\boldsymbol{\theta}_{-i})]$. In order for type θ_i' to be truthful, $E[x_i(\theta_i,\boldsymbol{\theta}_{-i})] \leq E[x_i(\theta_i',\boldsymbol{\theta}_{-i})]$ must hold. Hence in any incentive compatible mechanism $E[x_i(\theta_i,\boldsymbol{\theta}_{-i})]$ is constant in θ_i, for all i. Since any mechanism satisfying the latter is also incentive compatible, the condition is equivalent to incentive compatibility. Finally, if $E[x_i(\theta_i,\boldsymbol{\theta}_{-i})]$ is constant in θ_i then it must equal $E[x_i(\boldsymbol{\theta})]$. □

4.A.7 Proof of Proposition 4.6

Proof. First assume that (4.4) holds. It follows that there exist functions $u_i(\theta_i)$ such that $v_i(\theta) - v_n(\theta) = u_i(\theta_i) - u_n(\theta_n)$ for all i and θ. Recall that in an IC mechanism x, $\bar{x}_i := E[x_i(\theta_i,\boldsymbol{\theta}_{-i})]$ does not depend on θ_i. The principal's payoff from an incentive compatible mechanism x is therefore

$$\begin{aligned}
&\sum_i E[v_i(\boldsymbol{\theta})x_i(\boldsymbol{\theta})]\\
&=\sum_{i<n} E[(v_i(\boldsymbol{\theta})-v_n(\boldsymbol{\theta}))x_i(\boldsymbol{\theta})]+E[v_n(\boldsymbol{\theta})\sum_i x_i(\boldsymbol{\theta})]\\
&=\sum_{i<n} E[(u_i(\boldsymbol{\theta}_i)-u_n(\boldsymbol{\theta}_n))x_i(\boldsymbol{\theta})]+E[v_n(\boldsymbol{\theta})]\\
&=\sum_{i<n} E[u_i(\boldsymbol{\theta}_i)E[x_i(\boldsymbol{\theta})|\boldsymbol{\theta}_i]]-E[u_n(\boldsymbol{\theta}_n)E[\sum_{i<n} x_i(\boldsymbol{\theta})|\boldsymbol{\theta}_n]]+E[v_n(\boldsymbol{\theta})]\\
&=\sum_{i<n} E[u_i(\boldsymbol{\theta}_i)]\bar{x}_i-E[u_n(\boldsymbol{\theta}_n)(1-\bar{x}_n)]+E[v_n(\boldsymbol{\theta})]\\
&=\sum_i E[u_i(\boldsymbol{\theta}_i)]\bar{x}_i+E[v_n(\boldsymbol{\theta})-u_n(\boldsymbol{\theta}_n)]\\
&=\sum_i E[v_i(\boldsymbol{\theta})+u_n(\boldsymbol{\theta}_n)-v_n(\boldsymbol{\theta})]\bar{x}_i+E[v_n(\boldsymbol{\theta})-u_n(\boldsymbol{\theta}_n)]\\
&=\sum_i E[v_i(\boldsymbol{\theta})]\bar{x}_i.
\end{aligned}$$

Hence, if (4.4) holds then the principal's expected payoff from an incentive compatible mechanism x is the same as her expected payoff from the constant mechanism y given by $y_i(\theta)\equiv\bar{x}_i$. In particular, the principal cannot do better than allocating to her ex-ante preferred agent and so no profitable mechanism exists. Note that we have not used the unbiasedness assumption and so the following is true even if the principal is not unbiased: A profitable mechanism can only exist if (4.4) is violated.

Now let the principal be unbiased. Assume that (4.4) is violated. Then there do no exist functions $u_i(\theta_i)$ $(i=1,\dots,n)$ such that $v_i(\theta)-v_n(\theta)=u_i(\theta_i)-u_n(\theta_j)$ $(i<n)$. (If such functions did exist then it would follow that for any i,j: $v_i(\theta)-v_j(\theta)=v_i(\theta)-v_n(\theta)-(v_j(\theta)-v_n(\theta))=u_i(\theta_i)-u_n(\theta_n)-(u_j(\theta_j)-u_n(\theta_n))=u_i(\theta_i)-u_j(\theta_j)$ and so (4.4) would hold). We will now construct a profitable mechanism.

Let Ω be the vector space of functions from $\{1,\dots,n-1\}\times\Theta$ to $\mathbb{R}$ and let U_i be the set of functions from Θ_i to $\mathbb{R}$. Moreover, let $W\subset\Omega$ be the set of functions from $\{1,\dots,n-1\}\times\Theta$ to $\mathbb{R}$ for which there exist functions $u_i(\theta_i)$ with $w_i(\theta)=\pi(\theta)(u_i(\theta_i)-u_n(\theta_n))$ $\forall i<n\,\forall\theta$. Now consider the following minimization problem

$$\begin{aligned}
&\min_{u_1\in U_1,\dots,u_n\in U_n}\sum_{i<n}\sum_\theta[\pi(\theta)(v_i(\theta)-v_n(\theta))-\pi(\theta)(u_i(\theta)-u_n(\theta))]^2\\
&=\min_{w\in W}\sum_{i<n}\sum_\theta[\tilde{v}_i(\theta)-w_i(\theta)]^2,
\end{aligned}$$

where $\tilde{v}_i(\theta)=\pi(\theta)(v_i(\theta)-v_n(\theta))$. Note that W is a linear subspace of Ω and hence the solution $\hat{w}$ to the above minimization problem is the orthogonal projection of $\tilde{v}\in\Omega$ onto W (all spaces are finite-dimensional and so existence is not an issue). Let $\hat{\varepsilon}=\tilde{v}-\hat{w}$ be the projection residual. Note that the optimal value of the minimization problem is zero if and only if (4.4) holds. By assumption, (4.4) is violated and hence in particular $\hat{\varepsilon}$ must be nonzero. Moreover, since $\hat{\varepsilon}$ is orthogonal to W, for any $h \in W$ it must hold that

$$\sum_{i<n}\sum_{\theta} h_i(\theta)\hat{\varepsilon}(\theta) = 0.$$

We will now use $\hat{\varepsilon}$ to construct a profitable mechanism. Let $\underline{\hat{\varepsilon}} = \min_{i<n,\theta} \hat{\varepsilon}_i(\theta)$ and let

$$\hat{z}_i(\theta) = \hat{\varepsilon}_i(\theta) - \underline{\hat{\varepsilon}} \quad \forall i < n \,\forall \theta.$$

By construction, $\hat{z} \in \Omega$ is nonnegative. Define

$$\hat{x}_i(\theta) = \alpha \hat{z}_i(\theta),$$

where $\alpha > 0$ is chosen sufficiently small such that $\sum_{i<n} \hat{x}_i(\theta) \leq 1$ for all θ. Also, define $\hat{x}_n(\theta) = 1 - \sum_{i<n} \hat{x}_i(\theta)$. Then $\hat{x}$ is a feasible mechanism.

In the remainder of the proof we show that $\hat{x}$ is a profitable mechanism. We first verify that $\hat{x}$ is IC. Let $j < n$ be an agent. Then for any report θ_j' it holds that

$$\begin{aligned}
\sum_{\theta_{-j}} \pi_{-j}(\theta_{-j})\hat{x}_j(\theta_j', \theta_{-j}) &= \sum_{i<n}\sum_{\theta} \pi(\theta)\frac{1}{\pi_j(\theta_j)} 1(\theta_j = \theta_j') 1(i = j)\hat{\varepsilon}_i(\theta) - \alpha\underline{\hat{\varepsilon}} \\
&= -\alpha\underline{\hat{\varepsilon}},
\end{aligned}$$

because the function $\pi(\theta)\frac{1}{\pi_j(\theta_j)}1(\theta_j = \theta_j')1(i = j)$ lies in W and the function $\hat{\varepsilon}_i(\theta)$ is orthogonal to W. Since $-\alpha\underline{\hat{\varepsilon}}$ does not depend on θ_j', $\hat{x}$ is IC for agent j. It remains to check IC for agent n. Let θ_n' be a report. Then

$$\begin{aligned}
\sum_{\theta_{-n}} \pi_{-n}(\theta_{-n})\hat{x}_n(\theta_n', \theta_{-n}) &= \sum_{\theta_{-n}} \pi_{-n}(\theta_{-n})(1 - \sum_{i<n} \hat{x}_i(\theta_n', \theta_{-n})) \\
&= 1 + (n-1)\alpha\underline{\hat{\varepsilon}} - \alpha\sum_{i<n}\sum_{\theta}\pi(\theta)\frac{1}{\pi_n(\theta_n)}1(\theta_n = \theta_n')\hat{\varepsilon}_i(\theta) \\
&= 1 + (n-1)\alpha\underline{\hat{\varepsilon}},
\end{aligned}$$

because the function $\pi(\theta)\frac{1}{\pi_n(\theta_n)}1(\theta_n = \theta_n')$ lies in W and the function $\hat{\varepsilon}_i(\theta)$ is orthogonal to W. Hence $\hat{x}$ is an IC mechanism. It only remains to show that the principal's expected payoff from $\hat{x}$ is greater than $\hat{v}$.

The principal's expected payoff from $\hat{x}$ is

$$\begin{aligned}
&\sum_i\sum_\theta \pi(\theta)v_i(\theta)x_i(\theta) \\
=&\sum_{i<n}\sum_\theta \pi(\theta)(v_i(\theta) - v_n(\theta))\hat{x}_i(\theta) + \sum_\theta \pi(\theta)v_n(\theta)\sum_i \hat{x}_i(\theta) \\
=&\sum_{i<n}\sum_\theta \tilde{v}_i(\theta)\hat{x}_i(\theta) + \bar{v} \\
=&\alpha\sum_{i<n}\sum_\theta (\hat{w}_i(\theta) + \hat{\varepsilon}_i(\theta))\hat{\varepsilon}_i(\theta) - \alpha\sum_i\sum_\theta \tilde{v}_i(\theta)\underline{\hat{\varepsilon}} + \bar{v}.
\end{aligned}$$

By assumption, $\sum_\theta \pi(\theta)v_i(\theta)$ is the same for all i and hence for any $i<n$: $\sum_\theta \tilde{v}_i(\theta)=0$. This means that the second term in the last line above is zero. Because in addition $\hat{w}\in W$ and $\hat{\varepsilon}$ is orthogonal to W, the principal's expected payoff now simplifies to

$$\alpha\sum_{i<n}\sum_{\theta}\hat{\varepsilon}_i(\theta)^2+\bar{v}.$$

By assumption $\tilde{v}$ does not lie in W and so the projection residual $\hat{\varepsilon}$ is nonzero. It follows that the first term above is positive and therefore that the principal's expected payoff from $\hat{x}$ is greater than $\bar{v}$. That is to say, $\hat{x}$ is a profitable mechanism. □

4.A.8 Proof of Proposition 4.7

Proof. The result follows from Proposition 4.6 by interpreting the disposal option as an additional agent. Formally, let there be an agent 0 with a singleton type space $\Theta_0=\{\theta^0\}$ and $v_0\equiv 0$. A mechanism without disposal in this setting corresponds to a mechanism with disposal in the original setting. By Proposition 4.6, a profitable mechanism without disposal exists in the setting with the additional agent if and only if there do not exist functions $u_i(\theta_i)$ $(i=0,\dots,n)$ such that $v_i(\theta)-v_0(\theta)=u_i(\theta_i)-u_0(\theta_0)\ \forall i>0\,\forall\theta$. Since Θ_0 is a singleton and $v_0\equiv 0$ the condition simplifies the following: A profitable mechanism exists if and only there do not exist functions $u_1(\theta_1),\dots,u_n(\theta_n)$ and a constant c such that $v_i(\theta)=u_i(\theta_i)-c\ \forall i>0\,\forall\theta$. But the latter simply means that there does not exist an agent $i>0$ such that $v_i(\theta_i,\theta_{-i})$ is not constant in θ_{-i}. □

References

Aumann, Robert J. 1974. "Subjectivity and correlation in randomized strategies." *Journal of Mathematical Economics* 1 (1): 67–96. [142]

Aumann, Robert J. 1987. "Correlated equilibrium as an expression of Bayesian rationality." *Econometrica: Journal of the Econometric Society,* 1–18. [142]

Becker, Gary S. 1973. "A theory of marriage: Part I." *Journal of Political Economy* 81 (4): 813–46. [140]

Bergemann, Dirk, and Stephen Morris. 2016. "Bayes correlated equilibrium and the comparison of information structures in games." *Theoretical Economics* 11 (2): 487–522. [142]

Blackwell, David. 1951. "Comparison of Experiments." In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability,* 93–102. University of California Press. [142]

Blackwell, David. 1953. "Equivalent comparisons of experiments." *Annals of Mathematical Statistics,* 265–72. [142]

Börgers, Tilman, Angel Hernando-Veciana, and Daniel Krähmer. 2013. "When are signals complements or substitutes?" *Journal of Economic Theory* 148 (1): 165–95. [142]

Börgers, Tilman, and Peter Postl. 2009. "Efficient compromising." *Journal of Economic Theory* 144 (5): 2057–76. [141]

Brualdi, Richard A. 2006. *Combinatorial matrix classes.* Vol. 13. Cambridge University Press. [139]

Crémer, Jacques, and Richard P McLean. 1985. "Optimal selling strategies under uncertainty for a discriminating monopolist when demands are interdependent." *Econometrica* 53: 345–61. [129, 134, 135, 141]

Crémer, Jacques, and Richard P McLean. 1988. "Full extraction of the surplus in Bayesian and dominant strategy auctions." *Econometrica: Journal of the Econometric Society,* 1247–57. [129, 134, 141]

Feng, Tangren, and Qinggong Wu. 2019. "Getting Information from Your Enemies." [142]

Gershkov, Alex, Jacob K Goeree, Alexey Kushnir, Benny Moldovanu, and Xianwen Shi. 2013. "On the equivalence of Bayesian and dominant strategy implementation." *Econometrica* 81 (1): 197–220. [142]

Goldlücke, Susanne, and Thomas Tröger. 2020. "The multiple-volunteers principle." *Available at SSRN 3753985.* [142]

Hardy, Godfrey H, John E Littlewood, and George Pólya. 1952. *Inequalities.* Cambridge University Press. [140]

Hoffman, Alan J. 1963. "On simple linear programming problems." In *Proceedings of Symposia in Pure Mathematics,* 7: 317–27. [140]

Kattwinkel, Deniz, Axel Niemeyer, Justus Preusser, and Alexander Winter. 2022. "Mechanisms without transfers for fully biased agents." https://arxiv.org/abs/2205.10910. [129]

Kim, Semin. 2017. "Ordinal versus cardinal voting rules: A mechanism design approach." *Games and Economic Behavior* 104: 350–71. [141]

Manelli, Alejandro M, and Daniel R Vincent. 2010. "Bayesian and Dominant-Strategy Implementation in the Independent Private-Values Model." *Econometrica* 78 (6): 1905–38. [142]

Myerson, Roger B. 1981. "Optimal auction design." *Mathematics of Operations Research* 6 (1): 58–73. [141]

Rosenthal, Robert W. 1974. "Correlated equilibria in some classes of two-person games." *International Journal of Game Theory* 3 (3): 119–28. [142]

Villani, Cédric. 2009. *Optimal transport: old and new.* Vol. 338. Springer. [138]

Vince, A. 1990. “A rearrangement inequality and the permutahedron.” *American Mathematical Monthly* 97 (4): 319–23. [140]