ROBUST MECHANISM DESIGN AND DOMINANT STRATEGY VOTING RULES

TILMAN BÖRGERS AND DOUG SMITH

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

1. Introduction

Economic outcomes depend not only on market processes but also on political processes. Economists have therefore a long-standing interest in political decision making. Political decisions are often made through voting procedures. It is interesting to investigate which voting procedures perform *well* in the sense of helping to achieve some measure of economic welfare. One methodology that can be used to address this question is the theory of *mechanism design.* In this paper we consider the design of voting rules from the perspective of the theory of mechanism design.

Our starting point is a classic result on voting rules, due to Alan Gibbard [11] and Mark Satterthwaite [15]. According to this result the only dominant strategy voting rules for three or more alternatives are dictatorial voting rules. Gibbard and Satterthwaite assumed 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 (the *full domain assumption*). Gibbard and Satterthwaite then asked whether it is possible to construct a game form[1] that determines which alternative is chosen as a function of the strategies chosen by the voters, such that each voter has a dominant strategy whatever this voter's preferences are. A dominant strategy was defined to be a strategy that is always 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]

The interest in dominant strategy game forms is motivated by the fact a dominant strategy is a prediction of a rational voter's behavior that does

E-mail address: tborgers@umich.edu, dougecon@umich.edu.

Date: June 4, 2011.

[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 [1].

not require any assumption about the voter's beliefs about the other voters' strategy choices. If a voter does not have a dominant strategy, then that voter's optimal choice depends on her beliefs about other voters' behavior. These beliefs in turn may be derived from beliefs about other voters' preferences. It seems attractive to bypass such beliefs, and to construct a game form in 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 without making specific assumptions about these voters' beliefs about each other.*[3]

These two parts are logically independent. Part (A) seems more convincing: often voting schemes are constructed long before the precise context in which they will be used is known. It seems wise not to make any special assumptions about agents' knowledge about each other. Part (B) can perhaps be motivated by the idea that game forms in which voters' behavior can be uniquely predicted independent of their beliefs are simpler than game forms in which each voter's optimal choice depends on the voter's beliefs about other voters, but this point seems less compelling. The implicit idea of simplicity is just one of several conceivable notions of simplicity.

In this paper we present an investigation of the theory of voting rules that is based on the first part of the two part argument described above, but not on the second part. 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. Using the terminology of game theory, the fact that we do not make any assumptions about voters' beliefs about each other is reflected by the fact that we analyze 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.[4] However, we do not require each voter's choice, for given preference of that voter, to be the same for all type spaces.

Our main finding is that a mechanism designer who evaluates voting rules using the Pareto criterion, or a utilitarian welfare function, can improve on dictatorial mechanisms, even when not making any assumption abut voters' beliefs about each other. The fact that this result is true even if the mechanism designer only relies on the Pareto criterion appears paradoxical.

---

[3]Blin and Satterthwaite [2] 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.

[4]For the definitions of *type space* and *Bayesian equilibrium* see Fudenberg and Tirole [10, pp. 213-215].

How can one achieve a Pareto improvement on dictatorship? To explain, we need to describe the set-up of our paper in more detail.

In order to be able to use the notion of *Bayesian equilibrium* we use a framework that is slightly different from the framework that Gibbard and Satterthwaite used. We model voters' attitudes towards risk, adopting the assumption that voters evaluate risky prospects according to von Neumann Morgenstern utility theory. It then seems natural to allow voting rules to map profiles of von Neumann Morgenstern utility functions into probability distributions over outcomes. The first question that arises is whether a version of Gibbard and Satterthwaite's theorem holds for the setting just described. This question has been answered affirmatively by Aanund Hylland in 1980 in the unpublished [12]. When voters have von Neumann Morgenstern utilities, and lotteries are allowed as outcomes, then the only game forms that offer each agent always a dominant strategy, and that pick an alternative if it is unanimously preferred by all agents, are random dictatorships.[5] In random dictatorships each voter 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.

We can now state the two main results of this paper. Both results address whether there are game forms such that for all finite type spaces, there is at least one Bayesian 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.[6] 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. Indeed, there is no game form that increases the sum of players' expected utilities. This is our second main result. Our first main result thus indicates that a robust analysis of voting schemes can lead to more positive results if the requirement that voters' optimal strategies are independent of their beliefs is abandoned. Our second main result shows that such positive results are only available for some specific welfare criteria and not for others.

Our approach is related to Bergemann and Morris' [4] work on robust mechanism design. As we do, they consider Bayesian equilibria of mechanisms on *all* type spaces. Bergemann and Morris seek conditions under which the Bayesian implementability of a social choice correspondence on

[5]This result is Theorem 1* in Hylland [12]. It is also Theorem 1 in Dutta et. al. [8] (see also [9]) where an alternative proof is provided. Another proof is in Nandeibam [14].

[6]The notions of interim and ex post efficiency are due to Holmström and Myerson [13].

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 environments without transferrable payoffs considered by Gibbard and Satterthwaite, dominant strategy implementability may be a stronger requirement than Bayesian implementability on all type spaces.[7] Bergemann and Morris do not consider the problem of comparing different mechanisms from an efficiency or welfare point of view. Such comparisons are a focus in our work.

The approach of this paper are also closely related to Smith [16] who is concerned with the problem of designing a mechanism for public goods. Like we do in this paper, Smith considers the performance of different mechanisms 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 paper, by contrast, when considering the ex post perspective, we find that no mechanism can improve on dominant strategy mechanisms.

The spirit of our work in this paper is also related to Börgers [5] who showed in the Gibbard-Satterthwaite framework the existence of mechanisms for which the outcomes that result if all players chose a strategy from their sets of *undominated strategies* are Pareto efficient, and in a sense defined in that paper less biased than the outcomes of dictatorship. The set of undominated strategies is equal to the set of expected utility maximizing strategies that a rational agent might choose if one considers all possible beliefs. Thus, implicitly, [5] considered implementation on all type spaces with belief-dependent strategies, and contrasted this with Gibbard and Satterthwaite's dominant strategy requirement. However, Börgers used a framework in which agents' preferences were modeled using ordinal preferences rather than von Neumann Morgenstern utilities. Moreover, his approach can be considered an *implementation* approach, as he considered *all* undominated strategies, whereas our approach here is a *mechanism design* approach in the sense that we study for every type space *some* equilibrium, but not *all* Bayesian equilibria. We leave the further exploration of the implementation approach in our framework to future research.

Below, in Section 2, we explain the model and the definitions used in this paper. In Section 3 we explain the welfare criteria that we use to evaluate different game forms. In Section 4 we adapt Hylland's theorem on random dictatorship to our setting. In Section 5 we present our two main results.

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

Section 6 concludes. The proof of the second of the main results is in an Appendix.

2. The Voting Problem

There are two agents: $i \in \{1, 2\}$.[8] The agents have to choose one alternative from a finite set A of alternatives. We assume that A has at least three elements. The set of all probability distributions over A is $\Delta(A)$, where for $\delta \in \Delta(A)$ we denote by $\delta(a) \in [0, 1]$ the probability that δ assigns to alternative a. The two 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 each agent's von Neumann Morgenstern utility function is normalized such that $\min_{a \in A} u_i(a) = 0$ and $\max_{a \in A} u_i(a) = 1$. We also 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 $u_i(\delta) = \sum_{a \in A} u_i(a) \cdot \delta(a)$.

A mechanism designer has a 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 $G = (S_1, S_2, x)$ *consists of:*

(i) *a non-empty finite strategy set S_i for each agent $i \in \{1, 2\}$;*

We define: $S \equiv S_1 \times S_2$.

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

To make the exposition in this paper easier, we require in Definition 1 that the strategy sets are finite. However, our results will also hold when game forms are allowed to have infinite strategy sets.

Once the mechanism designer has announced a game form, the two agents choose simultaneously and independently their strategies. 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 specifying a *type space.*

Definition 2. *A* type space $\mathcal{T} = (T_1, T_2, \pi_1, \pi_2, u_1, u_2)$ *consists for each $i \in \{1, 2\}$ of:*

(i) *a nonempty, finite set T_i of types;*

We write $\Delta(T_i)$ for the set of all probability distributions over T_i.

[8] We restrict attention to only two agents for simplicity. We conjecture, but have not yet proven, that all our arguments extend to the case of more than two agents.

(ii) *a belief function* $\pi_i : T_i \to \Delta(T_j)$ *(where* $j \neq i$*);*

(iii) *a utility function* $u_i : T_i \times A \to [0,1]$.

We write $\pi_i(t_i)[t_j]$ *for the probability that type* i *assigns to player* j *being type* t_j *(where* $j \neq i$*). We write* $u_i(t_i)[a]$ *for the utility that* $u_i(t_i)$ *assigns to* a.[9] *The utility function satisfies for both* $i \in \{1,2\}$ *and all* $t_i \in T_i$ *the assumptions introduced earlier:*

(a) $\min_{a \in A} u_i(t_i)[a] = 0$ *and* $\max_{a \in A} u_i(t_i)[a] = 1$*;*

(b) $u_i(t_i)[a] \neq u_i(t_i)[b]$ *whenever* $a \neq b$.

In this definitions beliefs are purely 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 probability one to an event to which another agent attaches probability zero. Observe also that we assume type spaces to be finite. We thus avoid technical difficulties associated with infinite type spaces.

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 in which agents find themselves. We denote the set of all type spaces by Υ.

The mechanism designer proposes to agents how they might play the game. He might propose to agents to randomize. For $i = 1, 2$ we denote by $\Delta(S_i)$ the set of all probability distributions on S_i. For the agents to accept the mechanism designer's proposal, he must propose a *Bayesian equilibrium.* Because the mechanism designer does not know the true type space, he has to propose a *Bayesian equilibrium for every type space.*

Definition 3. *A* Bayesian equilibrium of game form G for every type space *is a pair* (σ_1, σ_2) *such that for every* $i \in \{1,2\}$*:*

(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* $\Delta(S_i)$.

We write $\sigma_i(\mathcal{T}, t_i)$ *for the mixed strategy assigned to* $t_i \in T_i$*, and* $\sigma_i(\mathcal{T}, t_i)[s_i]$ *for the probability that this mixed strategy assigns to* $s_i \in S_i$.

(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:*

$$\sum_{t_j \in T_j} \sum_{s_1 \in S_1, s_2 \in S_2} u_i(x(s_1, s_2)) \cdot \sigma_i[s_i] \cdot \sigma_j(\mathcal{T}, t_j)[s_j] \cdot \pi(t_i)[t_j], \tag{1}$$

where $j \neq i$.

[9] Observe that we suppress in the notation the dependence of π_i and u_i on the type space $\mathcal{T}$. We are not aware of any confusion that might arise from this simplification of our notation.

A restrictive requirement for the Bayesian equilibria that the mechanism designer proposes, and that is implicit in the work on dominant strategy mechanism design, is that equilibria be *belief independent*.

Definition 4. *A game form G and a Bayesian equilibrium of G for every type space, (σ_1, σ_2), is* belief independent *if for all $i \in \{1,2\}, \mathcal{T}, \widetilde{\mathcal{T}} \in \Upsilon, t_i \in T_i$ and $\tilde{t}_i \in \tilde{T}_i$ such that $u_i(t_i) = \tilde{u}_i(\tilde{t}_i)$ we have:*

$$\sigma_i(\mathcal{T}, t_i) = \sigma_i(\widetilde{\mathcal{T}}, \tilde{t}_i), \tag{2}$$

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

Our main interest in this paper is in relaxing the requirement of belief independence. We shall, however, not be able to 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 5. *A Bayesian equilibrium of game form G for every type space, (σ_1, σ_2), is* consistent *if for all type spaces $\mathcal{T}, \widetilde{\mathcal{T}} \in \Upsilon$ such that:*

(i) *for every $i \in \{1,2\}$: $\widetilde{T}_i \subseteq T_i$ (where $\widetilde{T}_i$ corresponds to $\widetilde{\mathcal{T}}$ and T_i corresponds to $\mathcal{T}$);*

(ii) *for every $i \in \{1,2\}$ and every $t_i \in T_i$: $\tilde{u}_i(t_i) = u_i(t_i)$ and $\tilde{\pi}(t_i) = \pi(t_i)$ (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 $i \in \{1,2\}$ and every $t_i \in T_i$:

(iii) $\sigma(\widetilde{\mathcal{T}}, t_i) = \sigma(\mathcal{T}, t_i)$.

Observe that the type t_i referred to in item (iii) of Definition 5 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 an agent's equilibrium choices should only depend on that agent's utility function and that agent's hierarchy of beliefs. This assumption seems reasonable because the type space, as opposed to the utility function and the hierarchy of beliefs, is really only a construction by the modeler, and not necessarily a construction that the agent is aware of. We don't explicitly formulate the stronger assumption that equilibrium choices should only depend on agents' utility functions and hierarchies of beliefs, but instead work with the weaker consistency requirement, because the consistency requirement is easier to formulate, and is sufficient for our purposes. We believe that our results would also go through if we made the more demanding assumption for equilibria.

3. Welfare

We postulate a mechanism designer who seeks to further the utility of the agents rather than his own utility. At different points we investigate the implications of different objectives for the mechanism designer. At times, we

shall only assume that the mechanism designer seeks to achieve a Pareto efficient decision. At other points in the paper, we consider the implications of the assumption that the mechanism designer seeks to maximize the welfare function: $u_1(a) + u_2(a)$. Because we have normalized utilities, this corresponds to the "relative utilitarian" welfare function that was axiomatized by Dhillon [6] and Dhillon and Mertens [7].

When evaluating the utility of the two agents for a realized type combination (t_1, t_2) the mechanism designer can either only consider the outcomes that result from the mixed strategies prescribed for these two types, or she may consider the expected utilities of these two 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. From this perspective, 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 are clearly preferable. We pursue both perspectives in this paper.

The considerations of the preceding two paragraphs lead to four possible formalizations of the mechanism designer's objectives. We present these in the four definitions that follow below. None of these definitions attributes a prior over type spaces in Υ or over types in each type space to the mechanism designer. Instead, we work with a dominance notion, that is prior free. Whatever the mechanism designer's prior is, if he has one, he will never choose a dominated game form in the sense described in the four definitions below.[10]

Definition 6. *The game form G with the consistent Bayesian equilibrium for all type spaces (σ_1, σ_2)* ex post Pareto dominates *the game form $\tilde{G}$ with the consistent Bayesian equilibrium for all type spaces $(\tilde{\sigma}_1, \tilde{\sigma}_2)$ if for all $i \in \{1, 2\}$, $\mathcal{T} \in \Upsilon$, and $(t_1, t_2) \in T_1 \times T_2$:*

$$\sum_{s_1 \in S_1, s_2 \in S_2} u_i(t_i)[x(s_1, s_2)] \cdot \sigma_1(\mathcal{T}, t_1)[s_1] \cdot \sigma_2(\mathcal{T}, t_2)[s_2] \geq$$

$$\sum_{s_1 \in \tilde{S}_1, s_2 \in \tilde{S}_2} u_i(t_i)[\tilde{x}(s_1, s_2)] \cdot \tilde{\sigma}_1(\mathcal{T}, t_1)[s_1] \cdot \tilde{\sigma}_2(\mathcal{T}, t_2)[s_2], \tag{3}$$

with strict inequality for at least one $i \in \{1, 2\}$, $\mathcal{T} \in \Upsilon$, and $(t_1, t_2) \in T_1 \times T_2$. A direct mechanism that is not ex post Pareto dominated will be called ex post Pareto undominated.

[10] The two main results of this paper use Definitions 7 and 8. We provide the other two definitions, and discuss the relations among the concepts introduced in these four definitions, to give the reader a better understanding of the context of our main results.

Definition 7. *The game form G with the consistent Bayesian equilibrium for all type spaces (σ_1, σ_2)* ex post utilitarian[11] dominates *the game form $\tilde{G}$ with the consistent Bayesian equilibrium for all type spaces $(\tilde{\sigma}_1, \tilde{\sigma}_2)$ if for all $\mathcal{T} \in \Upsilon$, and $(t_1, t_2) \in T_1 \times T_2$:*

$$\sum_{i\in\{1,2\}} \sum_{s_1\in S_1, s_2\in S_2} u_i(t_i)[x(s_1,s_2)] \cdot \sigma_1(\mathcal{T},t_1)[s_1] \cdot \sigma_2(\mathcal{T},t_2)[s_2] \geq$$

$$\sum_{i\in\{1,2\}} \sum_{s_1\in \tilde{S}_1, s_2\in \tilde{S}_2} u_i(t_i)[\tilde{x}(s_1,s_2)] \cdot \tilde{\sigma}_1(\mathcal{T},t_1)[s_1] \cdot \tilde{\sigma}_2(\mathcal{T},t_2)[s_2], \tag{4}$$

with strict inequality for at least one $\mathcal{T} \in \Upsilon$, and $(t_1, t_2) \in T_1 \times T_2$. A direct mechanism that is not ex post utilitarian dominated will be called ex post utilitarian undominated.

Note that the game form G with the consistent Bayesian equilibrium for all type spaces (σ_1, σ_2) ex post utilitarian dominates game form $\tilde{G}$ with the consistent Bayesian equilibrium for all type spaces $(\tilde{\sigma}_1, \tilde{\sigma}_2)$ if the former ex post Pareto dominates the latter.

Definition 8. *The game form G with the consistent Bayesian equilibrium for all type spaces (σ_1, σ_2)* interim Pareto dominates *the game form $\tilde{G}$ with the consistent Bayesian equilibrium for all type spaces $(\tilde{\sigma}_1, \tilde{\sigma}_2)$ if for all $i, j \in \{1, 2\}$ with $i \neq j$, $\mathcal{T} \in \Upsilon$, and $t_i \in T_i$:*

$$\sum_{t_j\in T_j} \pi_i(t_i)[t_j] \sum_{s_1\in S_1, s_2\in S_2} u_i(t_i)[x(s_1,s_2)] \cdot \sigma_1(\mathcal{T},t_1)[s_1] \cdot \sigma_2(\mathcal{T},t_2[s_2] \geq$$

$$\sum_{t_j\in T_j} \pi_i(t_i)[t_j] \sum_{s_1\in \tilde{S}_1, s_2\in \tilde{S}_2} u_i(t_i)[\tilde{x}(s_1,s_2)] \cdot \tilde{\sigma}_1(\mathcal{T},t_1)[s_1] \cdot \tilde{\sigma}_2(\mathcal{T},t_2)[s_2], \tag{5}$$

with strict inequality for at least one $i, j \in \{1, 2\}$ with $i \neq j$, $\mathcal{T} \in \Upsilon$, and $t_i \in T_i$. A direct mechanism that is not interim Pareto dominated will be called interim Pareto undominated.

Definition 9. *The game form G with the consistent Bayesian equilibrium for all type spaces (σ_1, σ_2)* interim utilitarian dominates *the game form $\tilde{G}$ with the consistent Bayesian equilibrium for all type spaces $(\tilde{\sigma}_1, \tilde{\sigma}_2)$ if for all $\mathcal{T} \in \Upsilon$ and $(t_1, t_2) \in T_1 \times T_2$:*

$$\sum_{i\in\{1,2\}} \sum_{t_j\in T_j} \pi_i(t_i)[t_j] \sum_{s_1\in S_1, s_2\in S_2} u_i(t_i)[x(s_1,s_2)] \cdot \sigma_1(\mathcal{T},t_1)[s_1] \cdot \sigma_2(\mathcal{T},t_2)[s_2] \geq$$

$$\sum_{i\in\{1,2\}} \sum_{t_j\in T_j} \pi_i(t_i)[t_j] \sum_{s_1\in \tilde{S}_1, s_2\in \tilde{S}_2} u_i(t_i)[\tilde{x}(s_1,s_2)] \cdot \tilde{\sigma}_1(\mathcal{T},t_1)[s_1] \cdot \tilde{\sigma}_2(\mathcal{T},t_2)[s_2], \tag{6}$$

with strict inequality for at least one $\mathcal{T} \in \Upsilon$ and $(t_1, t_2) \in T_1 \times T_2$. A direct mechanism that is not interim utilitarian dominated will be called interim utilitarian undominated.

Note that the game form G with the consistent Bayesian equilibrium for all type spaces (σ_1, σ_2) interim utilitarian dominates game form $\tilde{G}$ with the

[11]For simplicity, we use "utilitarian" rather than the more clumsy "relative utilitarian."

consistent Bayesian equilibrium for all type spaces $(\tilde{\sigma}_1, \tilde{\sigma}_2)$ if the former interim Pareto dominates the latter.

4. Belief Independent Equilibria: Hylland's Theorem

We begin by restating Hylland's version of the Gibbard Satterthwaite theorem in our setting. Hylland's theorem implies that all game forms and belief independent equilibria of these game forms that satisfy a unanimity requirement are random dictatorships. To define unanimity and random dictatorships we need some notation. If u is a utility function, we denote by $d(u)$ the element of A that maximizes u.[12]

Definition 10. *A game form G and a Bayesian equilibrium of G for every type space, (σ_1, σ_2), satisfy* unanimity *if for every $\mathcal{T} \in \Upsilon$, $(t_1, t_2) \in T_1 \times T_2$ and every $a \in A$:*

$$\sum_{s_1 \in S_1, s_2 \in S_2} \sigma_1(\mathcal{T}, t_1)[s_1] \cdot \sigma_2(\mathcal{T}, t_2)[s_2] \cdot x(s_1, s_2)[a] = 1 \tag{7}$$

whenever $d(u_1(t_1)) = d(u_2(t_2)) = a$.

Definition 11. *A game form G and a Bayesian equilibrium of G for every type space, (σ_1, σ_2), are a* random dictatorship *if there is some $p \in [0,1]$ such that for every $\mathcal{T} \in \Upsilon$, $(t_1, t_2) \in T_1 \times T_2$ and every $a \in A$:*

$$\sum_{s_1 \in S_1, s_2 \in S_2} x(s_1, s_2)[a] \cdot \sigma_1(\mathcal{T}, t_1)[s_1] \cdot \sigma_2(\mathcal{T}, t_2)[s_2] =$$

$$= \begin{cases} 1 & \text{if } d(u_1(t_1)) = a \text{ and } d(u_2(t_2)) = a, \\ p & \text{if } d(u_1(t_1)) = a \text{ and } d(u_2(t_2)) \neq a, \\ 1-p & \text{if } d(u_1(t_1)) \neq a \text{ and } d(u_2(t_2)) = a, \\ 0 & \text{if } d(u_1(t_1)) \neq a \text{ and } d(u_2(t_2)) \neq a. \end{cases} \tag{8}$$

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

Proposition 1. *A game form G and a Bayesian equilibrium of G for every type space, (σ_1, σ_2), are belief-independent and satisfy 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_1, σ_2) a "cardinal decision scheme" in the sense of Definition 1 in [8], and show that this cardinal decision scheme has the properties listed in Theorem 1 in [8] which is a version of Hylland's theorem. It then follows from Theorem 1 in [8] 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 properties that were introduced in

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

[13]Theorem 1* in Hylland [12]. Hylland's theorem does not assume that game forms are finite.

Definition 2. A cardinal decision scheme is a mapping $\phi : \mathcal{U}^2 \to \Delta(A)$. We can derive from G and (σ_1, σ_2) a cardinal decision scheme by setting for any $(u_1, u_2) \in \mathcal{U}^2$:

$$\phi(u_1, u_2) = \sum_{s_1 \in S_1, s_2 \in S_2} \sigma_1(\mathcal{T}, t_1)[s_1] \cdot \sigma_2(\mathcal{T}, t_2)[s_2] \cdot x(s_1, s_2), \tag{9}$$

where we can pick any $\mathcal{T} \in \Upsilon$ and any $(t_1, t_2) \in T_1 \times T_2$ such that $u_1(t_1) = u_1$ and $u_2(t_2) = u_2$. By belief-independence it does not matter which such $\mathcal{T}$ and $(t_1, t_2) \in T_1 \times T_2$ we choose. Then ϕ is a cardinal decision scheme as defined in Definition 1 of [8]. We can complete the proof by showing that ϕ has the two properties listed in Theorem 1 of [8]. The first property is unanimity: If $d(u_1) = d(u_2) = a$ then $\phi(u_1, u_2) = a$. This is implied by the assumption that G and (σ_1, σ_2) satisfy unanimity. The second is strategy proofness: If $(u_1, u_2) \in \mathcal{U}^2$ and $u_1' \in \mathcal{U}$, then $u_1(\phi(u_1, u_2)) \geq u_1(\phi(u_1', u_2))$ and the same condition also holds for agent 2. To prove this for agent 1 we pick $\mathcal{T} \in \Upsilon$, $t_1, t_1' \in T_1$ and $t_2 \in T_2$ such that $u_1(t_1) = u_1, u_1(t_1') = u_1'$, and $u_2(t_2) = u_2$. Moreover, $\pi_1(t_1)$ and $\pi_1(t_1')$ place probability 1 on t_2. Then the fact that σ_1 and σ_2 are Bayesian equilibria of G for the type space $\mathcal{T}$ implies:

$$\sum_{s_1 \in S_1, s_2 \in S_2} (u_1(x(s_1, s_2)) \cdot \sigma_1(\mathcal{T}, t_1)[s_1] \cdot \sigma_2(\mathcal{T}, t_2)[s_2]) \geq$$

$$\sum_{s_1 \in S_1, s_2 \in S_2} \big(u_1(x(s_1, s_2)) \cdot \sigma_1(\mathcal{T}, t_1')[s_1] \cdot \sigma_2(\mathcal{T}, t_2)[s_2]\big) \tag{10}$$

By the definition of ϕ, this is equivalent to: $u_1(\phi(u_1, u_2) \geq u_1(\phi(u_1', u_2))$, that is, strategy proofness. The proof of strategy proofness for agent 2 is analogous. □

From now on, when we refer to random dictatorship, we shall mean a specific game form G, and a specific equilibrium (σ_1, σ_2) of G for every type space.

Definition 12. *The following game form G and equilibrium (σ_1, σ_2) of G for every type space will be referred to as* p-random dictatorship*:*

(i) $S_1 = S_2 = A$;

(ii)

$$x(s_1, s_2)[a] = \begin{cases} 1 & \text{if } s_1 = s_2 = a; \\ p & \text{if } s_1 = a \text{ and } s_2 \neq a; \\ 1 - p & \text{if } s_1 \neq a \text{ and } s_2 = a; \\ 0 & \text{if } s_1 \neq a \text{ and } s_2 \neq a; \end{cases}$$

(iii) $\sigma_i(\mathcal{T}, t_i)[d(u_i(t_i))] = 1$ *for all* $i \in \{1, 2\}$, $\mathcal{T} \in \Upsilon$, *and* $t_i \in T_i$.

It is immediate that (σ_1, σ_2) is a 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 one described in Definition 12.

5. Game Forms that Dominate Random Dictatorship

We can now present the two main results of this paper. The first result examines interim Pareto dominance, while the second result concerns ex post utilitarian dominance. The first result says that for every $p \in [0,1]$ such that $p \neq 0$ and $p \neq 1$ there are a game form, and a Bayesian equilibrium of this game form for every type space, that interim Pareto dominate random dictatorship when the probability of agent 1 being dictator is p. We refer to the game form as *p-random dictatorship with compromise.*

Definition 13. *The following game form is called a* p-random dictatorship with compromise.

(i) *for every* $i \in \{1,2\}$*:*

$$S_i = 2^A \times A,$$

where 2^A *is the set of all non-empty subsets of* A*;*

(ii) *If* $s_1 = (\mathcal{A}_1, a_1)$*,* $s_2 = (\mathcal{A}_2, a_2)$*, and* $\mathcal{A}_1 \cap \mathcal{A}_2 = \emptyset$*, then:*

$$x(s_1,s_2)[a] = \begin{cases} 1 & \text{if } a_1 = a_2 = a; \\ p & \text{if } a_1 = a \text{ and } a_2 \neq a; \\ 1-p & \text{if } a_1 \neq a \text{ and } a_2 = a; \\ 0 & \text{if } a_1 \neq a \text{ and } a_2 \neq a; \end{cases}$$

(iii) *If* $s_1 = (\mathcal{A}_1, a_1)$*,* $s_2 = (\mathcal{A}_2, a_2)$*, and* $\mathcal{A}_1 \cap \mathcal{A}_2 \neq \emptyset$*, then there is some* $a \in \mathcal{A}_1 \cap \mathcal{A}_2$ *such that*

$$x(s_1,s_2)[a] = 1.$$

In words, this game form offers each agent i the opportunity to nominate one preferred alternative, a_i, and also a set $\mathcal{A}_i$ of "acceptable" alternatives. If there is exactly one alternative that both voters include in their set of acceptable alternatives, then that alternative is chosen with probability 1. If there is more than one alternative that both voters include in their set of alternatives, then one of those alternatives is chosen with probability 1. Otherwise, the mechanism reverts to random dictatorship. We refer to this game form as *random dictatorship with compromise* because it offers agents the opportunity to compromise on a mutually acceptable alternative in place of random dictatorship.

One Bayesian equilibrium of this game form is that both agents always choose a_i to be their most preferred alternative, and set $\mathcal{A}_i = \{a_i\}$. In this equilibrium, the possibility of a compromise is not used by either agent. This is an equilibrium because neither agent can unilaterally force a compromise. Any deviation that unilaterally alters the set of acceptable alternatives has no effect. However, the next proposition shows that *p-random dictatorship with compromise* also has a Bayesian equilibrium for all type spaces that

interim Pareto dominates random dictatorship. We also show that this equilibrium respects unanimity, to clarify that our result does indeed result from weakening the dominance requirement, and not weakening any other property listed in Proposition 1.

Proposition 2. *For all $p \in [0,1]$ such that $p \neq 0$ and $p \neq 1$, p-random dictatorship with compromise has a consistent equilibrium for all type spaces (σ_1, σ_2) that interim Pareto dominates p-random dictatorship and that respects unanimity.*

Interim Pareto dominance implies interim utilitarian dominance. Therefore, Proposition 2 also shows that p-random dictatorship with compromise has an equilibrium that interim utilitarian dominates p-random dictatorship. The main difficulty in the proof below is not so much showing interim Pareto dominance, but proving the existence of a consistent equilibrium. The argument in the proof below can be used to show the existence of consistent Bayesian equilibria for all type spaces of arbitrary finite games.

Proof. We construct the equilibrium (σ_1, σ_2). To ensure that the equilibrium satisfies unanimity we require each type's strategy to always include the alternative ranked top by that type in the set of acceptable alternatives. This restriction of the strategy space is innocuous, because any strategy that does not include the top ranked alternative in the set of acceptable alternatives is weakly dominated by a strategy that does include the top ranked alternative. Moreover, this restriction does indeed imply that if both agents rank the same alternative at the top, then that alternative is chosen with probability 1.

We now proceed inductively. We begin by considering type spaces $\mathcal{T}$ where for every $i \in \{1,2\}$ the set T_i has exactly one element. In such type spaces it is common belief 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 both i we have:

$$u_i(a) > pu_i(d(u_1(t_1))) + (1-p)u_i(d(u_2(t_2))). \tag{11}$$

Observe that the assumption $p \neq 0$ and $p \neq 1$ implies that some such type spaces exist. For such type spaces the strategies are:

$$\sigma_i(\mathcal{T}, t_i) = (\{d(u_i(t_i)), a\}, d(u_i(t_i))) \tag{12}$$

for $i \in \{1,2\}$. Note that these strategies constitute a Nash equilibrium of the complete information game in which agents' preferences are common knowledge, and that the outcome a strictly Pareto-dominates the outcome under random dictatorship. For all other type spaces with just a single element for each player the strategies are:

$$\sigma_i(\mathcal{T}, t_i) = (\{d(u_i(t_i))\}, d(u_i(t_i))) \tag{13}$$

Note that these strategies constitute a Nash equilibrium of the complete information game in which agents' preferences are common knowledge, and that the outcome is exactly the same as under random dictatorship.

Now suppose we had constructed the equilibrium for all type spaces $\mathcal{T}$ in which T_1 and T_2 have at most n elements. We first extend the construction to all type spaces $\mathcal{T}$ in which T_1 has at most $n+1$ elements and T_2 has at most n elements. Then we extend the construction to all type spaces $\mathcal{T}$ in which T_1 has at most $n+1$ elements and T_2 has at most $n+1$ elements.

Suppose first that we are considering a type space $\mathcal{T}$ in which T_1 has at most $n+1$ elements and T_2 has at most n elements. Consider all type spaces $\widetilde{\mathcal{T}}$ that are contained in $\mathcal{T}$, i.e. for which conditions (i) and (ii) of Definition 5 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 or for every $i \in \{1,2\}$ and every $t_i \in \widetilde{T}_i$:

$$\sigma_i\left(\mathcal{T}, t_i\right) = \sigma_i\left(\widetilde{\mathcal{T}}, t_i\right). \tag{14}$$

By the inductive hypothesis the right hand side of this equation has already been defined. Observe that this is well-defined. 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 5, 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. This strategic game has a Nash equilibrium in mixed strategies. 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 has at most $n+1$ elements and T_2 has at most $n+1$ elements in the same way as we extend it to all type spaces $\mathcal{T}$ in which T_1 has at most $n+1$ elements and T_2 has at most n elements.

To conclude the proof we note that this equilibrium 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 $\mathcal{A}_i = \{d(u_i(t_i))\}$. 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 (11) holds. □

Proposition 2 indicates that Pareto improvements on random dictatorship are possible if we focus on interim expected utilities of all types. However, interim expected utilities are determined by types' subjective beliefs, and these subjective beliefs may be wrong. For example, a player may expect that another player is of a certain type even though this player is of a different type. It is therefore interesting to consider instead ex post expected utility. We find in this case a negative result.

Proposition 3. *For all $p \in [0,1]$, there is no game form G that has a consistent equilibrium for all type spaces (σ_1, σ_2) that ex post utilitarian dominates p-random dictatorship.*

Because ex post utilitarian dominance implies ex post Pareto dominance, this result implies that there are no game form and equilibrium for all type spaces that ex post Pareto dominate p-random dictatorship. The proof of Proposition 3 is in the appendix. It is an indirect proof. We postulate the existence of a mechanism that ex post utilitarian dominates p-random dictatorship. We first show that such a mechanism cannot Pareto-dominate p-random dictatorship. Then we find a type pair such that one player is better off, and another player is better off under the postulated dominating mechanism than under p-random dictatorship. We then add types of the player who is better off, and use incentive compatibility arguments to show that we can make the increase in this player's utility arbitrarily small, while the utility loss of the other player remains bounded away from zero. This then contradicts utilitarian dominance. When we add types to the type space, but assume that for the existing types the strategy remains unchanged, we make use of the consistency assumption. We introduced this assumption to make precisely this argument in the proof of Proposition 3 possible.

6. Conclusion

Gibbard and Satterthwaite's theorem, and Hylland's version of this theorem in a cardinal utility setting, are central results of voting theory. We have argued that the insistence of the theorem on unique, belief independent strategy choices may be overly restrictive if a mechanism designer is considered who is primarily concerned either with Pareto improvements or with utilitarian welfare. 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 than are not dominated in one of the senses considered in this paper. In Smith [16] the analogous question is investigated for public goods mechanisms. Smith's work shows the subtleties of this problem. Another important step is the investigation of robust implementation as opposed to robust mechanism design. Implementation, unlike mechanism design, considers *all* equilibria of a given game form. One might ask whether there are mechanisms such that *all* equilibria on all type spaces dominate random dictatorship. We leave this question for future research.

References

[1] 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.

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

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

[6] Amrita Dhillon (1998), Extended Pareto Rules and Relative Utilitarianism, *Social Choice and Welfare* 15, 521-542.

[7] Amrita Dhillon and Jean Francois Mertens (1999), Relative Utilitarianism, *Econometrica* 67, 471-498.

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

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

[10] Drew Fudenberg and Jean Tirole (1991), *Game Theory*, Cambridge and London: The MIT Press.

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

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

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

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

[15] 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.

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

Appendix

Proof of Proposition 3. Step 1: We show for every game form G and every equilibrium of G for all type spaces, (σ_1, σ_2), if G and (σ_1, σ_2) ex post utilitarian dominate p-random dictatorship, then:

$$\sum_{s_1 \in S_1, s_2 \in S_2} x(s_1, s_2)[a] \cdot \sigma_1(\mathcal{T}, t_1)[s_1] \cdot \sigma_2(\mathcal{T}, t_2)[s_2] \leq 1 - p \tag{15}$$

for all $\mathcal{T} \in \Upsilon$, every $(t_1, t_2) \in T_1 \times T_2$, and every $a \in A$ such that $a \neq d(u_1(t_1))$, and

$$\sum_{s_1 \in S_1, s_2 \in S_2} x(s_1, s_2)[a] \cdot \sigma_1(\mathcal{T}, t_1)[s_1] \cdot \sigma_2(\mathcal{T}, t_2)[s_2] \leq p \tag{16}$$

for all $\mathcal{T} \in \Upsilon$, every $(t_1, t_2) \in T_1 \times T_2$, and every $a \in A$ such that $a \neq d(u_2(t_2))$. That is, any alternative that is not agent 1's preferred alternative can be chosen with a probability of at most $1-p$, and any alternative that is not agent 2's preferred alternative can be chosen with a probability of at most p. We prove this statement only for agent 1. The proof for agent 2 is analogous.

The proof is indirect. Suppose there were some type space $\mathcal{T}^*$, some $(t_1^*, t_2^*) \in T_1^* \times T_2^*$, and some alternative $a^* \in A$ such that $a^* \neq d(u_1(t_1^*))$, and yet:

$$\sum_{s_1 \in S_1, s_2 \in S_2} x(s_1, s_2)[a^*] \cdot \sigma_1(\mathcal{T}^*, t_1^*)[s_1] \cdot \sigma_2(\mathcal{T}^*, t_2^*)[s_2] > 1 - p. \tag{17}$$

We now construct a new type space, $\widehat{\mathcal{T}}$, and show that in this type space there is a vector of types such that the outcome prescribed by the equilibrium (σ_1, σ_2) yields lower ex post utilitarian welfare than p-random dictatorship. This contradicts the assumption that G and (σ_1, σ_2) ex post utilitarian dominate p-random dictatorship.

The type sets in $\widehat{\mathcal{T}}$ are given by: $\widehat{T}_1 = T_1^*$, and $\widehat{T}_2 = T_2^* \cup \{t_2(1), \ldots, t_2(K)\}$ where $K \in \mathbb{N}$ is large enough. We define later how large K needs to be. The types that are contained in T_1^* or T_2^* have the same utility function and beliefs in $\widehat{\mathcal{T}}$ as in $\mathcal{T}$. For types $t_2 \in \{t_2(1), t_2(2), \ldots t_2(K)\}$ the beliefs are given by:

$$\pi_2(t_2(k))[t_1^*] = 1. \tag{18}$$

The utility function of types $t_2 \in \{t_2(1), t_2(2), \ldots t_2(K)\}$ is:

$$u_2(t_2(k))[a] = \begin{cases} 1 & \text{if } a = a^*; \\ \frac{k}{K} & \text{if } a = d(u_1(t_1^*)); \\ 0 & \text{otherwise.} \end{cases} \tag{19}$$

This concludes the construction of $\widehat{\mathcal{T}}$.[14] By the consistency of the Bayesian equilibrium (σ_1, σ_2), for all types in T_1 and T_2, σ_1 and σ_2 have to prescribe the same strategies for $\widehat{\mathcal{T}}$ as for $\mathcal{T}^*$. For types $t_2 \in \{t_2(1), t_2(2), \ldots t_2(K)\}$ the strategy $\sigma_2(\widehat{\mathcal{T}}, t_2)$ must be a best response to $\sigma_1(\widehat{\mathcal{T}}, t_1^*)$.

[14]The construction violates our earlier assumption that there are no indifferences. The construction and the argument that follows below can easily be modified to comply with this assumption by assigning the bottom ranked alternatives *almost* the same, but not *exactly* the same utility.

We denote for every $k \in \{0, 1, 2, \ldots, K\}$ by $v_2(k)$ the expected utility of type $t_2(k)$ in the game form G if equilibrium (σ_1, σ_2) is played. By standard incentive compatibility arguments $v_2(k)$ is increasing in k. Observe that, for $k \in \{1, 2, \ldots, K\}$, the difference $v_2(k) - v_2(k-1)$ cannot be more than $1/K$ because, by adopting type $t_2(k)$'s strategy, type $t_2(k-1)$ can always get within $1/K$ of type $t_2(k)$'s expected utility. We also denote for $k \in \{0, 1, 2, \ldots, K\}$ by $r_2(k)$ the equilibrium expected utility of type $t_2(k)$ under random dictatorship. It is immediate that $r_2(k)$ is increasing in k, and that $r_2(k) - r_2(k-1) = 1/K$ for $k = 1, 2, \ldots, K$.

Now consider the difference: $v_2(k) - r_2(k)$. The observations of the previous paragraph imply that as k increases the change in the absolute value of this difference, $|(v_2(k) - r_2(k)) - (v_2(k-1) - r_2(k-1))|$, is at most $1/K$. Note that by choosing K large enough, we can make the step size of changes of this difference arbitrarily small. Observe that $v_2(0) > r_2(0)$ because, by the assumption of the indirect proof, in the game form G, type $t_2(k)$ has a strategy that implies that alternative a^* is chosen with a probability larger than $1 - p$, so that in equilibrium type $t_2(0)$ must obtain alternative a^* with at least that probability. By contrast, under random dictatorship, alternative a^* is chosen with probability $1 - p$ only. On the other hand, $v_2(K) \leq r_2(K)$, because random dictatorship yields for agent $t_2(K)$ at least one of his top alternatives with probability 1. What we have said so far implies that we can find some $k \in \{0, 1, 2, ..., K\}$ such that $v_2(k) - r_2(k)$ is strictly positive but arbitrarily close to zero, provided we choose K large enough.

Next we note that $v_2(k) > r_2(k)$ implies that

$$\sum_{s_1 \in S_1} x(s_1, s_2)[a^*] \cdot \sigma_1(\widehat{\mathcal{T}}, t_1^*)[s_1] > 1 - p \tag{20}$$

for every pure strategy $s_2 \in S_2$ in the support of $\sigma_2(\widehat{\mathcal{T}}, t_2(k))$. This is because $v_2(k) > r_2(k)$ implies that every strategy in the support of $\sigma_2(\widehat{\mathcal{T}}, t_2(k))$ must yield strictly higher expected utility for type $t_2(k)$ than p-random dictatorship would give to this type. Moreover, the only way in which type $t_2(k)$ can be better off under G and (σ_1, σ_2) than under p-random dictatorship, where $d(u_1(t_1^*))$ and a^* are chosen with probabilities p and $1 - p$ respectively, is by raising the probability of a^* above $1 - p$.

Next, we denote for every $k \in \{0, 1, 2, \ldots, K\}$ by $v_1(k)$ the expected utility of type t_1^* when he encounters type $t_2(k)$, and we denote by $r_1(k)$ the expected utility under p-random dictatorship of type t_1^* when he encounters type $t_2(k)$. We first observe that whenever $v_2(k) > r_2(k)$ we must have: $v_1(k) < r_1(k)$. This is because p-random dictatorship would give $d(u_1(t_1^*))$ and a^* with probability p and $1 - p$. By contrast, the game form G gives in equilibrium a^* with a probability that is larger than $1 - p$. Therefore, the outcome will be worse than random dictatorship for player 1. Now consider all pure strategies of player 2 that, matched with type t_1^*'s equilibrium strategy, yield a probability of a^* of more than $1 - p$. As observed before, $v_2(k) > r_2(k)$ implies that type $t_2(k)$ can only play such strategies with positive support. Against each of these strategies player 1 obtains a maximum utility strictly lower than $r_1(k)$. Therefore, there is $\ell > 0$ such that $v_2(k) > r_2(k)$ implies: $v_1(k) < r_1(k) - \ell$.

Now choose K large enough so that we can find a type $t_2(k)$ for whom $v_2(k) > r_2(k)$, but $v_2(k) < r_2(k) + \ell$. We then have: $v_1(k) < r_1(k) - \ell$, and therefore,

adding the last two inequalities: $v_1(k)+v_2(k) < r_1(k)+v_r(k)$. This contradicts the hypothesis that G and (σ_1, σ_2) ex post utilitarian dominate p-random dictatorship.

*S*tep 2: We now complete the proof by showing that no game form G and equilibrium (σ_1, σ_2) of G for all type spaces that has have the properties described in Step 1 can ex post utilitarian dominate p-random dictatorship. The proof is indirect. Suppose there were some game form G and some equilibrium (σ_1, σ_2) of G for all type spaces that have the properties described in Step 1 and that ex post utilitarian dominate p-random dictatorship. Then there must be some type space $\mathcal{T}^{**}$ and some $(t_1^{**}, t_2^{**}) \in T_1^{**} \times T_2^{**}$such that:

$$\sum_{i\in\{1,2\}} \sum_{s_1\in S_1, s_2\in S_2} u_i(t_i^{**})[x(s_1,s_2)] \cdot \sigma_1(\mathcal{T}^{**}, t_1^{**})[s_1] \cdot \sigma_2(\mathcal{T}^{**}, t_2^{**})[s_2] >$$

$$pu_1(d(u_1(t_1^{**}))) + (1-p)u_2(d(u_2(t_2^{**}))) \tag{21}$$

We now construct a new type space, $\widetilde{\mathcal{T}}$, and show that in this type space there is a vector of types such the outcome prescribed by the equilibrium (σ_1, σ_2) yields lower utilitarian welfare than p-random dictatorship. This contradicts the assumption that G and (σ_1, σ_2) ex post utilitarian dominate p-random dictatorship. The construction and the argument below are very similar to, but not identical to, the argument in Step 1.

Before we begin the construction we note that it must be that in equilibrium, at t^{**}, either $d(u_1(t^{**}))$ is chosen with probability strictly less than p, or $d(u_2(t^{**}))$ is chosen with probability strictly less than $1-p$, or both. Otherwise, the game form G with the equilibrium (σ_1, σ_2) could not yield strictly higher utilitarian welfare at t^{**} than p-random dictatorship. Without loss of generality, we focus on the case that $d(u_1(t^{**}))$ is chosen with probability strictly less than p. The other case can be dealt with by a symmetric argument. Let a^{**} be the second most preferred alternative of agent 1 at t_1^{**}.

We now construct $\widetilde{\mathcal{T}}$. The type sets are given by: $\widetilde{T}_1 = T_1^{**}$, and $\widetilde{T}_2 = T_2^{**} \cup \{t_2(1), \ldots, t_2(K)\}$ where $K \in \mathbb{N}$ is large enough. We define later how large K needs to be. The types that are contained in T_1^{**} or T_2^{**} have the same utility functions and beliefs in $\widetilde{\mathcal{T}}$ as in $\mathcal{T}^{**}$. For types $t_2 \in \{t_2(1), t_2(2), \ldots t_2(K)\}$ the beliefs are given by:

$$\pi_2(t_2(k))[t_1^{**}] = 1. \tag{22}$$

The utility function of types $t_2 \in \{t_2(1), t_2(2), \ldots t_2(K)\}$ is:

$$u_2^*(t_2(k))[a] = \begin{cases} 1 & \text{if } a = a^{**}; \\ \frac{k}{K} & \text{if } a \neq d(u_1(t_1^{**})) \text{ and } a \neq a^{**}; \\ 0 & \text{if } a = d(u_1(t_1^{**})). \end{cases} \tag{23}$$

This concludes the construction of $\widetilde{\mathcal{T}}$.[15] By the consistency of the Bayesian equilibrium (σ_1, σ_2), for all types in T_1^{**} and T_2^{**}, σ_1 and σ_2 have to prescribe the same

[15]The construction violates our earlier assumption that there are no indifferences. The construction and the argument that follows below can easily be modified to comply with this assumption by assigning to the middle ranked alternatives *almost* the same, but not *exactly* the same utility.

strategies for $\widetilde{\mathcal{T}}$ as for $\mathcal{T}^{**}$. For types $t_2 \in \{t_2(1), t_2(2), \dots t_2(K)\}$ the strategy $\sigma_2(\widetilde{\mathcal{T}}, t_2)$ must be a best response to $\sigma_1(\widetilde{\mathcal{T}}, t_1^{**})$.

We denote for every $k \in \{0, 1, 2, \dots, K\}$ by $v_2(k)$ the equilibrium expected utility of type $t_2(k)$ in the game form G with equilibrium (σ_1, σ_2). By standard incentive compatibility arguments $v_2(k)$ is increasing in k. Observe that, for $k \in \{1, 2, \dots, K\}$, the difference $v_2(k) - v_2(k-1)$ cannot be more than $1/K$ because, by adopting type $t_2(k)$'s strategy, type $t_2(k-1)$ can always get within $1/K$ of type $t_2(k)$'s expected utility. We also denote for $k \in \{0, 1, 2, \dots, K\}$ by $r_2(k)$ the equilibrium expected utility of type $t_2(k)$ under random dictatorship. It is immediate that $r_2(k) = 1 - p$ for all $k = 1, 2, \dots, K$.

Now consider the difference: $v_2(k) - r_2(k)$. The observations of the previous paragraph imply that as k increases the difference increases, and that moreover it can change by at most $1/K$. Note that by choosing K large enough, we can make the step size of changes of this difference arbitrarily small. Observe next that $v_2(0) \leq r_2(0)$. This is because under G and (σ_1, σ_2) alternative a^{**} can be chosen with a probability of at most $1 - p$, by Step 1 of this proof applied to player 1. Therefore, $v_2(0) \leq 1 - p = r_2(0)$. Finally, we show that $v_2(1) > r_2(1)$. Observe that $v_2(1) > 1 - p$, because, by assumption, the probability of $d(u_1(t_1^{**}))$ under G and (σ_1, σ_2) at (t_1^{**}, t_2^{**}) is strictly less than p. Thus the probability of all other alternatives together must be strictly more than $1 - p$. Type $t_2(1)$ can choose the same strategy as type t_2^{**}, and therefore, if type $t_2(1)$ chooses optimally, $v_2(1) > 1 - p = r_2(1)$. What we have said so far implies that we can find some $k \in \{0, 1, 2, ..., K\}$ such that $v_2(k) - r_2(k)$ is strictly positive but arbitrarily close to zero, provided we choose K large enough.

Next we note that $v_2(k) > r_2(k)$ implies that

$$\sum_{s_1 \in S_1} x(s_1, s_2)[d(u_1(t^{**}))] \cdot \sigma_1(\widetilde{\mathcal{T}}, t_1^{**})[s_1] < p \tag{24}$$

for every pure strategy $s_2 \in S_2$ in the support of $\sigma_2(\widetilde{\mathcal{T}}, t_2(k))$. This is because every strategy in the support of player 2's strategy $\sigma_2(\widetilde{\mathcal{T}}, t_2(k))$ must yield the same expected utility, and hence strictly higher expected utility than $r_2(k)$. But if such a strategy implements $d(u_1(t^{**}))$ with probability of p or more, then the remaining probability that is distributed among a^{**} and all other alternatives, is at most $1 - p$. Therefore, player 2's expected utility from such a strategy is no more than $1 - p = r_2(k)$, which contradicts our assumption that player 2's expected utility is more than $r_2(k)$.

Next, we denote for every $k \in \{0, 1, 2, \dots, K\}$ by $v_1(k)$ the expected utility of type t_1^{**} when he encounters type $t_2(k)$, and we denote by $r_1(k)$ the expected utility under random dictatorship of type t_1^{**} when he encounters type $t_2(k)$. We first observe that whenever $v_2(k) > r_2(k)$ we must have: $v_1(k) < r_1(k)$. This is because random dictatorship would give $d(u_1(t_1^*))$ and a^{**} with probability p and $1 - p$. By contrast, the game form G gives in equilibrium $d(u_1(t_1^{**}))$ with a probability that is less p. Therefore, the outcome will be worse than random dictatorship for player 1. Now consider all pure strategies of player 2 that, matched with type t_1^*'s equilibrium strategy, yield a probability of $d(u_1(t_1^{**}))$ of strictly less than p. As observed before, $v_2(k) > r_2(k)$ implies that type $t_2(k)$ can only play such strategies with positive support. Against each of these strategies player 1 obtains a maximum

utility strictly lower than $r_1(k)$. Therefore, there is $\ell > 0$ such that $v_2(k) > r_2(k)$ implies: $v_1(k) < r_1(k) - \ell$.

Now choose K large enough so that we can find a type $t_2(k)$ for whom $v_2(k) > r_2(k)$, but $v_2(k) < r_2(k) + \ell$. We then have: $v_1(k) < r_1(k) - \ell$, and therefore, adding the last two inequalities: $v_1(k) + v_2(k) < r_1(k) + v_r(k)$. This contradicts the hypothesis that G and (σ_1, σ_2) ex post utilitarian dominate p-random dictatorship. □