Robust mechanism design and dominant strategy voting rules

Borgers, Tilman and Smith, Doug
University of Michigan

03. November 2011

ROBUST MECHANISM DESIGN AND DOMINANT STRATEGY VOTING RULES

TILMAN BÖRGERS

Department of Economics, University of Michigan, Ann Arbor, MI

DOUG SMITH

Federal Trade Commission, Washington, DC

Abstract. We develop an analysis of voting rules that is robust in the sense that we do not make any assumption regarding voters' knowledge about each other. In dominant strategy voting rules, voters' behavior can be predicted *uniquely* without making any such assumption. However, on full domains, the only dominant strategy voting rules are random dictatorships. We show that the designer of a voting rule can achieve Pareto improvements over random dictatorship by choosing rules in which voters' behavior can depend on their beliefs. The Pareto improvement is achieved for all possible beliefs. The mechanism that we use to demonstrate this result is simple and intuitive, and the Pareto improvement result extends to all equilibria of the mechanism that satisfy a mild refinement. We also show that the result only holds for voters' interim expected utilities, not for their ex post expected utilities.

1. Introduction

In this paper we consider the design of voting rules from the perspective of the theory of robust mechanism design. Our starting point is the classic result due to Gibbard [15] and Satterthwaite [21] according to which the only dominant strategy voting rules for three or more alternatives are dictatorial voting rules. Gibbard and Satterthwaite assumed

---

E-mail addresses: tborgers@umich.edu, dsmith2@ftc.gov.

Date: November 3, 2011.

This paper is a chapter of Doug Smith's dissertation at the University of Michigan. We thank Stephan Lauermann and Arunava Sen for helpful discussions. The views expressed in this article are those of the authors and do not necessarily reflect those of the Federal Trade Commission.

the number of alternatives to be finite. Preferences were modeled as complete and transitive orders of the set of alternatives. For every voter the range of relevant preferences was taken to be the set of *all* possible preferences over the alternatives. Gibbard and Satterthwaite then asked whether it is possible to construct a game form[1] that determines which alternative is selected as a function of the strategies chosen by the voters, such that each voter has a dominant strategy whatever that voter's preferences are. A dominant strategy was defined to be a strategy that is a best reply to each of the other voters' strategy combinations. Gibbard and Satterthwaite showed that the only game forms that offer each voter for all preferences a dominant strategy are game forms that leave the choice of the outcome to just one individual, the dictator.[2]

One motivation for the interest in dominant strategy mechanisms is that dominant strategies predict rational voters' behavior without relying on any assumption about the voters' beliefs about each others' preferences or behavior. If a voter does not have a dominant strategy, then that voter's optimal choice depends on his beliefs about other voters' behavior which in turn may be derived from beliefs about other voters' preferences. It seems at first sight attractive to bypass such beliefs, and to construct a game form for which a prediction can be made that is independent of beliefs.

On closer inspection, this argument can be seen to consist of two parts:

> (A) *The design of a good game form for voting should not be based on specific assumptions about voters' beliefs about each other.*
> (B) *A good game form for voting should allow us to predict rational voters' choices uniquely from their preferences, without making specific assumptions about these voters' beliefs about each other.*[3]

Both parts to the argument have their own appeal. Voting schemes are often constructed long before the particular contexts in which they will be used are known. It seems wise not to make any special assumptions about agents' knowledge about each other, motivating part (A)

---

[1]We use the terms *game form* and *mechanism* synonymously.

[2]The literature that builds on Gibbard and Satterthwaite's seminal work is voluminous. For a recent survey see Barberà [2].

[3]Blin and Satterthwaite [7] emphasize the interpretation of the Gibbard Satterthwaite theorem as a result about voting procedures in which each voter's choice depends only on their preferences, and not on their beliefs about others' preferences.

of the argument. Part (B) can perhaps be motivated by the idea that game forms in which voters' behavior can be uniquely predicted independent of their beliefs confront voters with simpler strategic problems than game forms in which voters' rational behavior is belief dependent.

As the Gibbard-Satterthwaite theorem shows, (A) and (B) together impose strong restrictions on a voting scheme. In this paper we maintain (A), but drop (B). In other words, we examine game forms for voting without making assumptions about voters' beliefs about each other, but we do not restrict attention to game forms for which voters' equilibrium strategies are independent of voters' beliefs.[4] In our formal model we avoid any assumptions about voters' beliefs abut each other by analyzing any proposed game form for *all* possible type spaces. For each type space we look for a Bayesian equilibrium of the given game form for that type space.[5] However, we do not require each voter's choice, for a given preference of that voter, to be the same for all type spaces.

In order to be able to use the notion of *Bayesian equilibrium* we need to introduce a framework that is slightly different from Gibbard and Satterthwaite's framework. We model voters' attitudes towards risk, assuming that they maximize expected utility. A version of Gibbard and Satterthwaite's theorem for expected utility maximizers has been shown by Hylland [17]. Hylland assumed that voters have von Neumann Morgenstern utilities, and that lotteries are allowed as outcomes. He characterized game forms that offer each agent for every utility function a dominant strategy, that pick an alternative with probability 1 if it is unanimously preferred by all agents, and that pick an alternative with probability 0 if it is unanimously ranked lowest by all agents. He showed that the only such game forms are random dictatorships.[6] In random dictatorships each voter i gets to be dictator with a probability p_i that is independent of all preferences. If voter i is dictator, then the outcome that voter i ranks highest is chosen.

The two main results of this paper address whether there are game forms such that for all finite type spaces, there is at least one Bayesian

[4]Our work is thus in a sense complementary to the work surveyed by Barberà [2] that insists on dominant strategies (B), but seeks to obtain more positive results than Gibbard and Satterthwaite by restricting the domain of preferences that is considered.

[5]For the definition of *type spaces* see Bergemann and Morris [5].

[6]This result is Theorem 1* in Hylland [17]. It is also Theorem 1 in Dutta et. al. [12] (see also [13]) where an alternative proof is provided. Another proof is in Nandeibam [20].

equilibrium of the game form that yields all voters' types the same expected utility, and in some type spaces, for some voters' types, strictly higher expected utility than random dictatorship. Obviously, the answer to this question can be positive only when each voter's probability of being dictator is strictly less than one. In our first main result we show that in this case the answer to our question is indeed positive, provided that we consider *interim* expected utility, that is, each voter's expected utility is calculated when that voter's type is known, but the other voters' types are not yet known.[7] If an *ex post* perspective is adopted instead, that is, if voters' expected utility is considered conditional on the vector of *all* voters' types, then no voting game form Pareto improves on random dictatorship. This is our second main result.

We show the first main result using a simple game form that allows voters to avoid random dictatorship and implement a compromise whenever all voters agree that the compromise is preferable to random dictatorship. It will be easy to see that our first main result can be extended, and we can show that not just one, but *all* Bayesian equilibria of the game form that we are proposing, if they satisfy a mild refinement, yield for all voters at least as high expected utility as random dictatorship.[8] The compromise option may not turn out to be a Pareto improvement ex post as agents may compromise because they think it likely that the compromise will improve on random dictatorship, but ex post discover to have a type vector that appeared ex ante unlikely, and for which the compromise is not a Pareto improvement. The second main result shows that *any* game form will sometimes make some type worse off in comparison to random dictatorship.

Section 2 discusses related literature. Section 3 explains the model and the definitions used in this paper. In Section 4 we adapt Hylland's theorem on random dictatorship to our setting. In Section 5 we explain how we relax the requirement that voters' choices, for given preferences, are the same in all type spaces. Sections 6 and 7 contain our two main results. Section 8 concludes.

---

[7]The notions of interim and ex post efficiency are due to Holmström and Myerson [18].

[8]Our first, positive result thus is thus in the spirit of the literature on *full implementation*, which considers all equilibria of a game form, whereas our second, negative result is in the spirit of the literature on *mechanism design* which considers only some equilibrium of a given game form. Both results are stronger than they would be if the respectively other approach were used. For the distinction between implementation and mechanism design see, for example, Jackson [19].

2. Related Literature

Our approach builds on Bergemann and Morris's [4] seminal work on robust mechanism design.[9] They consider, as we do, Bayesian equilibria of mechanisms on *all* type spaces. Bergemann and Morris seek conditions under which the Bayesian implementability of a social choice correspondence on all type spaces implies dominant strategy implementability (or, more generally, implementability in *ex post equilibria*). The conditions that they find apply to *separable environments*, the prime example of which are environments in which each agent's utility depends on some physical allocation and this agent's monetary transfer. Bergemann and Morris point out [4, Section 6.3] that in non-separable environments, such as the environment without transferrable payoffs considered by Gibbard and Satterthwaite, dominant strategy implementability may be a stronger requirement than Bayesian implementability on all type spaces.[10] Bergemann and Morris do not consider the problem of comparing different mechanisms if the mechanism designer cannot fully achieve his or her objectives. Such comparisons are the focus of our work.

The approach that we take in this paper to comparing different mechanisms is based on Smith [23] who studies the design of a mechanism for public goods. Smith considers the performance of different mechanisms in a Bayesian equilibrium on all type spaces. He focuses on an ex post perspective, and demonstrates that a mechanism designer can improve efficiency using a more flexible mechanism than a dominant strategy mechanism. In our setting, by contrast, we find that no mechanism can improve on dominant strategy mechanisms ex post, but that such an improvement is possible from an interim perspective.

Chung and Ely [11] describe an auctioneer of a single object who designs an auction to maximize expected revenues. The auctioneer considers equilibria of different auction mechanisms on the universal type space, and evaluates different mechanisms using a maximin criterion: taking the distribution of the agents' valuations, but not the agents' beliefs, as given, for each mechanism the auctioneer determines the probability distribution on the universal type space for which that mechanism yields the lowest expected revenue. The auctioneer then chooses a mechanism that maximizes the lowest expected revenue. Aside from

[9] The literature on robust mechanism design and implementation was recently surveyed by Bergemann and Morris [5].

[10] The discussion paper version [3] of Bergemann and Morris [4] also includes a general characterization of Bayesian implementability on all type spaces, however we do not make use of this characterization.

the obvious differences in setting, the main conceptual difference from our work is that our mechanism designer has only a partial order of mechanisms, whereas Chung and Ely's mechanism designer has a complete order. Our order is based on comparing mechanisms on every type space, and ranking one mechanism above another if it performs according to the designer's objectives at least as good on all type spaces, and on some better. For this order we find, unlike Chung and Ely, that in our setting there are mechanisms that perform better than dominant strategy mechanisms.

Whereas the papers cited so far are concerned with mechanism design, in the sense that for any given mechanism and type space only one Bayesian equilibrium is considered, there is also a literature on robust implementation, in which for any given mechanism and type space all Bayesian equilibria are taken into account. Bergemann and Morris [6] provide conditions for a social choice function to be implementable on every type space.

A recent paper by Yamashita [24] is related to the idea of robust implementation. Yamashita considers a bilateral trade setting, and evaluates mechanisms on the basis of the lowest expected welfare among all outcomes that can result if agents use strategies that are not weakly dominated. Expected welfare is calculated on the basis of the mechanism designer's subjective prior over agents' types. Yamashita's work is similar to work on implementation because he considers all outcomes, not just some outcomes, that can result under a solution concept. A predecessor to Yamashita [24] is Börgers [8] who considered in the Gibbard-Satterthwaite framework the existence of mechanisms for which the outcomes that result if all players choose a strategy that is not weakly dominated are Pareto efficient, and (in a sense defined in that paper) less one-sided than the outcomes of dictatorship. Börgers showed the existence of such mechanisms. Börgers used a framework in which agents' preferences were modeled using ordinal preferences rather than von Neumann Morgenstern utilities.

Bayesian mechanism design approaches to voting are surprisingly rare in the literature. For the case of independent types, Azrieli and Kim [1] have recently considered interim and ex ante efficiency in a setting with two alternatives and independent types. Schmitz and Tröger [22] consider the same issue and allow correlated types. Börgers and Postl [9] study ex ante welfare maximization in a setting with three alternatives. The type space in their paper is very small, with the ordinal ranking of alternatives being common knowledge, and only the cardinal utility functions private information.

The game form that we use to prove our first main result, that random dictatorship can be improved upon from an interim perspective, is almost identical to the *Full Consensus or Random Ballot Fall-Back* game form that Heitzig and Simmons [16] have introduced. While their motivation, like ours, is to consider voting systems that are more flexible than dictatorial voting systems, and that allow for compromises, the focus of their formal analysis is on complete information, correlated equilibria that are in some sense coalition proof. In our paper the focus is on analyzing Bayesian equilibria in arbitrary type spaces.

3. The Voting Problem

There are n agents: $i \in I = \{1, 2, \ldots, n\}$. The agents have to choose one alternative from a finite set A of alternatives which has at least three elements. The set of all probability distributions over A is $\Delta(A)$, where for $\delta \in \Delta(A)$ and $a \in A$ we denote by $\delta(a) \in [0, 1]$ the probability that δ assigns to a. The agents are commonly known to be expected utility maximizers. We denote agent i's von Neumann Morgenstern utility function by $u_i : A \to \mathbb{R}$. We assume that $a \neq b \Rightarrow u_i(a) \neq u_i(b)$, i.e., there are no indifferences. We define the expected utility for probability distributions $\delta \in \Delta(A)$ by $u_i(\delta) = \sum_{a \in A} u_i(a)\delta(a)$.

A mechanism designer has a (possibly incomplete) ranking of the alternatives in A that may depend on the agents' utility functions. We shall be more specific about the designer's objectives later. The mechanism designer does not know the agents' utility functions, nor does she know what the agents believe about each other. To implement an outcome that potentially depends on the agents' utility functions the mechanism designer asks the agents to play a *game form.*

Definition 1. *A* game form $G = (S, x)$ *consists of:*

(i) *a set* $S \equiv \prod_{i \in I} S_i$ *where for every* $i \in I$ *the set* S_i *is non-empty and finite;*

(ii) *a function* $x : S \to \Delta(A)$.

The set S_i is the set of (pure) strategies available to agent i in the game form G. We focus on finite sets of pure strategies, while allowing mixed strategies, to ease exposition. Our results also hold when the sets S_i of pure strategies are allowed to be infinite. The function x assigns to every combination of pure strategies s the potentially stochastic outcome $x(s)$ that is implemented when agents choose that combination of pure strategies. We write $x(s, a)$ for the probability that $x(s)$ assigns to alternative a.

Once the mechanism designer has announced a game form, the agents choose their strategies simultaneously and independently. Because the agents don't necessarily know each others' utility functions or beliefs, this game may be a game of incomplete information. A hypothesis about the agents' utility functions and their beliefs about each other can be described by a *type space.*

Definition 2. *A* type space $\mathcal{T} = (T, \pi, u)$ *consists of:*

(i) *a set* $T \equiv \prod_{i \in I} T_i$*, where for every* $i \in I$ *the set* T_i *is non-empty and finite;*

(ii) *an array* $\pi = (\pi_1, \pi_2, \dots, \pi_n)$ *of functions* $\pi_i : T_i \to \Delta(T_{-i})$ *where* $\Delta(T_{-i})$ *is the set of all probability distributions over* $T_{-i} \equiv \prod_{j \neq i} T_j$*;*

(iii) *an array* $u = (u_1, u_2, \dots, u_n)$ *of functions* $u_i : T_i \times A \to \mathbb{R}$ *such that* $a \neq b \Rightarrow u_i(t_i, a) \neq u_i(t_i, b)$ *for all* $t_i \in T_i$.

The set T_i is the set of types of agent i. We assume type spaces to be finite to avoid technical difficulties, in particular in the proof of Proposition 2. Finiteness rules out the universal type space. Our construction has the advantage that, unlike the universal type space, spurious types, i.e. multiple types with identical hierarchies of beliefs, are allowed. Ely and Pesky [14, Section 1.2] point out the possible importance of spurious types when analyzing Bayesian equilibria.

Agent i privately observes his type. The function π_i describes for every type $t_i \in T_i$ the beliefs that agent i has about the other agents' types when agent i himself is of type t_i. We write $\pi_i(t_i, t_{-i})$ for the probability that type t_i assigns to the other players types being t_{-i}. Beliefs are subjective. There may or may not be a common prior for a particular type space. Different agents' beliefs may be incompatible with each other in the sense that one agent may attach positive probability to an event to which another agent attaches probability zero. The function $u_i(t_i)$ describes player i's utility when i is of type t_i. We write $u_i(t_i, a)$ for the utility that $u_i(t_i)$ assigns to alternative a. The utility functions $u_i(t_i)$ satisfy the assumption that we introduced earlier that there are no indifferences.[11]

We assume that the mechanism designer has no knowledge of the agents' utility functions or their beliefs. Therefore, the mechanism designer regards all type spaces as possible descriptions of the environment. We denote the set of all type spaces by Υ.

---

[11]Observe that we suppress in the notation the dependence of π_i and u_i on the type space $\mathcal{T}$. No confusion should arise from this simplification of our notation.

The mechanism designer proposes to agents how they might play the game. For the agents to accept the mechanism designer's proposal, she must propose a *Bayesian equilibrium.* Because the mechanism designer does not know the true type space, she has to propose a *Bayesian equilibrium for every type space.*

Definition 3. *A* Bayesian equilibrium of game form G for every type space *is an array* $\sigma^* = (\sigma_1^*, \sigma_2^*, \ldots, \sigma_n^*)$ *such that for every* $i \in I$*:*

(i) σ_i^* *is a family of functions* $(\sigma_i^*(\mathcal{T}))_{\mathcal{T}\in\Upsilon}$ *where for every* $\mathcal{T} \in \Upsilon$ *the function* $\sigma_i^*(\mathcal{T})$ *maps the type space* T_i *corresponding to* $\mathcal{T}$ *into* $\Delta(S_i)$*, the set of all probability distributions on* S_i*;*

and, writing $\sigma_i^*(\mathcal{T}, t_i)$ *for the mixed strategy assigned to* t_i*, and writing* $\sigma_i^*(\mathcal{T}, t_i, s_i)$ *for the probability that this mixed strategy assigns to* $s_i \in S_i$*, we have for every* $\mathcal{T} \in \Upsilon$*,* $i \in I$*, and* $t_i \in T_i$ *(where* T_i *corresponds to* $\mathcal{T}$*):*

(ii) $\sigma_i^*(\mathcal{T}, t_i)$ *maximizes the expected utility of type* t_i *among all mixed strategies in* $\Delta(S_i)$*, where expected utility for any mixed strategy* $\sigma_i \in \Delta(S_i)$ *is:*

$$\sum_{t_{-i}\in T_{-i}} \pi_i(t_i, t_{-i}) \sum_{s\in S} u_i(t_i, x(s)) \cdot \sigma_i(s_i) \cdot \prod_{j\neq i} \sigma_j^*(\mathcal{T}, t_j, s_j). \tag{1}$$

The mechanism designer evaluates different mechanisms and their equilibria using the Pareto criterion. When evaluating the agents' utility for a realized type combination t the mechanism designer can either only consider the outcomes that result from the mixed strategies prescribed for these types, or she may consider the expected utilities of these types, based on the types' own subjective beliefs. In other words, the mechanism designer may adopt an *ex post* or an *interim* perspective when evaluating agents' utilities. The interim perspective respects agents' own perception of their environment. The ex post perspective has a paternalistic flavor. On the other hand, for example when agents' beliefs are incompatible with each other, the mechanism designer may be justified in discarding agents' beliefs, on the basis that at least some of them have to be wrong, as agents themselves will discover at some point. Thus neither the interim nor the ex post perspective seem clearly preferable. We pursue both perspectives in this paper.

Definition 4. *The game form* G *and the Bayesian equilibrium for all type spaces* σ^* interim Pareto dominate *the game form* $\tilde{G}$ *and the Bayesian equilibrium for all type spaces* $\tilde{\sigma}^*$ *if for all* $\mathcal{T} \in \Upsilon$*,* $i \in I$*, and*

$t_i \in T_i$:

$$\sum_{t_{-i}\in T_{-i}} \pi_i(t_i,t_{-i}) \sum_{s\in S} u_i(t_i,x(s)) \cdot \prod_{j\in I} \sigma_j^*(\mathcal{T},t_j,s_j) \geq$$
$$\sum_{t_{-i}\in T_{-i}} \pi_i(t_i,t_{-i}) \sum_{s\in S} u_i(t_i,\tilde{x}(s)) \cdot \prod_{j\in I} \tilde{\sigma}_j^*(\mathcal{T},t_j,s_j) \tag{2}$$

with strict inequality for at least one $\mathcal{T} \in \Upsilon$, $i \in I$, and $t_i \in T_i$.

Definition 5. *The game form G and the Bayesian equilibrium for all type spaces σ^** ex post Pareto dominate *the game form $\tilde{G}$ and the Bayesian equilibrium for all type spaces $\tilde{\sigma}^*$ if for all $\mathcal{T} \in \Upsilon$, $i \in I$, and $t \in T$:*

$$\sum_{s\in S} u_i(t_i,x(s)) \cdot \prod_{j\in I} \sigma_j^*(\mathcal{T},t_j,s_j) \geq$$
$$\sum_{s\in S} u_i(t_i,\tilde{x}(s)) \cdot \prod_{j\in I} \tilde{\sigma}_j^*(\mathcal{T},t_j,s_j) \tag{3}$$

with strict inequality for at least one $\mathcal{T} \in \Upsilon$, $i \in I$, and $t \in T$.

Our main interest in this paper is in exploring how the mechanism designer'a ability to achieve her objective depends on additional conditions that Bayesian equilibria of the mechanism designer's proposed game form have to satisfy. In the next section, we consider the very restrictive requirement of belief independence. In the subsequent sections, we relax this requirement.

4. Belief Independent Equilibria: Hylland's Theorem

We begin by exploring the consequences of a restrictive requirement for the Bayesian equilibrium that the mechanism designer proposes. This requirement is implicit in the work on dominant strategy mechanism design. It is that equilibria be *belief independent*. Using the notion of belief independent equilibria, we can restate Hylland's version of the Gibbard Satterthwaite theorem in our setting.

Definition 6. *A Bayesian equilibrium for every type space, σ^*, of a game form G is* belief independent *if for all $i \in I, \mathcal{T}, \widetilde{\mathcal{T}} \in \Upsilon, t_i \in T_i$ and $\tilde{t}_i \in \tilde{T}_i$ such that $u_i(t_i) = \tilde{u}_i(\tilde{t}_i)$ we have:*

$$\sigma_i^*(\mathcal{T},t_i) = \sigma_i^*(\widetilde{\mathcal{T}},\tilde{t}_i), \tag{4}$$

where T_i, u_i correspond to $\mathcal{T}$ and $\tilde{T}_i, \tilde{u}_i$ correspond to $\widetilde{\mathcal{T}}$.

The reformulation of Hylland's theorem presented below says that all game forms and belief independent equilibria of these game forms that satisfy two unanimity requirements are random dictatorships. To define the two unanimity requirements and random dictatorship we need some notation. If u is a utility function, we denote by $b(u)$ the element of A that maximizes u, and by $w(u)$ the element of A that minimizes u.[12]

Definition 7. *A game form G and a Bayesian equilibrium of G for every type space, σ^*, satisfy*

(i) positive unanimity *if for every $\mathcal{T} \in \Upsilon$, $t \in T$, and $a \in A$ such that $b(u_i(t_i)) = a$ for all $i \in I$, we have:*

$$\sum_{s \in S} \prod_{i \in I} \sigma_i^*(\mathcal{T}, t_i, s_i) \cdot x(s,a) = 1; \tag{5}$$

(ii) negative unanimity *if for every $\mathcal{T} \in \Upsilon$, $t \in T$, and $a \in A$ such that $w(u_i(t_i)) = a$ for all $i \in I$, we have:*

$$\sum_{s \in S} \prod_{i \in I} \sigma_i^*(\mathcal{T}, t_i, s_i) \cdot x(s,a) = 0. \tag{6}$$

Positive and negative unanimity are implied by, but weaker than ex post Pareto efficiency. Next, we provide the formal definition of random dictatorship that we need for our reformulation of Hylland's theorem.

Definition 8. *A game form G and a Bayesian equilibrium of G for every type space, σ^*, are a* random dictatorship *if there is some $p \in [0,1]^n$ such that for every $\mathcal{T} \in \Upsilon$, $t \in T$, and $a \in A$:*

$$\sum_{s \in S} \prod_{i \in I} \sigma_i^*(\mathcal{T}, t_i, s_i) \cdot x(s,a) = \sum_{\{i \in I : b(u_i(t_i)) = a\}} p_i. \tag{7}$$

The following is implied by Hylland's theorem.[13]

Proposition 1. *A game form G and a belief-independent Bayesian equilibrium of G for every type space, σ^*, satisfy positive and negative unanimity if and only if they are a random dictatorship.*

Proof. The "if-part" is obvious. To prove the "only if-part" we derive from G and σ^* a "cardinal decision scheme" in the sense of Definition 1 in [12], and show that this cardinal decision scheme has the properties listed in Theorem 1 in [12] and the correction in [13]. It then follows

---

[12]Recall that we have assumed that there are no indifferences. Therefore, there is a unique element of A that maximizes u, and a unique element of A that minimizes u.

[13]Theorem 1* in Hylland [17]. We use here the version of Hylland's theorem that is Theorem 1 in [12] with the correction in [13].

from Theorem 1 in [12] that the cardinal decision scheme is a random dictatorship. This then implies the "only if-part" of our Proposition 1.

Denote by $\mathcal{U}$ the set of all utility functions that have the property of no indifferences (see Definition 2). A cardinal decision scheme is a mapping $\phi : \mathcal{U}^n \to \Delta(A)$. We can derive from G and σ^* a cardinal decision scheme by setting for any $(u_1, u_2, \ldots, u_n) \in \mathcal{U}^n$ and $a \in A$ the probability $\phi(u_1, u_2, \ldots, u_n, a)$ that $\phi(u_1, u_2, \ldots, u_n)$ assigns to a as:

$$\phi(u_1, u_2, \ldots, u_n, a) = \sum_{s \in S} \prod_{i \in I} \sigma_i^*(\mathcal{T}, t_i, s_i) \cdot x(s, a), \tag{8}$$

where we can pick any $\mathcal{T} \in \Upsilon$ and any $t \in T$ such that $u_i(t_i) = u_i$ for all $i \in I$. By belief-independence it does not matter which such $\mathcal{T}$ and $t \in T$ we choose. Then ϕ is a cardinal decision scheme as defined in Definition 1 of [12].

We can complete the proof by showing that ϕ has the two properties listed in Theorem 1 of [12] and the additional property listed in the correction [13]. The first property is unanimity: If $b(u_i) = a$ for all $i \in I$, then $\phi(u_1, u_2, \ldots, u_n, a) = 1$. This is implied by the assumption that G and σ^* satisfy positive unanimity.

The second property is strategy proofness: If $(u_1, u_2, \ldots, u_n) \in \mathcal{U}^n$ and $u_i' \in \mathcal{U}$, then $u_i(\phi(u_i, u_{-i})) \geq u_i(\phi(u_i', u_{-i}))$, where u_{-i} is the array $(u_1, u_2, \ldots, u_n)$ leaving out u_i. To prove this we pick $\mathcal{T} \in \Upsilon$, $t_i, t_i' \in T_i$ and $t_{-i} \in \prod_{j \neq i} T_j$ such that $u_i(t_i) = u_i, u_i(t_i') = u_i'$, and $u_j(t_j) = u_j$ for all $j \neq i$. Moreover, $\pi_i(t_i)$ and $\pi_i(t_i')$ place probability 1 on t_{-i}. Then the fact that σ^* is a Bayesian equilibrium of G for the type space $\mathcal{T}$ implies:

$$\begin{aligned} &\sum_{s \in S} u_i(t_i, x(s)) \cdot \sigma_i^*(\mathcal{T}, t_i, s_i) \cdot \prod_{j \neq i} \sigma_j^*(\mathcal{T}, t_j, s_j) \geq \\ &\sum_{s \in S} u_i(t_i, x(s)) \cdot \sigma_i^*(\mathcal{T}, t_i', s_i) \cdot \prod_{j \neq i} \sigma_j^*(\mathcal{T}, t_j, s_j) \end{aligned} \tag{9}$$

By the definition of ϕ, this is equivalent to: $u_i(\phi(u_i, u_{-i}) \geq u_i(\phi(u_i', u_{-i}))$, that is, strategy proofness.

The third property, introduced in the correction [13], is a property labelled (*) in [13]: If $w(u_i) = a$ for all $i \in I$, then $\phi(u_1, u_2, \ldots, u_n, a) = 0$. This is implied by the assumption that G and σ^* satisfy negative unanimity. □

From now on, when we refer to random dictatorship, we shall mean a specific game form G and a specific equilibrium σ^* of G for every type space.

Definition 9. *For any vector $p \in [0,1]^n$ such that $\sum_{i \in I} p_i = 1$ the following game form G and equilibrium σ^* of G for every type space will be referred to as p-*random dictatorship*:*

(i) $S_i = A$ *for all* $i \in I$*;*

(ii) $x(s,a) = \sum_{\{i \in I : b(u_i(t_i)) = a\}} p_i$ *for all* $s \in S$ *and* $a \in A$*;*

(iii) $\sigma_i^*(\mathcal{T}, t_i, b(u_i(t_i))) = 1$ *for all* $i \in I$, $\mathcal{T} \in \Upsilon$, *and* $t_i \in T_i$.

It is immediate that σ^* is a belief-independent Bayesian equilibrium of G for every type space, and that G and this equilibrium are a random dictatorship. There are other game forms and equilibria that are random dictatorships, but it is without loss of generality to only consider the ones described in Definition 9.

5. Consistent Equilibria

Our main interest in this paper is in considering the implications of relaxing the requirement of belief independence for the Bayesian equilibria of the game form that the mechanism designer chooses. We do not, however, completely dispense with any link between players' strategies in different type spaces. The Bayesian equilibria that we shall investigate need to satisfy a *consistency* requirement. This requirement is implied by, but does not imply belief independence.

Definition 10. *A Bayesian equilibrium of game form G for every type space, σ^*, is* consistent *if for all type spaces $\mathcal{T}, \widetilde{\mathcal{T}} \in \Upsilon$ such that:*

(i) *for every $i \in I$: $\widetilde{T}_i \subseteq T_i$ (where $\widetilde{T}_i$ corresponds to $\widetilde{\mathcal{T}}$ and T_i corresponds to $\mathcal{T}$);*

(ii) *for every $i \in I$ and every $t_i \in T_i$: $\tilde{u}_i(t_i) = u_i(t_i)$ and $\tilde{\pi}_i(t_i) = \pi_i(t_i)$ (where $\tilde{u}_i, \tilde{\pi}_i$ correspond to $\widetilde{\mathcal{T}}$, and u_i, π_i correspond to $\mathcal{T}$),*

we have for every $i \in I$ and every $t_i \in T_i$:

(iii) $\sigma_i^*(\widetilde{\mathcal{T}}, t_i) = \sigma_i^*(\mathcal{T}, t_i)$.

Observe that the type t_i referred to in item (iii) of Definition 10 has the same utility function and hierarchy of beliefs in type space $\mathcal{T}$ and in type space $\widetilde{\mathcal{T}}$. Therefore, the consistency requirement is implied by the assumption that a type's equilibrium choice only depends on that type's utility function and hierarchy of beliefs. One problem with this stronger assumption would be that we would rule out that equilibria in which spurious types play different strategies, the possible importance of which we mentioned in Section 3. Our weaker consistency requirement does allow Bayesian equilibria in which agents with identical hierarchies of beliefs make different choices. Our results would also go through if we made the more demanding assumption.

6. A Game Form that Interim Pareto Dominates Random Dictatorship

The first main result of this paper examines interim Pareto dominance, while the second main result concerns ex post Pareto dominance. The first result says that for every $p \in [0,1]^n$ such that $\sum_{i\in I} p_i = 1$ and $p_i < 1$ for all $i \in I$ there are a game form, and a Bayesian equilibrium of this game form for every type space, that interim Pareto dominate p-random dictatorship. We refer to the dominating game form as *p-random dictatorship with compromise*:

Definition 11. *For every $p \in [0,1]^n$ such that $\sum_{i\in I} p_i = 1$ the following game form is called a p-*random dictatorship with compromise*.*

(i) *For every $i \in I$:*

$$S_i = \mathcal{A} \times \mathcal{R},$$

where $\mathcal{A}$ is the set of all non-empty subsets of A, and $\mathcal{R}$ is the set of all complete strict ordinal rankings of A; we write $s_i = (A_i, R_i) \in S_i$ for a strategy for agent i.

(ii) *If $\bigcap\limits_{i\in I} A_i = \emptyset$ then for all $a \in A$:*

$$x(s,a) = \sum_{\{i\in I: aR_ia' \ \ \forall a'\in A\}} p_i.$$

(iii) *If $\bigcap\limits_{i\in I} A_i \neq \emptyset$ then for all $a \in \bigcap A_i$:*

$$x(s,a) = \sum_{\{i\in I: aR_ia' \ \ \forall a'\in \bigcap\limits_{i\in I} A_i\}} p_i.$$

In words, this game form offers each agent i the opportunity to provide a complete ranking of outcomes R_i, and also a set A_i of "acceptable" alternatives. If there is at least one common element among the sets of acceptable alternatives for all agents, then the mechanism implements random dictatorship (with the preferences described by the R_i) but with the restriction that the dictator can only choose an outcome from the unanimously acceptable alternatives. Otherwise, the mechanism reverts to random dictatorship (with outcomes determined by the highest ranked elements of the R_i). We refer to this game form as *p-random dictatorship with compromise* because it offers agents the opportunity to replace the outcome of p-random dictatorship by a compromise on a mutually acceptable alternative.[14]

It is elementary to verify that a strategy of player i such that for some type t_i we have $b(u_i(t_i)) \notin A_i$ is weakly dominated by the same strategy in which A_i is replaced by $A_i \cup \{b(u_i(t_i))\}$. Moreover, any strategy of some player i such that for some type t_i we have that R_i is not type t_i's true preference over A_i as described by $u_i(t_i)$, is weakly dominated by a strategy such that A_i is left unchanged, but R_i is replaced by a preference ordering that reflects t_i's true preference over A_i. Preferences which player i indicates for alternatives $A \setminus A_i$ are irrelevant for the outcome of the game. These considerations motivate us to restrict attention to "truthful strategies" which we define to be strategies such that $b(u_i(t_i)) \in A_i$, and such that R_i is the true preference according to $u_i(t_i)$, for all types t_i. Note that we have ruled out some, but not necessarily all weakly dominated strategies. In any case, it seems eminently plausible that all players will choose truthful strategies.

In a Bayesian equilibrium for all type spaces in which all players choose truthful strategies, any type's interim expected utility is not smaller than the interim expected utility from p-random dictatorship. This is because a type can always force an outcome that gives at least as high interim expected utility as p-random dictatorship by choosing the truthful strategy for which $A_i = \{b(u_i(t_i))\}$. Note also that all players choosing this strategy for all types will always be a consistent Bayesian equilibrium for all type spaces.

We now show that *p-random dictatorship with compromise* also has a Bayesian equilibrium for all type spaces, in which all players choose truthful strategies, and that interim Pareto dominates random dictatorship. We further show that this equilibrium respects positive and negative unanimity. This clarifies that our result is indeed a consequence of weakening the belief independence requirement, and not of weakening any other property listed in Proposition 1.

---

[14]This game form was inspired by *Approval Voting* (see Brams and Fishburn [10]), which, like our game form, allows voters to indicate "acceptable" alternatives. However, in approval voting the alternative that the largest number of agents regards as acceptable is selected, whereas our game form requires unanimity. Moreover, our game form uses random dictatorship as a fallback, whereas approval voting does not have any such fallback. When p is the uniform distribution, the game form that we consider is closely related to the *Full Consensus or Random Ballot Fall-Back* game form that Heitzig and Simmons [16] introduced. Heitzig and Simmons require the sets A_i to be singletons.

Proposition 2. *For every $p \in [0,1]^n$ such that $\sum_{i \in I} p_i = 1$ and $p_i < 1$ for all $i \in I$, p-random dictatorship with compromise has a consistent Bayesian equilibrium for all type spaces σ^* that interim Pareto dominates p-random dictatorship and that satisfies positive and negative unanimity.*

Proof. We construct the equilibrium σ^* inductively. We begin by considering type spaces $\mathcal{T}$ where for every $i \in I$ the set T_i has exactly one element. In such type spaces it is common knowledge among the agents that agent i has utility function $u_i(t_i)$. We distinguish two cases. The first is that there is some alternative $a \in A$ such that for all $i \in I$ we have:

$$u_i(t_i, a) > \sum_{j \in I} p_j u_i(t_i, b(u_j(t_j))). \tag{10}$$

Observe that the assumption $p_i < 1$ for all $i \in I$ implies that some such type spaces exist. For such type spaces the strategies are:

$$\sigma_i^*(\mathcal{T}, t_i) = (\{b(u_i(t_i)), a\}, R_i) \tag{11}$$

for all $i \in I$, where R_i is agent i's true preference, and where a is some alternative for which (10) holds.[15] These strategies obviously constitute a Nash equilibrium of the complete information game in which agents' preferences are common knowledge. Note that the outcome a then results, and that this outcome strictly Pareto-dominates the outcome of random dictatorship.

For all other type spaces with just a single element for each player the strategies are:

$$\sigma_i^*(\mathcal{T}, t_i) = (\{b(u_i(t_i))\}, R_i) \tag{12}$$

for all $i \in I$ where R_i is again agent i's true preference. We noted already earlier that these strategies constitute a Nash equilibrium, and that the outcome is the same as under p-random dictatorship.

Now suppose we had constructed the equilibrium for all type spaces $\mathcal{T}$ in which all the sets T_i have at most k elements. We first extend the construction to all type spaces $\mathcal{T}$ in which T_1 has at most $k+1$ elements and for $j > 1$ the set T_j has at most k elements. Then we extend the construction to all type spaces $\mathcal{T}$ in which T_1 and T_2 have at most $k+1$ elements and for $j > 2$ the set T_j has at most k elements. The construction can then inductively be continued until it is extended to all type spaces $\mathcal{T}$ in which all the sets T_i have at most $k+1$ elements.

[15] Note that a must be the same for all players.

Suppose first that we are considering a type space $\mathcal{T}$ in which T_1 has at most $k+1$ elements and for $j > 1$ the set T_j has at most k elements. Consider all type spaces $\widetilde{\mathcal{T}}$ that are contained in $\mathcal{T}$, i.e. for which conditions (i) and (ii) of Definition 10 hold, and such that at least for one agent the type set has fewer elements than in $\mathcal{T}$. For such type spaces we define for every $i \in I$ and every $t_i \in \widetilde{T}_i$:

$$\sigma_i^* \left(\mathcal{T}, t_i\right) = \sigma_i^* \left(\widetilde{\mathcal{T}}, t_i\right). \tag{13}$$

By the inductive hypothesis the right hand side of this equation has already been defined. Observe that the right hand side does not depend on the particular choice of $\tilde{\mathcal{T}}$. If a type t_i of player i is contained in player i's type set in two different type spaces $\widetilde{\mathcal{T}}$ and $\widehat{\mathcal{T}}$ that are contained in $\mathcal{T}$ in the sense of Definition 10, then the intersection of these type spaces is also a type space, and by consistency the same strategy is assigned to type t_i in $\widetilde{\mathcal{T}}$ and in $\widehat{\mathcal{T}}$.

If the previous step defines the equilibrium strategy for all types in $\mathcal{T}$, then the inductive step is completed. Otherwise, it remains to define strategies for types t_i that are not contained in any type set of a type space that is a subspace of $\mathcal{T}$. We consider the strategic game in which each such type is a separate player, and expected utilities are calculated keeping the strategies of types that have already been dealt with in the previous paragraph fixed, and using each type's subjective beliefs to calculate that type's expected payoff. We restrict attention to truthful strategies, and strategies such that $w(u_i(t_i)) \notin A_i$. This strategic game has a Nash equilibrium in mixed strategies, and this Nash equilibrium is also a Nash equilibrium of the game with unrestricted strategy spaces. We define for each type t_i that still has to be dealt with the strategy $\sigma_i^*(\mathcal{T}, t_i)$ to be type t_i's equilibrium strategy.

By construction these strategies satisfy the consistency requirement. Also, they are by construction interim Bayesian equilibria: For types in typesets that correspond to a smaller type space the Bayesian equilibrium property carries over from the smaller type space. For all other types, their choices maximize expected utility by construction. We extend the construction to all type spaces $\mathcal{T}$ in which T_1 and T_2 have at most $k+1$ elements and for $i > 2$ the set T_i has at most k elements in the same way, and continue inductively, until the construction is extended to all type spaces $\mathcal{T}$ in which all the sets T_i have at most $k+1$ elements.

The equilibrium that we have constructed interim Pareto dominates random dictatorship. First, we note that no type can have lower expected utility than under random dictatorship. This is because each type can guarantee themselves an outcome that is at least as good as the random dictatorship outcome by choosing $A_i = \{b(u_i(t_i))\}$. Second, each type's expected utility is increased on type spaces in which each player's type set has just a single element, and for which inequality (10) holds.

The equilibrium that we have constructed satisfies positive unanimity because all players include their most preferred alternative in the set A_i. If all players have the same most preferred alternative, the sets A_i will have a non-empty intersection, and the random dictator will select the alternative that is most preferred by everyone. The equilibrium also satisfies negative unanimity. We have assumed that no player includes their least preferred alternative in the set A_i. Therefore, independent of whether these sets have a non-empty intersection or not, the random dictator will not select the agents' least preferred alternative, if they all have the same least preferred alternative. □

It is immediate that the proof of Proposition 2 also proves a claim that is slightly different from Proposition 2, and that we state below as Remark 1.

Remark 1. *If $p_i < 1$ for all $i \in I$, then every Bayesian equilibrium of p-random dictatorship for all type spaces in which all players choose truthful strategies, and in which players choose strategies of the form (11) in type spaces in which each player's type set has just a single element and for which inequality (10) holds,*[16] *interim Pareto-dominates p-random dictatorship.*

Note that the set of truthful strategies includes in particular the set of all non-weakly dominated strategies. Therefore, Remark 1 applies to all equilibria in non-weakly dominated strategies. As we explained in the Introduction, the importance of this result is that it shows that p-random dictatorship interim Pareto dominates p random dictatorship not just in the sense of mechanism design (where we only have to find one equilibrium with the desired properties) but also in the sense of implementation (where all equilibria (or, as in our case, all that satisfy some refinement) are considered). Observe, incidentally, that for the

[16]In words this condition says: if utility functions are common knowledge, and some alternative Pareto-dominates random dictatorship, then players pick their preferred alternative, and some alternative that Pareto dominates random dictatorship, the same for all players, as their set of acceptable alternatives.

result of Remark 1 the Bayesian equilibrium for all type spaces does not need to be consistent.

7. No Game Form Ex Post Pareto Dominates Random Dictatorship

In this section we show that the result of the previous section does not hold if utilities are evaluated ex post. The following proposition shows that in fact no mechanism ex post Pareto dominates p-random dictatorship.

Proposition 3. *For every $p \in [0,1]^n$ such that $\sum_{i \in I} p_i = 1$ there is no game form G that has a consistent equilibrium for all type spaces σ^* that ex post Pareto dominates p-random dictatorship.*

To prove this result we consider an equilibrium of a game form where, for some type vector, the type of some players $i \in I$ obtains her preferred alternative ex post with a probability that is smaller than the probability with which this alternative would be chosen under random dictatorship. We then use incentive compatibility arguments to infer how the ex post allocation will change as we change the types of players $j \notin I$, leaving for $i \in I$ player i's type unchanged. Through a sequence of changes, we can arrive at a type vector for which players $i \in I$ unambiguously prefer ex post random dictatorship over the postulated alternative mechanism, and thus the postulated game form does not ex post Pareto dominate random dictatorship. The incentive compatibility arguments cannot be used when we consider interim expected utilities. If we change the type of player $j \notin I$, then the expectations of the types of players $i \in I$ remain unchanged.

Proof. Indirect. Suppose for some $p \in [0,1]^n$ such that $\sum_{i \in I} p_i = 1$ there were a game form G and a consistent Bayesian equilibrium of G for all type spaces σ^* that ex post Pareto dominated p-random dictatorship. For the outcome resulting from G and σ^* to be different from p-random dictatorship, there must be some $\widehat{\mathcal{T}} \in \Upsilon$, $\hat{t} \in \widehat{T}$, and $\hat{a} \in A$ such that:

$$\sum_{s \in S} x(s, \hat{a}) \cdot \prod_{i \in I} \sigma_i^*(\widehat{\mathcal{T}}, \hat{t}_i, s_i) < \sum_{\left\{i \in I : b(u_i(\hat{t}_i)) = \hat{a}\right\}} p_i. \tag{14}$$

That is, alternative $\hat{a}$ is chosen with a probability that is strictly smaller than the probability with which it is chosen under random dictatorship. Let $\widehat{I}$ be the set $\left\{i \in I : b(u_i(\hat{t}_i)) = \hat{a}\right\}$. Notice that we must have $\emptyset \neq \hat{I} \neq I$. If $\hat{I} = \emptyset$, the right hand side of (14) would be zero. If $\hat{I} = I$, then G and σ^* would be ex post Pareto worse than random dictatorship at $\hat{t}$.

To complete the proof we construct a new type space $\widetilde{\mathcal{T}}$, and infer from (14) that in this type space there is a type vector such that the types of all players in $\widehat{I}$ strictly prefer the outcome of p-random dictatorship conditional on this type vector to the outcome in G resulting from the equilibrium σ^* conditional on this type vector. Therefore, G and σ^* do not ex post Pareto-dominate p-random dictatorship.

The type sets in $\widetilde{\mathcal{T}}$ are given by: $\widetilde{T}_i = \widehat{T}_i$ for all $i \in \widehat{I}$, and $\widetilde{T}_i = \widehat{T}_i \cup \{\tilde{t}_i\}$ for all $i \notin \widehat{I}$. For all $i \in I$ the types in $\widehat{T}_i$ have the same utility functions and beliefs in $\widetilde{\mathcal{T}}$ as in $\widehat{\mathcal{T}}$. For all $i \notin \widehat{I}$ type $\tilde{t}_i$'s beliefs are given by:

$$\pi_i(\tilde{t}_i)\left[\left(\left(\hat{t}_j\right)_{j\in\widehat{I}}, \left(\tilde{t}_j\right)_{\substack{j\notin\widehat{I}\\ j<i}}, \left(\hat{t}_j\right)_{\substack{j\notin\widehat{I}\\ j>i}}\right)\right] = 1, \tag{15}$$

and type $\tilde{t}_i$'s utility function is:

$$\tilde{u}_i(\tilde{t}_i, a) = \begin{cases} 1 & \text{if } a = \tilde{a}; \\ 1-\varepsilon_a & \text{if } a \notin \{\hat{a}, \tilde{a}\}; \\ 0 & \text{if } a = \hat{a}; \end{cases} \tag{16}$$

where $\tilde{a}$ denotes the second most preferred alternative of some player k's type $\hat{t}_k$, where $k \in \widehat{I}$. We assume that $0 < \varepsilon_a < \bar{\varepsilon}$ for all $a \notin \{\hat{a}, \tilde{a}\}$ for some $\bar{\varepsilon} \in (0,1)$, and that $a, a' \notin \{\hat{a}, \tilde{a}\}$ and $a \neq a'$ implies $\varepsilon_a \neq \varepsilon_{a'}$. This assumption ensures that the utility functions satisfy the condition of no indifferences. Moreover, by letting $\bar{\varepsilon}$ tend to zero, we can ensure that all ε_a tend to zero, which is the case that we shall focus on.

We now show that for $\bar{\varepsilon}$ sufficiently small at type vector $((\hat{t}_i)_{i\in\hat{I}}, (\tilde{t}_i)_{i\notin\widehat{I}})$ the alternatives other than $\hat{a}$ are in equilibrium σ^* chosen with a probability larger than $1 - \sum_{i\in\widehat{I}} p_i$. Note that the proof of Proposition 3 is concluded once this assertion is established. This is because random dictatorship gives for every $k \in \hat{I}$ player k's type $\hat{t}_k$ his top alternative $\hat{a}$ with probability $\sum_{i\in\widehat{I}} p_i$, and type $\hat{t}_k$'s second most preferred alternative $\tilde{a}$ with probability $1 - \sum_{i\in\widehat{I}} p_i$. By contrast, G and σ^* yield $\hat{a}$ with probability less than $\sum_{i\in\widehat{I}} p_i$, and some other alternative, not necessarily type $\hat{t}_k$'s second most preferred alternative, with a probability larger than $1 - \sum_{i\in\widehat{I}} p_i$. Therefore, type $\hat{t}_k$ strictly prefers random dictatorship.

Consider the player $i \notin \widehat{I}$ for whom i is smallest. We denote this player by $i1$. This player, when type $\tilde{t}_{i1}$, expects with probability 1 that the other players' type vector is $\hat{t}_{-i1}$. Because σ^* is consistent, type $\tilde{t}_{i1}$ expects the types $\hat{t}_{-i1}$ to choose the same in $\widetilde{T}$ as in $\widehat{T}$. By the assumption of the indirect proof, type $\hat{t}_{i1}$ has a strategy available that yields alternatives other than $\hat{a}$ with probability of more than $1 - \sum_{i \in \widehat{I}} p_i$. Type $\tilde{t}_{i1}$ will not necessarily choose the same strategy as type $\hat{t}_{i1}$. But, for small enough $\bar{\varepsilon}$, only a strategy that yields an alternative other than $\hat{a}$ with some probability $\tilde{p} > 1 - \sum_{i \in \widehat{I}} p_i$ can be optimal. Choosing such a strategy yields for type $\tilde{t}_{i1}$ expected payoff greater than $\tilde{p}(1 - \bar{\varepsilon})$ whereas any other pure strategy yields a a payoff that is no more than $1 - \sum_{i \in \widehat{I}} p_i < \tilde{p}$. For small enough $\bar{\varepsilon}$ the former expected payoff is larger than the latter.

Now consider the player $i \notin \widehat{I}$ for whom i is second smallest. We denote this player by $i2$. This player, when type $\tilde{t}_{i2}$, expects with probability 1 the other players' types to be $\hat{t}_{-i2}$ except for player $i1$ whom $i2$ expects with probability 1 to be type $\tilde{t}_{i1}$. By the step of the previous paragraph, if $\tilde{t}_{i2}$ chose the same strategy as $\hat{t}_{i2}$ does in equilibrium, $\tilde{t}_{i2}$ would expect an outcome other than $\hat{a}$ with probability larger than $1 - \sum_{i \in \widehat{I}} p_i$. He might choose in equilibrium some other strategy, but, for small enough $\bar{\varepsilon}$, he will never make a choice that yields an outcome other than $\hat{a}$ with a probability that is not larger than $1 - \sum_{i \in \widehat{I}} p_i$.

The step of the previous paragraph can be iterated until we arrive at the player $i \notin \widehat{I}$ for whom i is largest. We denote this player by $i(n-1)$. This player expects the other players to be of type $\tilde{t}_{-(i(n-1))}$ except for types $i \in \widehat{I}$, whom this player expects to be of type $\hat{t}_i$. By the same argument used in the previous two paragraphs, type $\tilde{t}_{i(n-1)}$ chooses in equilibrium a strategy that he expects to yield an outcome other than $\hat{a}$ with probability larger than $1 - \sum_{i \in \widehat{I}} p_i$. But at type vector $((\hat{t}_i)_{i \in \hat{I}}, (\tilde{t}_i)_{i \notin \widehat{I}})$ this type has correct expectations, and therefore at this type vector the equilibrium strategies do indeed yield an outcome other than $\hat{a}$ with probability larger than $1 - \sum_{i \in \widehat{I}} p_i$. As explained above, this concludes the proof. □

8. Conclusion

Gibbard and Satterthwaite's impossibility theorem, and Hylland's version of this theorem in a setting with stochastic outcomes, are central results of voting theory. We have argued that the insistence of these theorems on belief independent strategy choices may be overly restrictive if a mechanism designer is concerned with Pareto improvements. Such a mechanism designer can find voting schemes that are superior to random dictatorship if agents' choices are allowed to depend on their beliefs. Whatever those beliefs are, the outcomes will be at least as good as under random dictatorship, and sometimes better. Such an improvement is only possible if agents' subjective beliefs are accepted, and an interim perspective is adopted. From an ex post perspective, such unambiguous improvements are not possible.

An important problem left open by our paper is the characterization of voting rules that are not dominated in one of the senses considered in this paper. One can take a mechanism design or an implementation approach to this question, depending on whether one considers just one, or all consistent Bayesian equilibria on all type spaces of a given game form. In Smith [23] the analogous question is investigated for public good mechanisms, using a mechanism design approach. Smith proves for one particular mechanism that it is not dominated. Smith's work shows the subtleties involved in such proofs. We leave the question as applied to voting rules for future research.

References

[1] Yaron Azrieli and Semin Kim (2011), Pareto Efficiency and Weighted Majority Rules, unpublished, Ohio State University.

[2] Salvador Barberà (2010), Strategy-proof Social Choice, Chapter 25 in: K. J. Arrow, A. K. Sen and K. Suzumura (eds.), *Handbook of Social Choice and Welfare*, Amsterdam: North-Holland.

[3] Dirk Bergemann and Stephen Morris (2003), Robust Mechanism Design, Cowles Foundation Discussion Paper No. 1421.

[4] Dirk Bergemann and Stephen Morris (2005), Robust Mechanism Design, *Econometrica* 73, 1771-1813.

[5] Dirk Bergemann and Stephen Morris (2011), Robust Mechanism Design: An Introduction, mimeo., Yale and Princeton, New Haven and Princeton.

[6] Dirk Bergemann and Stephen Morris (2011), Robust Implementation in General Mechanisms, *Games and Economic Behavior* 71, 261-281.

[7] Jean-Marie Blin and Mark Satterthwaite (1977), On Preferences, Beliefs, and Manipulation within Voting Situations, *Econometrica* 45, 881-888.

[8] Tilman Börgers (1991), Undominated Strategies and Coordination in Normalform Games, *Social Choice and Welfare* 8, 65-78.

[9] Tilman Börgers and Peter Postl (2009), Efficient Compromising, *Journal of Economic Theory* 155, 2057-2076.

[10] Steven Brams and Peter Fishburn (2007), *Approval Voting*, second edition, Heidelberg etc.: Springer.

[11] Kim-Sau Chung and Jeff Ely (2007), Foundations of Dominant Strategy Mechanisms, *Review of Economic Studies* 74, 447-476.

[12] Bhaskar Dutta, Hans Peters, and Arunava Sen (2007), Strategy-Proof Cardinal Decision Schemes, *Social Choice and Welfare* 28, 163-179.

[13] Bhaskar Dutta, Hans Peters, and Arunava Sen (2008), Strategy-Proof Cardinal Decision Schemes (Erratum), *Social Choice and Welfare* 30, 701-702.

[14] Jeffrey Ely and Marcin Peski (2006), Hierarchies of Belief and Interim Rationalizability, *Theoretical Economics* 1, 19-65.

[15] Alan Gibbard (1973), Manipulation of Voting Schemes: A General Result, *Econometrica* 41, 587-602.

[16] Jobst Heitzig and Forrest Simmons (2010), Some Chance for Consensus: Voting Methods For Which Consensus is an Equilibrium, *Social Choice and Welfare*, forthcoming.

[17] Aanund Hylland (1980), Strategy Proofness of Voting Procedures with Lotteries as Outcomes and Infinite Sets of Strategies, mimeo., University of Oslo, Institute of Economics.

[18] Bengt Holmström and Roger Myerson (1983), Efficient and Durable Decision Rules with Incomplete Information, *Econometrica* 51, 1799-1819.

[19] Matthew Jackson (2001), A Crash Course in Implementation Theory, *Social Choice and Welfare* 18, 655-708.

[20] Shasikanta Nandeibam (2004), The Structure of Decision Schemes with von Neumann Morgenstern Preferences, discussion paper, University of Bath.

[21] Mark Satterthwaite (1975), Strategy-Proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions, *Journal of Economic Theory* 1975, 187-217.

[22] Patrick W. Schmitz and Thomas Tröger (2011), The (Sub-) Optimality of the Majority Rule, *Games and Economic Behavior*, forthcoming.

[23] Doug Smith (2010), A Prior Free Efficiency Comparison of Mechanisms for the Public Goods Problem, mimeo., University of Michigan, Ann Arbor.

[24] Takuro Yamashita (2011), Robust Welfare Guarantees in Bilateral Trading Mechanisms, mimeo., Stanford University, Palo Alto.