Manipulability, Decomposability, and Rationalizability

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit Maastricht
op gezag van de Rector Magnificus,
prof. mr. G.P.M.F. Mols
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen op
dinsdag 21 december 2010 om 10.00 uur

door

Souvik Roy

Promotor:

Prof. dr. H.J.M. Peters

Copromotor:

Dr. A.J.A. Storcken

Beoordelingscommissie:

Prof. dr. A.M. Riedl (voorzitter)

Dr. M. Lombardi

Prof. dr. F. Maniquet (CORE, Louvain-La-Neuve)

To my father Dinobondhu Roy, mother Mridula Roy, and brother Sourav Roy

Preface

I would like to express a sense of love to my father Dinobondhu Roy, mother Mridula Roy, and brother Sourav Roy and friend Kumarjit Saha who have been a source of unwavering love, quiet patience, and tremendous encouragement over the years of my life. I am heartily thankful to my advisor, Prof. dr. H.J.M. Peters, whose encouragement, guidance and support from the initial to the final level enabled me to develop an understanding of the subject. I owe my deepest gratitude to my co-advisor, Dr. A.J.A. Storcken for his meticulous guidance and constant inspiration. I am indebted to Prof. Arunava Sen and Prof. Andrés Perea for their and unwavering support and guidance. I also offer my regards and blessings to all of those who supported me in any respect during the completion of the thesis. Finally, I thank the Department of Quantitative Economics and METEOR for providing financial support for this project.

Contents

Page

CHAPTER 1

Introduction

1. Introduction

Chapters 2-4 of this thesis deal with social choice theory and more specifically with voting, and Chapter 5 with a topic from epistemic game theory.

Social choice theory, as the name suggests, deals with techniques for finding an alternative for a society respecting their preferences over the set of alternatives. Of course, such a technique must satisfy some desirable properties such as strategy-proofness and unanimity. Strategy-proofness ensures that the individuals can not be better off by misrepresenting their true preferences, whereas unanimity implies that if all agents report the same preference, then the rule selects the top of that common preference. However, the classic results of Gibbard (1973) and Satterthwaite (1975) have shown that if we allow for all possible preferences of the individuals then the only rule that satisfies these properties is the dictatorial one. As all the non-dictatorial rules are manipulable, the natural question arises, which one is least manipulable, i.e., manipulable at minimum number of profiles. Furthermore, this impossibility result leaves another question open as to whether in a more restricted context rules other than dictatorships can be strategy-proof. We address these two fundamental questions in the first four chapters of this thesis.

Epistemic game theory is a different approach towards game theory. This theory analyzes different ways a player may reason about his opponents' behavior to make a decision.

Chapter 2 considers approval voting rules. Here we characterize all preference profiles at which the approval (voting) rule is manipulable, under three extensions of preferences to sets of alternatives: by comparison of worst alternatives, best alternatives, or by comparison based on stochastic dominance. We perform a similar exercise for k-approval rules, where voters approve of a fixed number k of alternatives. These results can be used to compare (k-)approval rules with respect to their manipulability. Analytical results are obtained for the case of two voters, specifically, the values of k for which the k-approval rule is minimally manipulable – has the smallest number of manipulable preference profiles – under the various preference extensions are determined. For the number of voters going to infinity, an asymptotic result is that the k-approval rule with k around half the number of alternatives is minimally manipulable among all scoring rules. Further results are obtained by simulation and indicate that k-approval rules may improve on the approval rule as far as manipulability is concerned.

In Chapter 3, we turn to collective decision problems with a finite number of agents who have single-peaked preferences on the real line. H. Moulin (Public Choice 35 (1980), 437-455) has characterized the class of unanimous and strategy-proof deterministic rules in this framework. Here we focus on the probabilistic aspect of the problem. A probabilistic decision scheme assigns a probability distribution over the set of alternatives to every profile of reported preferences. Hereby we show that any unanimous and strategy-proof probabilistic rule can be expressed as a probability mixture, i.e., a convex combination of deterministic rules. Thus we characterize the class of unanimous and strategy-proof probabilistic schemes as a closed and convex set with the extreme points as deterministic rules. This characterization is of great use in solving many other related problems such as finding the mechanism that maximizes ex-ante total expected utility of all agents.

Chapter 4 deals with the characterization of the class of dominant-strategy incentive-compatible (or strategy-proof) random social choice functions in the standard multi-dimensional voting model where voter preferences over the various dimensions (or components) are separable when there are two voters. We show that these social choice functions (which we call generalized random dictatorships) are induced by probability distributions on voter sequences of length equal to the number of components. They induce a fixed probability distribution on the product set of voter peaks. The marginal probability distribution over every component is a random dictatorship. Our results generalize the classic random dictatorship result in Gibbard (1977b) and also show that the decomposability results for strategy-proof deterministic social choice functions for multi-dimensional models with separable preferences obtained in LeBreton and Sen (1999), do not extend straightforwardly to random social choice functions.

The thesis concludes with *Proper rationalizability* (Schuhmacher (1999), Asheim (2001)). Proper rationalizability is a concept in epistemic game theory that is based on two assumptions: (1) every player is *cautious,* i.e., does not exclude any opponent's choice from consideration, and (2) every player *respects* the opponent's preferences, i.e., deems one opponent's choice to be infinitely more likely than another whenever he believes the opponent to prefer the one to the other. In this chapter, we provide a new foundation for proper rationalizability, by assuming that players have incomplete information about the opponents' utilities. We show that, if the uncertainty of each player about the opponents' utilities vanishes gradually in some regular manner, then the choices he can rationally make under common belief in rationality are all properly rationalizable in the original game with no uncertainty about the opponents' utilities.

CHAPTER 2

On the manipulability of approval voting and related scoring rules

1. Introduction

Approval voting is a well accepted voting procedure.[1] In approval voting each voter can approve of as many alternatives as he wants. It is well known (Brams and Fishburn, 1983, and the references therein) that this procedure is strategy-proof (non-manipulable) if preferences are dichotomous, that is, each voter distinguishes only between a set of good and a set of bad alternatives. With more refined preferences, however, strategy-proofness no longer holds.

In this chapter we study the manipulability of the approval (voting) rule and of a related procedure called k-approval (voting) rule. In a k-approval rule each voter approves of exactly k alternatives. This procedure is less flexible than the approval rule – voters can provide less information about their preferences – but tends to be also less manipulable, as we will argue. Therefore, k-approval rules may offer a good compromise between the approval rule and scoring rules such as Borda count.

In Section 2 we introduce the approval rule and next we study its manipulability. Since the approval rule (and also each k-approval rule) is a social choice correspondence and can be multi-valued, we need to make assumptions about extending the preferences (weak orderings) of voters over alternatives to sets of alternatives. We do this in three ways: by comparing the worst alternatives of a set, or by comparing the best alternatives of a set, or by comparing sets on the basis of stochastic dominance using equal chances. In Section 3 we characterize the non-manipulable preference profiles under approval voting for worst, best, and stochastic dominance comparison. The special cases of strict preferences follow as corollaries. Strategy-proofness under dichotomous preferences follows as a special case as well.

In Section 4 we characterize the non-manipulable profiles under k-approval rules, again for worst, best, and stochastic dominance comparison. We also include a brief consideration of a lexicographic refinement of worst and best comparison. For

[1]It is used, for instance, to select candidates for councils of scientific communities such as the Society for Social Choice and Welfare and the Game Theory Society.

technical reasons attention in Section 4 is restricted to strict preferences.

The main purpose of all these exercises is to compare the approval rule and k-approval rules for different values of k with respect to manipulability and under different assumptions about the voters' preferences on sets of alternatives. This comparison is based on a simple measure, namely the number of manipulable preference profiles. The implicit assumption is therefore that all profiles are equally likely. This is called 'impartial culture' in the literature. Unfortunately, a complete analytical comparison is out of the question due to the combinatorial complexity of the problem. For this reason, our comparative results are mainly based on simulations and, thus, they are conjectures and suggestions rather than theorems. A selection of the results of these simulations is presented in Section 5. They give rise to some prudent conclusions concerning the manipulability of the approval and k-approval rules under different assumptions on preference extensions. In particular, they give support to the conjecture that k-approval rules for specific values of k may be less susceptible to manipulation than the approval rule.

Nevertheless, we also present some analytical comparison results. In Section 4.4 we consider the two-voter case and compute the optimal k for different preference extensions, that is, the value of k for which the k-approval rule is minimally manipulable. For $k = 1$, the k-approval rule is just plurality voting. In the two-voter case, this is non-manipulable (strategy-proof) under any reasonable preference extension, including those considered in this chapter. Plurality voting, however, has a serious drawback. If (the) two voters agree on a good second-ranked alternative but disagree on the first, then under plurality voting this compromise is not chosen; it would be chosen, however, under any other k-approval rule. Therefore, for each of the three mentioned preference comparisons and for $k \neq 1$ we have established the overall optimal value of k, and the optimal value under the restriction $k \leq m/2$, where m is the total the number of alternatives. The latter restriction is justified by the desirable property of 'citizen sovereignty': for each alternative there is a preference profile resulting in that alternative as the unique outcome. For $2 \leq k \leq m/2$ we find $k = 2$ as the optimal value in case of best or stochastic dominance comparison, and $k \approx \sqrt{m}$ in case of worst comparison.

On the other extreme, in Section 4.6 we let the number of voters go to infinity and show that even among *all scoring rules* the k-approval rule with $k \in \{(m-1)/2, (m+1)/2\}$ if m is odd, and with $k = m/2$ if m is even, is minimally manipulable. Of course, this result should be interpreted with care, since the probability of manipulability by a single voter is very small anyway if the number of voters is large. The basic intuition for this result is that the (statistical) variance in scores is maximal for the mentioned value(s) of k, so that any single voter's probability of being able to change the outcome is minimal.

Related literature In most voting situations agents have the possibility to manipulate the outcome of the vote by not voting according to their true preferences. The classical theorem of Gibbard (1973) and Satterthwaite (1975) formalizes this fact for social choice functions, which assign a unique alternative to every preference profile, but it also holds for social choice correspondences under various assumptions on preference extensions to sets (e.g., Barberà, Dutta, and Sen, 2001). The present chapter belongs to the strand of literature, initiated by Kelly (1988, 1989), which accepts this phenomenon as a matter of fact and looks for social choice rules which are second best in this respect, i.e., least manipulable. Other references include Fristrup and Keiding (1998) and Aleskerov and Kurbanov (1999). Maus et al. (2007) contains a brief overview of this literature.

Of course, counting the non-manipulable profiles is just one way of measuring the degree of (non-)manipulability of voting rules. Many other approaches are possible (e.g., Saari (1990) or recently Campbell and Kelly, 2008). As already mentioned, our measure of non-manipulability reflects 'impartial culture': each preference profile is implicitly regarded as equally likely. The characterizations of the sets of non-manipulable profiles derived in this chapter, however, are also needed when considering ‘partial culture’.

Notation We denote the cardinality of a set D by $|D|$.

2. Approval Voting

The set of voters is $N = \{1, ..., n\}$ with $n \geq 2$ and the (finite) set of alternatives is A with $|A| = m \geq 3$. A *preference* is a weak ordering on A, i.e., a complete, reflexive, and transitive binary relation on A. By W we denote the set of all preferences. A *preference profile* w is a function from N to W, i.e., an element of W^N. For a preference profile w, $w(i)$ is the preference of voter $i \in N$. For a non-empty subset B of A, $w(i)_{|B}$ denotes the restriction of $w(i)$ to the set B, i.e., $w(i)_{|B} = \{(x, y) \in B \times B \mid (x, y) \in w(i)\}$. Obviously, $w(i)_{|A} = w(i)$.

We next introduce some further notation. Let w be a preference profile and $i \in N$. Let $1 \leq \ell \leq m$ and suppose there exists a set of alternatives B with $|B| = \ell$, $(x, y) \in w(i)$ and $(y, x) \notin w(i)$ for all $x \in B$ and $y \in A \backslash B$. Then we denote this set by $\beta_\ell(w(i))$. Observe that $\beta_\ell(w(i))$ exists if and only if there are ℓ alternatives strictly preferred to the remaining $m - \ell$ alternatives according to $w(i)$; that is, $\beta_\ell(w(i))$ contains only full indifference classes of $w(i)$.

Also, for a subset B of A, by $\beta(w(i)_{|B})$ we denote the set of best elements of B according to $w(i)$, that is, $\beta(w(i)_{|B}) = \{x \in B \mid (x, y) \in w(i) \text{ for all } y \in B\}$. Similarly, $\omega(w(i)_{|B})$ denotes the set of worst elements of B according to $w(i)$, that is, $\omega(w(i)_{|B}) = \{x \in B \mid (y, x) \in w(i) \text{ for all } y \in B\}$. The *lower contour set* of $a \in A$ at $w(i)$ is the set $L(a, w(i)) = \{x \in A \mid (a, x) \in w(i)\}$. Observe that $a \in L(a, w(i))$ by reflexivity of $w(i)$.

In approval voting, each voter $i \in N$ approves of $k(i)$ alternatives, where $1 \leq k(i) \leq m$ is the choice of the voter. The outcome of the vote is the set of those alternatives that receive the largest number of votes. (Observe that excluding $k(i) = 0$ is without loss of generality since the option $k(i) = m$ is available.) To formalize this, a *report* of voter i is a pair $r(i) = (w(i), k(i)) \in W \times \{1, \ldots, m\}$ such that $\beta_{k(i)}(w(i))$ exists. This implies that if a voter approves of an alternative x he also has to approve of all alternatives which are indifferent or strictly preferred to x according to $w(i)$. By R we denote the set of all reports, and by R^N the set of all (report) *profiles*. We denote by

$$\text{score}(x, r) = |\{i \in N \mid x \in \beta_{k(i)}(w(i))\}|$$

the number of voters who approve of alternative $x \in A$ at profile $r = (w, k) = ((w(i), k(i)))_{i \in N} \in R^N$. The *approval rule* φ, defined by

$$\varphi(r) = \{x \in A \mid \text{score}(x, r) \geq \text{score}(y, r) \text{ for all } y \in A\}, \quad r \in R^N,$$

assigns to each profile r the subset of alternatives with maximal score.

We need a few more notations. For $r = (w, k) \in R^N$ and $i \in N$, $\varphi(r_{-i})$ denotes the set of alternatives assigned by the approval rule to the restricted profile $r_{-i} = (r_1, ..., r_{i-1}, r_{i+1}, ..., r_n) \in R^{N \setminus \{i\}}$, that is,

$$\varphi(r_{-i}) = \{x \in A \mid \text{score}(x, r_{-i}) \geq \text{score}(y, r_{-i}) \text{ for all } y \in A\} ,$$

where $\text{score}(x, r_{-i}) = |\{j \in N \setminus \{i\} \mid x \in \beta_{k(j)}(w(j))\}|$. Finally, for (any) $a \in \varphi(r_{-i})$,

$$\varphi^{-}(r_{-i}) = \{x \in A \mid \text{score}(x, r_{-i}) = \text{score}(a, r_{-i}) - 1\}$$

is the (possibly empty) set of those alternatives that have score one less than the elements of $\varphi(r_{-i})$. We call the alternatives in $\varphi(r_{-i})$ *quasi-winners* and those in $\varphi^{-}(r_{-i})$ *almost quasi-winners*.[2] These notations are convenient in view of the following straightforward observation, which will be used throughout the next section:

$$\varphi(r) = \begin{cases} \varphi(r_{-i}) \cap \beta_{k(i)}(w(i)) & \text{if } \varphi(r_{-i}) \cap \beta_{k(i)}(w(i)) \neq \emptyset, \\ \varphi(r_{-i}) \cup [\varphi^{-}(r_{-i}) \cap \beta_{k(i)}(w(i))] & \text{if } \varphi(r_{-i}) \cap \beta_{k(i)}(w(i)) = \emptyset. \end{cases} \tag{2.1}$$

In order to define (non-)manipulability of the approval rule at particular profiles we need to be able to extend individual preferences to preferences over non-empty

[2] Assuming that there is no confusion about the identity of the voter whose vote is left out.

subsets of alternatives. For a voter i in N and a preference $w(i) \in W$, we say that a binary relation $\succeq_{w(i)}$ on $2^A \setminus \{\emptyset\}$ *extends* $w(i)$ if $\{x\} \succeq_{w(i)} \{y\} \Leftrightarrow (x,y) \in w(i)$ holds for all $x, y \in A$. We write $B \succeq_{w(i)} C$ instead of $(B,C) \in \succeq_{w(i)}$. Also, $\succ_{w(i)}$ and $\sim_{w(i)}$ denote the asymmetric and symmetric parts of $\succeq_{w(i)}$, respectively.

In this paper we will consider (three or even more) different ways to extend $w(i)$ over alternatives to a binary relation over non-empty sets of alternatives. Suppose that $\succeq_{w(i)}$ extends $w(i)$ for all $i \in N$. For $i \in N$ and $r, s \in R^N$, we say that r and s are *i-deviations* if $r_{-i} = s_{-i}$. In that case, clearly, $\varphi(r_{-i}) = \varphi(s_{-i})$ and $\varphi^-(r_{-i}) = \varphi^-(s_{-i})$. The approval rule φ *is manipulable by voter i at $r = (w,k)$ towards s* if r and s are i-deviations and $\varphi(s) \succ_{w(i)} \varphi(r)$. The approval rule φ is *not manipulable at r* if for all voters i there is no i-deviation s such that φ is manipulable by i at r towards s.

3. Manipulability of approval voting

The purpose of this section is to characterize the (report) profiles at which the approval rule is not manipulable, for three different preference extensions.

3.1. *Worst comparison*

In this subsection we extend preferences to sets by considering worst alternatives of those sets. Let $i \in N$ and $w(i) \in W$, then we define the extension $\succeq_{w(i)}$ by

$$B \succeq_{w(i)} C \Leftrightarrow (x,y) \in w(i) \text{ for every } x \in \omega\left(w(i)_{|B}\right) \text{ and } y \in \omega\left(w(i)_{|C}\right)$$

for all non-empty sets $B, C \subseteq A$. Thus, B is weakly preferred to C whenever every worst element of B is (weakly) preferred, according to $w(i)$, to every worst element of C.

Theorem 3.1 *Let $r = (w,k) \in R^N$. The approval rule φ is not manipulable at r under worst comparison if and only if for each voter i at least one of the following two statements holds:*

(*a*) $\varphi(r_{-i}) \cap \beta_{k(i)}(w(i)) \neq \emptyset$ *and* $\{x\} \sim_{w(i)} \{y\}$ *for all* $x, y \in \varphi(r_{-i}) \cap \beta_{k(i)}(w(i))$.

(*b*) $\{x\} \sim_{w(i)} \{y\}$ *for all* $x, y \in \varphi(r_{-i})$.

In words, condition (a) requires that if among the quasi-winners there are alternatives belonging to the $k(i)$ highest ranked alternatives of voter i, then i is indifferent between those alternatives; and (b) requires that voter i is indifferent between all quasi-winners.

Proof of Theorem 3.3. For the if-part, let s be an i-deviation of r.

In case (a), it follows by (5.1) that $\varphi(r) = \varphi(r_{-i}) \cap \beta_{k(i)}(w(i))$. By the assumption in (a), $\varphi(r) = \beta\left(w(i)|_{\varphi(r_{-i})}\right)$. Again by (5.1), $\varphi(s) \cap \varphi(r_{-i}) \neq \emptyset$, so for every $x \in \varphi(r) = \beta\left(w(i)|_{\varphi(r_{-i})}\right)$ it follows that $\{x\} \succeq_{w(i)} \varphi(s)$. So, $\varphi(r) \succeq_{w(i)} \varphi(s)$.

Now consider case (b) and assume $\varphi(r_{-i}) \cap \beta_{k(i)}(w(i)) = \emptyset$ otherwise we are done by (a). Then $\varphi(r_{-i}) \subseteq \varphi(r)$, so we have $\omega\left(w(i)_{|\varphi(r)}\right) = \varphi(r_{-i})$. Since, by (5.1), $\varphi(s) \cap \varphi(r_{-i}) \neq \emptyset$, we have again $\varphi(r) \succeq_{w(i)} \varphi(s)$.

For the only if-part, suppose that there is an voter i for whom (a) nor (b) holds. It is sufficient to prove that φ is manipulable at profile r by voter i. Observe that either there exist $x, y \in \varphi(r_{-i}) \cap \beta_{k(i)}(w(i))$ such that $\{x\} \succ_{w(i)} \{y\}$, or $\varphi(r_{-i}) \cap \beta_{k(i)}(w(i)) = \emptyset$ and there exist $x, y \in \varphi(r_{-i})$ such that $\{x\} \succ_{w(i)} \{y\}$. In both cases, by (5.1), $x, y \in \varphi(r)$. Now consider the report $s(i) = (w'(i), 1)$ of voter i such that $\beta(w'(i)) = \{x\}$. Then, by (5.1) again, $\varphi(s) = \{x\} \succ_{w(i)} \varphi(r)$. $\square$

We now consider the subclass of strict of preferences. This will enable us to compare approval voting to k-approval voting, which is studied in the next section.

3.1.1. *Strict preferences.* A preference $w(i)$ is *strict* (or a *linear ordering*) if it is antisymmetric, i.e., $(x, y) \in w(i)$ implies $(y, x) \notin w(i)$ for all $x, y \in A$ with $x \neq y$. Let P denote the set of all linear orderings on A, and S the set of all reports $(w(i), k(i))$ with $w(i) \in P$. The following result considers manipulability of the approval rule φ when restricted to S^N.

Corollary 3.2 *Let $r = (w, k) \in S^N$. The approval rule φ, restricted to S^N, is not manipulable at r under worst comparison if and only if for each voter i at least one of the following two statements holds:*

(a) $|\varphi(r_{-i}) \cap \beta_{k(i)}(w(i))| = 1$.

(b) $|\varphi(r_{-i})| = 1$.

Proof. For the only-if direction, note that if voter i can manipulate via a preference in W, then i can also manipulate by a strict preference, by strictifying the weak preference in any arbitrary way. Thus, the only-if direction follows from Theorem 3.3. The if-direction is immediate from Theorem 3.3. □

3.2. *Best comparison*

In this subsection we extend preferences to sets by considering best alternatives of those sets. Let $i \in N$ and $w(i) \in W$, then we define the extension $\succeq_{w(i)}$ by[3]

$$B \succeq_{w(i)} C \Leftrightarrow (x, y) \in w(i) \text{ for every } x \in \beta\left(w(i)_{|B}\right) \text{ and } y \in \beta\left(w(i)_{|C}\right)$$

for all non-empty sets $B, C \subseteq A$. Thus, B is weakly preferred to C whenever every best element of B is (weakly) preferred, according to $w(i)$, to every best element of C.

Theorem 3.3 *Let $r = (w, k) \in R^N$. The approval rule φ is not manipulable at r under best comparison if and only if for each voter i at least one of the following two statements holds:*

(a) $(x, y) \in w(i)$ *for all* $x \in \beta\left(w(i)_{|\varphi(r_{-i})}\right)$ *and all* $y \in \varphi^{-}(r_{-i})$.

(b) $\varphi(r_{-i}) \cap \beta_{k(i)}(w(i)) = \emptyset$ *and* $\varphi^{-}(r_{-i}) \cap \beta_{k(i)}(w(i)) \neq \emptyset$.

[3] In order to avoid cumbersome notation we will use the same symbols for different preference extensions in this paper.

In words, condition (a) requires that any best alternative among the quasi-winners is preferred by i over all almost quasi-winners; and (b) requires that none of the quasi-winners is among his $k(i)$ highest ranked alternatives, but some of the almost quasi-winners are among his $k(i)$ highest ranked alternatives.

Proof of Theorem 3.3. For the if-part, let s be an i-deviation of r.

In case (a), it follows by (5.1) that there exists $x \in \varphi(r)$ with $x \in \beta\left(w(i)_{|\varphi(r_{-i})}\right)$. So by (a), $\{x\} \succeq_{w(i)} \{y\}$ for all $y \in \varphi(r_{-i}) \cup \varphi^{-}(r_{-i})$. This implies $\varphi(r) \succeq_{w(i)} \varphi(s)$.

In case (b), by (5.1), $\varphi(r) = \varphi(r_{-i}) \cup \left[\varphi^{-}(r_{-i}) \cap \beta_{k(i)}(w(i))\right]$. So $\varphi(r) \succeq_{w(i)} \{x\}$ for all $x \in \varphi(r_{-i}) \cup \varphi^{-}(r_{-i})$. This implies again $\varphi(r) \succeq_{w(i)} \varphi(s)$.

For the only if-part, suppose that there is an voter i for whom (a) nor (b) holds. It is sufficient to prove that φ is manipulable at profile r by voter i. Observe that either (i) $\varphi(r_{-i}) \cap \beta_{k(i)}(w(i)) \neq \emptyset$ and there exists $y \in \varphi^{-}(r_{-i})$ such that $\{y\} \succ_{w(i)} \beta\left(w(i)_{|\varphi(r_{-i})}\right)$; or (ii) $\varphi^{-}(r_{-i}) \cap \beta_{k(i)}(w(i)) = \emptyset$ and there exists $y \in \varphi^{-}(r_{-i})$ such that $\{y\} \succ_{w(i)} \beta\left(w(i)_{|\varphi(r_{-i})}\right)$. Note that, in both cases, $\varphi(r) \subseteq \varphi(r_{-i})$. For both cases, consider the report $s(i) = (w'(i), 1)$ of voter i such that $\beta(w'(i)) = \{y\}$. Then by (5.1), $\varphi(s) = \varphi(r_{-i}) \cup \{y\}$, which implies $\varphi(s) \succ_{w(i)} \varphi(r)$. □

For strict preferences we have the following corollary. The proof is straightforward and therefore omitted.

Corollary 3.4 *Let $r = (w, k) \in S^N$. The approval rule φ, restricted to S^N, is not manipulable at r under best comparison if and only if for each voter i at least one of the following two statements holds:*

(a) $(x, y) \in w(i)$ *for all* $y \in \varphi^{-}(r_{-i})$, *where* $\{x\} = \beta\left(w(i)_{|\varphi(r_{-i})}\right)$.

(b) $\varphi(r_{-i}) \cap \beta_{k(i)}(w(i)) = \emptyset$ *and* $\varphi^{-}(r_{-i}) \cap \beta_{k(i)}(w(i)) \neq \emptyset$.

3.3. *Stochastic dominance comparison*

In this subsection comparisons of sets of alternatives are based on stochastic dominance. To formalize this we need some further notions. Let u be a function from

A to $\mathbb{R}$. Then u is said to be a *utility function representing preference* $w(i)$ of voter i, if for all alternatives x and y in A

$$(x, y) \in w(i) \text{ if and only if } u(x) \geq u(y) .$$

Let B and C be two nonempty subsets of alternatives. Voter i is said to *prefer* B *to* C *according to stochastic dominance at preference* $w(i)$, denoted as $B \succeq_{w(i)} C$, if

$$\sum_{a \in B} \frac{1}{|B|} u(a) \geq \sum_{a \in C} \frac{1}{|C|} u(a) \text{ for every utility function } u \text{ representing } w(i) .$$

This preference extension[4] is based on the idea that, if we attach equal probabilities to the alternatives in each set, then the expected utility of the resulting lottery over B should be at least as high as the expected utility of the resulting lottery over C, for *each* utility function representing $p(i)$. Clearly, and in contrast to worst and best comparison in the preceding sections, this preference extension is not complete: many sets are incomparable. Observe that our notion of manipulability implies that a voter manages to obtain a preferred and thus comparable set.

In the following theorem we characterize the non-manipulable profiles under the stochastic dominance preference extension. To understand the proof, it is sometimes convenient to keep in mind the familiar characterization (or definition) of stochastic dominance involving only probabilities. This characterization says that a lottery ℓ is preferred over another lottery ℓ' if it can be obtained by shifting probability in ℓ' to preferred alternatives.

Theorem 3.5 *Let* $r = (w, k) \in R^N$. *The approval rule* φ *is not manipulable at* r *under stochastic dominance if and only if for each voter* i *at least one of the following three statements holds:*

(a) $\varphi(r_{-i}) \subseteq [A \backslash \beta_{k(i)}(w(i))]$ *and* $\varphi^{-}(r_{-i}) \cap \beta_{k(i)}(w(i)) \neq \emptyset$.

[4]The stochastic dominance criterion to compare sets has been used before, see e.g. Barberà, Dutta and Sen (2001).

(b) $\varphi(r_{-i}) \cap \beta_{k(i)}(w(i)) \neq \emptyset$ *and* $\{x\} \sim_{w(i)} \{y\}$ *for all* $x, y \in \varphi(r_{-i}) \cap \beta_{k(i)}(w(i))$ *and* $[A \backslash \beta_{k(i)}(w(i))] \cap \varphi(r_{-i}) \neq \emptyset$.

(c) $\{x\} \sim_{w(i)} \{y\}$ *for all* $x, y \in \varphi(r_{-i})$ *and* $\varphi^{-}(r_{-i}) \subseteq L(x, w(i))$ *for some* $x \in \varphi(r_{-i})$.

In words, these three cases can be described as follows. In case (a), no quasi-winner but at least one almost quasi-winner belongs to the $k(i)$ highest ranked alternatives. In case (b) there are quasi-winners among the $k(i)$ highest ranked alternatives and voter i is indifferent between them, but there are also lower ranked quasi-winners. In case (c) voter i is indifferent between the quasi-winners, and all almost quasi-winners are lower ranked than some of the quasi-winners.

For a proof of this theorem see the Appendix.

The following corollary (proof omitted) applies to strict preferences.

Corollary 3.6 *Let* $r = (w, k) \in S^N$*. The approval rule* φ*, restricted to* S^N*, is not manipulable at* r *under stochastic dominance comparison if and only if for each voter* i *at least one of the following three statements holds:*

(a) $\varphi(r_{-i}) \subseteq [A \backslash \beta_{k(i)}(w(i))]$ *and* $\varphi^{-}(r_{-i}) \cap \beta_{k(i)}(w(i)) \neq \emptyset$.

(b) $\varphi(r_{-i}) \cap \beta_{k(i)}(w(i)) = \{x\}$ *for some* $x \in A$ *and* $[A \backslash \beta_{k(i)}(w(i))] \cap \varphi(r_{-i}) \neq \emptyset$.

(c) $\varphi(r_{-i}) = \{x\}$ *for some* $x \in A$ *and* $\varphi^{-}(r_{-i}) \subseteq L(x, w(i))$.

3.4. *Dichotomous preferences*

A preference $w(i) \in W$ is *dichotomous* if it has two indifference classes, i.e., there are disjoint subsets $B_1 \neq \emptyset$ and B_2 of A such that $A = B_1 \cup B_2$, $(x, y), (y, x) \in w(i)$ for all $x, y \in B_1$ and for all $x, y \in B_2$, and $(x, y) \in w(i)$, $(y, x) \notin w(i)$ for all $x \in B_1$ and $y \in B_2$. Let $D \subseteq W$ denote the subset of all dichotomous preferences. A report $r(i) = (w(i), k(i))$ is in R_d if $w(i)$ is dichotomous and $k(i)$ is the cardinality of the higher indifference class of $w(i)$, i.e., $k(i) = |B_1|$ in the notation above[5]. A report $r(i) \in R_d$ is called *dichotomous* as well. In the following corollary we show that the approval rule is *strategy-proof* when restricted to dichotomous report profiles, under all three preference extensions considered in this paper: this means that φ is manipulable at no $r \in R_d^N$ under any of these preference extensions. This result confirms well known results on approval voting, see Brams and Fishburn (1983) and the references therein.

Corollary 3.7 *The approval rule φ, restricted to R_d^N, is strategy-proof under the worst, best, and stochastic dominance preference extensions.*

Proof. Let $r \in R_d^N$, $r(i) = (w(i), k(i))$ for all $i \in N$.

Suppose that for some $j \in N$ statement (b) in Theorem 3.3 does not hold. Then there is $x \in \varphi(r_{-j})$ with $x \in \beta_{k(j)}(w(j))$, and, clearly, $\{x\} \sim_{w(j)} \{y\}$ for all $x, y \in \varphi(r_{-j}) \cap \beta_{k(j)}(w(j))$. Hence, (a) holds for j. Thus, φ is strategy-proof under worst comparison.

Next, suppose that (a) in Theorem 3.3 does not hold for some $i \in N$. Then there is a $y \in \varphi^{-}(r_{-i})$ with $(y, x) \in w(i)$ and $(x, y) \notin w(i)$ for all $x \in \varphi(r_{-i})$. This implies that (b) holds for i. Thus, φ is strategy-proof under best comparison.

Finally, suppose (c) in Theorem 3.5 does not hold for some $i \in N$. There are two cases. If the first statement in (c) does not hold, then $\varphi(r_{-i}) \cap \beta_{k(i)}(w(i)) \neq \emptyset$ and $\varphi(r_{-i}) \cap \left[A \setminus \beta_{k(i)}(w(i))\right] \neq \emptyset$, so that (b) holds. If the second statement in (c) does not hold, then there is $y \in \varphi^{-}(r_{-i})$ with $\{y\} \succ_{w(i)} \{x\}$ for some $x \in \varphi(r_{-i})$. In this case, if $\varphi(r_{-i}) \subseteq \left[A \setminus \beta_{k(i)}(w(i))\right]$ then (a) holds, and otherwise (b) holds. Thus, φ is strategy-proof under stochastic dominance comparison. $\square$

4. Manipulability of k-approval voting

A variation on approval voting is obtained by fixing the number of alternatives that has to be approved by each voter. Specifically, for a profile $p \in P^N$ of strict

[5]Observe that, in this case, $k(i)$ is uniquely determined by $w(i)$.

preferences, an alternative $x \in A$, and a number $k \in \{1, \ldots, m-1\}$, we denote by the *k-score*

$$\text{score}_k(x, p) = |\{i \in N \mid x \in \beta_k(p(i))\}|$$

the total number of voters for who alternative x is among the k first ranked alternatives at a profile p. The *k-approval rule* φ_k, defined by

$$\varphi_k(p) = \{x \in A \mid \text{score}_k(x, p) \geq \text{score}_k(y, p) \text{ for all } y \in A\}, \quad p \in P^N$$

assigns to each profile p the subset of alternatives with maximal k-score.[6]

Observe that it is, indeed, convenient to restrict attention to strict preferences, since otherwise we might have to split up indifference classes due to the fact that the number of alternatives to be approved is now fixed.

The sets $\varphi_k(p_{-i})$ and $\varphi_k^-(p_{-i})$ of quasi-winners and almost quasi-winners are defined analogously as for the approval rule. Also, we have the following useful observation:

$$\varphi_k(p) = \begin{cases} \varphi_k(p_{-i}) \cap \beta_k(p(i)) & \text{if } \varphi_k(p_{-i}) \cap \beta_k(p(i)) \neq \emptyset, \\ \varphi_k(p_{-i}) \cup \left[\varphi_k^-(p_{-i}) \cap \beta_k(p(i))\right] & \text{if } \varphi_k(p_{-i}) \cap \beta_k(p(i)) = \emptyset, \end{cases} \tag{2.2}$$

for all $p \in P^N$, $i \in N$, and $1 \leq k \leq m-1$.

In what follows we characterize the profiles of preferences at which the k-approval rule is not manipulable for different preference extensions, starting with worst, best, and stochastic dominance comparison. The definitions of (non)-manipulability of φ_k at a profile p are completely analogous to those for the approval rule.

4.1. *Worst comparison*

For the definition of the worst comparison preference extension see Section 3.1.

The following theorem characterizes all profiles at which the k-approval rule is not manipulable under worst comparison.

[6]Unlike the approval rule the k-approval rule is a scoring rule, see Section 4.6. Note, further, that the case $k = m$ is uninteresting.

Theorem 4.1 *Let $p \in P^N$. The k-approval rule φ_k is not manipulable at p under worst comparison if and only if for each voter i at least one of the following three statements holds:*

(a) $|\varphi_k(p_{-i}) \cap \beta_k(p(i))| = 1$.

(b) $|\varphi_k(p_{-i})| = 1$.

(c) $A \setminus \beta_k(p(i)) \subsetneq \varphi_k(p_{-i})$.

In words, condition (a) requires that exactly one of that voter i's k highest ranked alternatives is a quasi-winner; (b) requires that there is a unique quasi-winner; and (c) requires that the quasi-winners are a strict subset of the $m-k$ lowest ranked alternatives.

Proof of Theorem 4.1. For the if-part, let $i \in N$ and let q be an i-deviation of p. Assume that at least one of the cases (a), (b), and (c) holds. We show that voter i cannot manipulate from p to q.

In case (a), let $\{x\} = \varphi_k(p_{-i}) \cap \beta_k(p(i))$. By (2.2), $\varphi_k(p) = \{x\}$. Again by (2.2), either $\varphi_k(q) \subseteq \varphi_k(p_{-i})$ or $\varphi_k(p_{-i}) \subseteq \varphi_k(q)$. In the first case, if $x \in \varphi_k(q)$, then $\varphi_k(p) = \{x\} \succeq_{p(i)} \varphi_k(q)$; if $x \notin \varphi_k(q)$ then $\varphi_k(q) \subseteq A \setminus \beta_k(p(i))$ so that again $\varphi_k(p) = \{x\} \succeq_{p(i)} \varphi_k(q)$. In the second case, $\varphi_k(p) = \{x\} \subseteq \varphi_k(q)$, hence $\varphi_k(p) \succeq_{p(i)} \varphi_k(q)$.

In case (b), let $\varphi_k(p_{-i}) = \{x\}$ for some alternative x. If $x \in \beta_k(p(i))$ we are done by case (a). If $x \notin \beta_k(p(i))$ then by (2.2), $\varphi_k(p) = \{x\} \cup \left[\varphi_k^-(p_{-i}) \cap \beta_k(p(i))\right]$ and, thus, $\omega(p(i)_{|\varphi_k(p)}) = x$. Further, also by (2.2), $\varphi_k(q) = \{x\}$ or $\varphi_k(q) = \{x\} \cup \left[\varphi_k^-(p_{-i}) \cap \beta_k(q(i))\right]$; in both cases, $(x, \omega(p(i)_{|\varphi_k(q)})) \in p(i)$ and, thus, $\varphi_k(p) \succeq_{p(i)} \varphi_k(q)$.

In case (c), by (2.2) we have $\varphi_k(p) = \varphi_k(p_{-i}) \cap \beta_k(p(i))$ and $\varphi_k(q) = \varphi_k(q_{-i}) \cap \beta_k(q(i)) = \varphi_k(p_{-i}) \cap \beta_k(q(i))$. If $\beta_k(q(i)) = \beta_k(p(i))$ then $\varphi_k(p) = \varphi_k(q)$. Otherwise, since $A \setminus \beta_k(p(i)) \subsetneq \varphi_k(p_{-i})$, there is a $y \in [A \setminus \beta_k(p(i))] \cap \varphi_k(q)$. Hence, $\varphi_k(p) \succeq_{p(i)} \varphi_k(q)$.

For the only-if part, suppose that there is a voter $i \in N$ such that none of the three cases (a), (b), and (c) holds. It is sufficient to prove that φ_k is manipulable at profile p by voter i. For this, in turn, it is sufficient to prove that i can manipulate at profile p for the following two cases.

Case (i): $\varphi_k(p_{-i}) \cap \beta_k(p(i)) = \emptyset$ and $|\varphi_k(p_{-i})| \geq 2$.

Let $b = \beta\left(p(i)_{|\varphi_k(p_{-i})}\right)$. Take $q(i)$ such that the positions in $p(i)$ of b and one of the alternatives in $\beta_k(p(i))$ are swapped. Then $\varphi_k(q) = \{b\}$ and $\varphi_k(q) \succ_{p(i)} \varphi_k(p)$, hence voter i can manipulate at profile p towards q.

Case (ii): $|\varphi_k(p_{-i}) \cap \beta_k(p(i))| \geq 2$ and $[A \setminus \beta_k(p(i))] \nsubseteq \varphi_k(p_{-i})$.

Let $w = \omega\left(p(i)_{|\varphi_k(p_{-i}) \cap \beta_k(p(i))}\right)$ and $y \in A \setminus [\beta_k(p(i)) \cup \varphi_k(p_{-i})]$. Let $q(i)$ be obtained from $p(i)$ by swapping the positions of the alternatives w and y. By (2.2), $\varphi_k(p) = \varphi_k(p_{-i}) \cap \beta_k(p(i))$ and $\varphi_k(q) = \varphi_k(p_{-i}) \cap \beta_k(p(i)) \setminus \{w\}$ it follows that $\varphi_k(q) \succ_{p(i)} \varphi_k(p)$, proving that φ_k is manipulable by voter i at profile p towards q.$\square$

4.2. *Best comparison*

For the definition of the best comparison preference extension see Section 3.2.

The following theorem characterizes all profiles at which the k-approval rule is not manipulable under best comparison.

Theorem 4.2 *Let $p \in P^N$. The k-approval scoring rule φ_k is not manipulable at p under best comparison if and only if for each voter i at least one of the following three statements holds:*

(*a*) $\left(\beta(p(i)_{|\varphi_k(p_{-i})}), x\right) \in p(i)$ *for all* $x \in \varphi_k^-(p_{-i})$.

(*b*) $\varphi_k(p_{-i}) \cap \beta_k(p(i)) = \emptyset$ *and* $\varphi_k^-(p_{-i}) \cap \beta_k(p(i)) \neq \emptyset$.

(*c*) $|\varphi_k(p_{-i}) \cap \beta_k(p(i))| > |A \setminus [\beta_k(p(i)) \cup \varphi_k(p_{-i})]|$.

In words, condition (a) requires that the best alternative among the quasi-winners is preferred over all almost quasi-winners; (b) requires that no quasi-winner is among his k first ranked alternatives, but some of the almost quasi-winners are; and (c) requires that the number of the voter's k highest ranked alternatives among the quasi-winners is larger than the number of alternatives that are neither among his k highest ranked nor among the quasi-winners.

Proof of Theorem 4.2. For the if-part, let q be an i-deviation of p. Note that $\varphi_k(p_{-i}) = \varphi_k(q_{-i})$ and $\varphi_k^-(p_{-i}) = \varphi_k^-(q_{-i})$. Assume that at least one of the cases (a), (b), and (c) holds. We show that voter i cannot manipulate from p to q.

In case (a), for both cases occurring in (2.2), we obtain $\beta\left(p(i)_{|\varphi_k(p)}\right) = \beta\left(p(i)_{|\varphi_k(p_{-i})}\right)$. Since $\beta\left(p(i)_{|\varphi_k(q)}\right) \in \varphi_k(p_{-i}) \cup \varphi_k^-(p_{-i})$ and by the assumption for case (a), we conclude that $\varphi_k(p) \succeq_{p(i)} \varphi_k(q)$.

In case (b), again using (2.2), we have $\varphi_k(p) = \varphi_k(p_{-i}) \cup [\varphi_k^-(p_{-i}) \cap \beta_k(p(i))]$, hence $\beta\left(p(i)_{|\varphi_k(p)}\right) = \beta\left(p(i)_{|\varphi_k^-(p_{-i})\cap\beta_k(p(i))}\right)$; and $\varphi_k(q) \in \varphi_k(p_{-i}) \cup [\varphi_k^-(p_{-i}) \cap \beta_k(q(i))]$. By the assumptions for this case, $\varphi_k(p) \succeq_{p(i)} \varphi_k(q)$.

In case (c), it is easy to see that $|A \setminus \varphi_k(p_{-i})| < |\beta_k(p(i))| = k = |\beta_k(q(i))|$, hence $\beta_k(q(i)) \cap \varphi_k(p_{-i}) \neq \emptyset$. Therefore, by (2.2) we have $\varphi_k(p) = \varphi_k(p_{-i}) \cap \beta_k(p(i))$ and $\varphi_k(q) = \varphi_k(p_{-i}) \cap \beta_k(q(i)) \subseteq \varphi_k(p_{-i})$. Thus, also in this case $\varphi_k(p) \succeq_{p(i)} \varphi_k(q)$.

For the only-if part, suppose that there is a voter $i \in N$ such that none of the three cases (a), (b), and (c) holds. It is sufficient to prove that φ_k is manipulable at profile p by voter i. For this, in turn, it is sufficient to prove that i can manipulate at profile p for the following two cases.

Case (i): There is an $x \in \varphi_k^-(p_{-i})$ such that $(x, b) \in p(i)$, where $b = \beta\left(\varphi_k(p_{-i}), p(i)_{|\varphi_k(p_{-i})}\right)$; $\varphi_k(p_{-i}) \cap \beta_k(p(i)) \neq \emptyset$; and $|\varphi_k(p_{-i}) \cap \beta_k(p(i))| \leq |A \setminus [\beta_k(p(i)) \cup \varphi_k(p_{-i})]|$.

For this case, note that $x \in \beta_k(p(i))$. By the assumptions for this case we can take a $q(i) \in P$ with $x \in \beta_k(q(i))$ and $\varphi_k(p_{-i}) \cap \beta_k(q(i)) = \emptyset$. Hence, $x \in \varphi_k(q) \setminus \varphi_k(p)$ and, thus, $\varphi_k(q) \succ_{p(i)} \varphi_k(p)$. So i can manipulate at profile p towards q.

Case (ii): There is an $x \in \varphi_k^-(p_{-i})$ such that $(x, b) \in p(i)$, where $b = \beta\left(p(i)_{|\varphi_k(p_{-i})}\right)$; and $\varphi_k^-(p_{-i}) \cap \beta_k(p(i)) = \emptyset$.

In this case, $\varphi_k(p) = \varphi_k(p_{-i})$. Note that the sets $\beta_k(p(i))$, $\varphi_k(p_{-i})$, and $\varphi_k^-(p_{-i})$ are pairwise disjoint. So we can take $q(i) \in P$ such that $x \in \beta_k(q(i))$ and $\varphi_k(p_{-i}) \cap \beta_k(q(i)) = \emptyset$. Then $\varphi_k(q) \supseteq \varphi_k(p_{-i}) \cup \{x\}$, so $x \in \varphi_k(q) \setminus \varphi_k(p)$, thus $\varphi_k(q) \succ_{p(i)} \varphi_k(p)$ and i can manipulate at profile p towards q. □

4.3. *Stochastic dominance comparison*

For the definition of the stochastic dominance comparison preference extension see Section 3.3.

The following theorem characterizes all profiles at which the k-approval rule is not manipulable under stochastic dominance comparison. Its proof is placed in the Appendix.

Theorem 4.3 *Let $p \in P^N$. The k-approval scoring rule φ_k is not manipulable at p under stochastic dominance comparison if and only if for all voters i at least one of the following five statements holds:*

(*a*) $A \setminus \beta_k(p(i)) \subsetneq \varphi_k(p_{-i})$.

(*b*) $\varphi_k(p_{-i}) \subseteq [A \setminus \beta_k(p(i))]$ *and* $\varphi_k^-(p_{-i}) \cap \beta_k(p(i)) \neq \emptyset$.

(*c*) $\varphi_k(p_{-i}) \cap \beta_k(p(i)) = \{w\}$ *for some* $w \in A$ *and* $[A \setminus \beta_k(p(i))] \cap \varphi_k(p_{-i}) \neq \emptyset$.

(*d*) $\varphi_k(p_{-i}) = \{w\}$ *for some* $w \in A$ *and* $\varphi_k^-(p_{-i}) \subseteq L(w, p(i))$.

(*e*) $\varphi_k(p_{-i}) = \{w\}$ *for some* $w \in A$ *and* $|\varphi_k^-(p_{-i}) \cap L(w, p(i))| > m - k$.

In words, these five cases can be described as follows. In case (a), at least one of voter i's k highest ranked alternatives and all of his lower ranked alternatives are quasi-winners. In case (b), no quasi-winner but at least one almost quasi-winner is among his k highest ranked alternatives. In case (c) there is a unique quasi-winner among voter i's k highest ranked alternatives, but there are lower ranked quasi-winners as well. In case (d) there is a unique quasi-winner, which is preferred by i to all almost quasi-winners. In case (e) there is again a unique quasi-winner, and among the almost quasi-winners there are more than $m-k$ alternatives worse than the unique quasi-winner.

4.4. *The two-voter case*

In this subsection we concentrate on the two-voter case and consider the following question: which k-approval rule is least (or minimally) manipulable, under various assumptions on preference extensions as studied in the preceding sections?

We start with a simple theorem, which will be derived from Theorems 4.1, 4.2, and 4.3, but also easily follows directly. It states that φ_1 is *strategy-proof*, i.e., not manipulable at any profile p.

Theorem 4.4 *Let $n = 2$. Then the 1-approval rule φ_1 is strategy-proof under worst, best, and stochastic dominance comparison.*

Proof. Let $p = (p(1), p(2))$ be a preference profile and let $k = 1$. Note that (b) in Theorem 4.1 is always satisfied: this shows strategy-proofness under worst comparison. In Theorem 4.2, (a) reduces to $\beta(p(1)) = \beta(p(2))$ and (b) to $\beta(p(1)) \neq \beta(p(2))$: this shows strategy-proofness under best comparison. Finally, in Theorem 4.3, (b) reduces to $\beta(p(1)) \neq \beta(p(2))$ and (d) to $\beta(p(1)) = \beta(p(2))$: this shows strategy-proofness under stochastic dominance comparison. □

This observation might make our quest for minimally manipulable rules futile, were it not the case that the 1-approval rule (i.e., plurality rule) is not unambiguously attractive. As an example, consider the case where voter 1 has preference $p(1)$: $xz\ldots y$ and voter 2 has preference $p(2)$: $yz\ldots x$ (notations obvious). Then $\varphi_1(p) = \{x, y\}$ but $\varphi_2(p) = \{z\}$. So it seems that φ_2 offers a better compromise in this case than φ_1.

Moreover, for more than two voters and apart from a few particular cases, Theorem 4.4 no longer holds.

We will now consider the three cases (worst, best, and stochastic dominance comparison) separately.

4.4.1. *Worst comparison for two voters.* The non-manipulable profiles for two voters under worst comparison are easily described using Theorem 4.1.

Corollary 4.5 *Let $n = 2$ and $2 \leq k < m$. Let $p \in P$ and consider worst comparison.*

(*a*) *If $k \leq (m+1)/2$, then φ_k is not manipulable at p if and only if $|\varphi_k(p)| = 1$, or equivalently,*

$$|\beta_k(p(i)) \cap \beta_k(p(2))| = 1 \ .$$

(*b*) *If $k > (m+1)/2$, then φ_k is not manipulable at p if and only if $|\varphi_k(p)| = 2k-m$, or equivalently,*

$$|\beta_k(p(i)) \cap \beta_k(p(2))| = 2k - m \ .$$

Proof. Case (b) in Theorem 4.1 does not apply. If case (a) in Theorem 4.1 applies then we have $|\beta_k(p(1)) \cap \beta_k(p(2))| = 1$ (or, equivalently, $|\varphi_k(p)| = 1$), but this is possible if and only if $k \leq (m+1)/2$. If case (c) in Theorem 4.1 applies then we have $|\beta_k(p(1)) \cap \beta_k(p(2))| = 2k - m$ (or, equivalently, $|\varphi_k(p)| = 2k - m$), but this is possible if and only if $k \geq (m+1)/2$; but for $k = (m+1)/2$ we have $2k - m = 1$, so that we are back in case (a). □

Denote by $\eta(m,k)$ the number of profiles (for two voters) at which φ_k is not manipulable. By straightforward counting we obtain the following result for the number of manipulable profiles for two voters under worst comparison.

Theorem 4.6 *Let $n = 2$ and $2 \leq k < m$. Consider worst comparison. Then*

$$\eta(m,k) = \begin{cases} m!\, k \binom{m-k}{k-1} k!\, (m-k)! & \text{if } k \leq (m+1)/2 \\ m! \binom{k}{2k-m} k!\, (m-k)! & \text{if } k > (m+1)/2 \ . \end{cases}$$

From this theorem we derive the following corollary (see the Appendix for a proof), which states some facts about k as far as non-manipulability is concerned. (The exact meaning of k^* being close to $\sqrt{m}$ in part (a) is explained in the proof.)

Corollary 4.7

(a) *$\eta(m,k)$ increases in k between 2 and an integer k^*, which is close to $\sqrt{m}$, and decreases between k^* and $\frac{1}{2}(m-1)$.*

(b) *$\eta(m,k)$ increases between $\frac{1}{2}(m-1)$ and $(m-1)$.*

(c) *The $(m-1)$-approval scoring rule is second best since $\eta(m,(m-1)) > \eta(m,k)$ for all $m-1 > k \geq 2$.*

The first-best value of k is $k = 1$ (Theorem 4.4), but φ_1 has the drawback that it does not give much opportunity for compromises. Among other values of k, the value $k = m-1$ is best. We might, however, prefer to have $k \leq (m+1)/2$, for the following reason. Call φ_k *citizen-sovereign* if for every alternative $x \in A$ there is a profile $p \in P$ with $\varphi_k(p) = \{x\}$. It is not difficult to see that φ_k is citizen-sovereign for any number of voters $n \geq 2$ if $k \leq (m+1)/2$. For $n = 2$ and $k > (m+1)/2$, however, φ_k is not citizen-sovereign. Hence, if we restrict ourselves to citizen-sovereign rules with $k \geq 2$, then the best rule is φ_{k^*}, where k^* is close to $\sqrt{m}$.

4.4.2. *Best comparison for two voters.* The non-manipulable profiles for two voters under best comparison can be derived from Theorem 4.2.

Corollary 4.8 *Let $n = 2$ and $2 \leq k < m$. Let $p \in P$ and consider best comparison.*

(a) *If $k \leq m/2$ then φ_k is not manipulable at $p \in P$ if and only if either*

$$\beta(p(1)) \in \beta_k(p(2)) \text{ and } \beta(p(2)) \in \beta_k(p(1))$$

or

$$\beta_k(p(1)) \cap \beta_k(p(2)) = \emptyset\ .$$

(b) *If $k > m/2$ then φ_k is not manipulable at any $p \in P$.*

Proof. If $k > m/2$ then case (c) in Theorem 4.2 applies to all $p \in P$, and if $k \leq m/2$ then case (c) applies to no $p \in P$. This implies part (b) of the corollary, and it also implies that for $k \leq m/2$ we only have to consider cases (a) and (b) in Theorem 4.2. It is easily seen that these cases result in the two cases in part (a) of the corollary.$\square$

The number of non-manipulable profiles $\eta(m,k)$ if $k \leq m/2$ is computed in the following theorem.

Theorem 4.9 *Let $n = 2$ and $2 \leq k \leq m/2$. Consider best comparison. Then*

$$\eta(m,k) = m!\,(m-2)!\,(k-1)^2 + m!\,(m-1)! + m!\,[(m-k)!]^2/(m-2k)!\ .$$

Proof. The first case in (a) in Corollary 4.8 with $\beta(p(1)) \neq \beta(p(2))$ results in

$$m!\,(k-1)\binom{m-2}{k-2}(k-1)!\,(m-k)!$$

different non-manipulable profiles. This yields the first term of $\eta(m,k)$ in the theorem. If $\beta(p(1)) = \beta(p(2))$ then this number is simply equal to $m!\,(m-1)!$, which yields the second term. The second case in (a) in Corollary 4.8 results in

$$m!\binom{m-k}{k}k!\,(m-k)!$$

different non-manipulable profiles, which simplifies to the third term for $\eta(m,k)$ in the theorem. $\square$

If we require $k \neq 1$ and citizen-sovereignty, i.e., $k \leq m/2$, then the optimal value of k with respect to non-manipulability, i.e., the value of k that maximizes $\eta(m,k)$, is equal to 2.

To see this, note that by Theorem 4.9 and some elementary calculations we have for $2 < k \leq \frac{m}{2}$:

$$\begin{aligned} &\eta(m,2) > \eta(m,k) \\ &\Leftrightarrow (m-2)(m-3) > k(k-2) + \frac{\overbrace{(m-k)(m-k-1)\cdot\ldots\cdot(m-2k+1)}^{k \text{ factors}}}{\underbrace{(m-2)(m-3)\cdot\ldots\cdot(m-k+1)}_{k-2 \text{ factors}}}\ . \end{aligned}$$

Since $k > 2$ it is therefore sufficient to prove that

$$(m-2)(m-3) > k(k-2) + (m-2k+2)(m-2k+1) \ .$$

This simplifies to $(4k-8)m > 5k^2 - 8k - 4$. Since $m \geq 2k$, it is sufficient to show that $3k^2 - 8k + 4 > 0$, which indeed holds for $k > 2$.

4.4.3. *Stochastic dominance comparison for two voters.* The non-manipulable profiles for two voters under stochastic dominance comparison can be derived from Theorem 4.3.

Corollary 4.10 *Let $n = 2$ and $2 \leq k < m$. Let $p \in P$ and consider stochastic dominance comparison. Then φ_k is not manipulable at p if and only if at least one of the following holds.*

(*a*) $\beta_k(p(1)) \cap \beta_k(p(2)) = \emptyset$.

(*b*) $|\beta_k(p(1)) \cap \beta_k(p(2))| = 1$.

(*c*) $\beta_k(p(1)) \cap \beta_k(p(2)) \neq \emptyset$ *and* $[A \setminus \beta_k(p(1))] \cap [A \setminus \beta_k(p(2))] = \emptyset$.

Proof. For $n = 2$ and $k \geq 2$ cases (d) and (e) in Theorem 4.3 are not possible. Case (c) in Theorem 4.3 reduces to case (b) above, and case (a) in Theorem 4.3 reduces to case (c) above. Finally, case (b) in the theorem reduces to case (a) above. □

From this description we can again derive the number of manipulable profiles $\eta(m, k)$.

Theorem 4.11 *Let $n = 2$, $k \geq 2$, and consider stochastic dominance comparison.*

(*a*) *If $k \leq m/2$ then*

$$\eta(m,k) = m!\,[(m-k)!]^2/(m-2k)! + m!\,k^2\,[(m-k)!]^2/(m-2k+1)! \ .$$

(*b*) *If $k > m/2$ then*

$$\eta(m,k) = m!\,[k!]^2/(2k-m)! \ .$$

Proof. If $k \le m/2$ then (c) in Corollary 4.10 is not possible, and cases (a) and (b) in the corollary are mutually exclusive. In case (a) of Corollary 4.10 there are

$$m! \binom{m-k}{k} k!\,(m-k)!$$

non-manipulable profiles, resulting in the first term for $\eta(m,k)$, and in case (b) of the corollary there are

$$m!\,k \binom{m-k}{k-1} k!\,(m-k)!$$

non-manipulable profiles, resulting in the second term for $\eta(m,k)$.

If $k > m/2$ then case (a) of Corollary 4.10 is not possible, and (b) is a special case of (c). For the latter case, we just have to count the number of profiles for which $[A \setminus \beta_k(p(1))] \cap [A \setminus \beta_k(p(2))] = \emptyset$, since the other condition is always fulfilled. This number is equal to

$$m! \binom{k}{m-k} k!\,(m-k)!$$

which is equal to $m!\,[k!]^2/(2k-m)!$. □

About the value of k that maximizes $\eta(m,k)$, so the value of k that is optimal with respect to non-manipulability, we can say the following.

1. For $2 \le k \le \frac{m}{2}$, the number of non-manipulable profiles decreases with k, and thus $k = 2$ is optimal.

2. For $\frac{m}{2} < k \le m-1$, the number of non-manipulable profiles increases with k, and thus $k = m-1$ is optimal.

3. $\eta(m,2) > \eta(m,m-1)$ for $m \ge 4$, so $k = 2$ is the overall optimal value between 2 and $m-1$.

To prove these statements, first assume $k \le \frac{m}{2}$. Then, using Theorem 4.11(a) and simplifying, we derive

$$\eta(m,k+1) < \eta(m,k) \Leftrightarrow 3k^2 - 2km - 1 < 0 \ ,$$

and it is easily seen that the right hand side holds for all $2 \le k \le \frac{m}{2}$. Next, assume $\frac{m}{2} < k \le m-1$. Then, using Theorem 4.11(b) and simplifying, we derive

$$\eta(m, k+1) > \eta(m, k) \Leftrightarrow 3k^2 + k(4-4m) + m^2 - 3m + 1 < 0 \ .$$

The roots of the quadratic expression in k at the right hand side are $\frac{2}{3}(m-1) \pm \frac{1}{3}\sqrt{m^2+m+1}$; the smaller root is smaller than $\frac{m}{2}$, whereas the larger root is larger than $m-1$. Thus, the right hand side holds for all $\frac{m}{2} < k \le m-1$. Finally, by Theorem 4.11 again,

$$\eta(m, 2) > \eta(m, m-1) \Leftrightarrow m > 3 \ ,$$

so that $k = 2$ is the overall optimal value of k for $2 \le k \le m-1$.

4.5. *Lexicographic worst and best comparison*

In this subsection we briefly consider a natural extension of worst and best comparison, namely lexicographic worst and best comparison. These preference extensions to sets are given by the following recursive definition. For two subsets B and C of alternatives, we say that B is (weakly) *preferred to C under lexicographic worst comparison* by voter i with preference $p(i)$ if

1. $C = \emptyset$, or
2. B and C are non-empty and $\big(\omega(p(i)_{|B}), \omega(p(i)_{|C})\big) \in p(i)$, or
3. $\omega(p(i)_{|B}) = \omega(p(i)_{|C}) =: w$ and $B \setminus \{w\}$ is preferred to $C \setminus \{w\}$ under lexicographic worst comparison by voter i with preference $p(i)$.

The definition for lexicographic best comparison is obtained simply by replacing the worst alternative by the best alternative, i.e., by replacing $\omega(\cdot)$ by $\beta(\cdot)$. Thus, under lexicographic worst comparison a voter first considers the worst elements of B and C. If these are different, then he prefers the set with the better worst element. Otherwise, the voter considers the second worst elements. If these are different, then he prefers the set with the better second worst element. Otherwise, he considers the third worst elements, etc. Similarly, of course, for lexicographic best comparison.

Complete characterizations of the non-manipulable profiles for both lexicographic worst and lexicographic best comparison can be given but are rather technical (even more so than for stochastic dominance comparison) and therefore not included.

Note that any profile that is manipulable under worst [best] comparison is also manipulable under lexicographic worst [best] comparison. Hence, the set of non-manipulable profiles under lexicographic worst [best] comparison is always a subset of the set of non-manipulable profiles under worst [best] comparison. It is not very difficult to check (we omit the proof for the sake of briefness) that all the profiles listed in Corollary 2.3, that is, all two-voter profiles that are non-manipulable under worst comparison, are also non-manipulable under lexicographic worst comparison, so that in this case considering lexicographic worst comparison instead of just worst comparison does not make any difference. The non-manipulable profiles coincide, and the optimal value of k as far as non-manipulability is concerned, is the same as in Section 4.4.1.

For two voters and lexicographic best comparison the situation is different and the set of non-manipulable profiles is a strict subset of the set of non-manipulable profiles under best comparison, that is, the set of profiles described in Corollary 4.8. To be precise, we have the following result, which can be derived from Corollary 4.8 (the proof is again left to the reader).

Corollary 4.12 *Let $n = 2$ and $2 \leq k < m$. Let $p \in P$ and consider lexicographic best comparison.*

(*a*) *If $k \leq m/2$ then φ_k is not manipulable at $p \in P$ if and only if either*

$$\{\beta(p(1))\} = \{\beta(p(2))\} = \beta_k(p(1)) \cap \beta_k(p(1))$$

or

$$\beta_k(p(1)) \cap \beta_k(p(2)) = \emptyset\ .$$

(*b*) *If* $k > m/2$ *then* φ_k *is not manipulable at any* $p \in P$.

In this case, the total number of non-manipulable profiles for $2 \leq k \leq m/2$ is equal to

$$\eta(m,k) = \frac{m!\,[(m-k)!]^2\,(m-2k+2)}{(m-2k+1)!}$$

and this number is decreasing in k, so that $k = 2$ is the value of k that minimizes manipulability subject to $2 \leq k \leq m/2$, just as in the best comparison case. The proofs of these facts are somewhat simpler than for the best comparison case. For the sake of briefness we omit them.

4.6. *An asymptotic result*

We start with defining the class of all scoring rules. A (normalized) *scoring vector* is a vector $s = (s_1, s_2, \ldots, s_m) \in \mathbb{R}^m$ with $1 = s_1 \geq s_2 \geq \ldots \geq s_m = 0$. For a preference $\pi \in P$ and an alternative $x \in A$ let $t(\pi, x)$ denote the *rank* of x in the preference π, i.e., $t(\pi, x) = k$ where $k = 1$ if $\{x\} = \beta(\pi)$ and $\{x\} = \beta_k(\pi) \setminus \beta_{k-1}(\pi)$ otherwise.

For a scoring vector s, a profile $p \in P^N$, and an alternative $x \in A$, we denote by the *s-score*

$$\mathrm{score}_s(x,p) = \sum_{i \in N} s_{t(p(i),x)}$$

the total score that x obtains under profile p and score vector s. The *scoring rule with scoring vector* s is defined by

$$\varphi_s(p) = \{x \in A \mid \mathrm{score}_s(x,p) \geq \mathrm{score}_s(y,p) \text{ for all } y \in A\}, \quad p \in P^N .$$

Clearly, a k-approval rule is a scoring rule with scoring vector s such that $s_1 = \ldots = s_k = 1$ and $s_{k+1} = \ldots = s_m = 0$.

Now, in what follows, we fix the number of alternatives m and let the number of voters go to infinity. We will show, formally, that then any two scoring rules lead to the same expected values of the highest score, second highest score, and so on, up to a multiplicative constant proportional to the standard deviations of the scoring vectors: the higher this standard deviation the larger the differences between the expected scores. Since the standard deviation is maximal for k-approval rules with k around $m/2$, we can conclude by the law of large numbers that the proportion of manipulable profiles is smallest for this rule.[7]

In order to derive the announced result, assume that voter preferences are drawn from the uniform distribution over P – that is, according to 'impartial culture'. Let $Y = (Y_\pi)_{\pi \in P}$ denote the random vector giving the numbers of voters for each preference (so $\sum_{\pi \in P} Y_\pi = n$). Then Y has a multinomial distribution with mean $(n/m!)\mathbf{1}$, where $\mathbf{1}$ is a vector with all entries equal to 1. Write $A = \{x_1, \ldots, x_m\}$, then for a scoring vector s the random vector Y gives rise to a random vector of scores $X^s = (X_1^s, \ldots, X_m^s)$ where $X_j^s = \sum_{\pi \in P} Y_\pi s_{t(\pi, x_j)}$ for $j = 1, \ldots, m$. Let

$$\sigma(s) = \sqrt{\frac{s_1^2 + \ldots + s_m^2}{m} - \bar{s}^2}$$

denote the standard deviation of the scoring vector s, where $\bar{s} = (s_1 + \ldots + s_m)/m$ is the mean of s. Proposition 2 in Pritchard and Wilson (2009) asserts that $(X^s - n\bar{s}\mathbf{1})/\sqrt{n}$ converges in distribution to $Z^s := \sigma(s)\left(\frac{m}{m-1}\right)^{1/2}(Z - \bar{Z}\mathbf{1})$, where $Z = (Z_1, \ldots, Z_m)$ is a vector of independent standard normal random variables and $\bar{Z} = (1/m)\sum_{j=1}^m Z_j$. In words, this means that the limit distributions of the vectors of normalized random variables X^s differ only in a multiplicative constant, namely the standard deviation $\sigma(s)$. This implies $Z^s = (\sigma(s)/\sigma(s'))Z^{s'}$ for any two scoring vectors s and s'. In particular, this also holds for the associated order statistics $(Z_{(1)}^s, \ldots, Z_{(m)}^s)$ and $(Z_{(1)}^{s'}, \ldots, Z_{(m)}^{s'})$, with $Z_{(1)}^s$ and $Z_{(1)}^{s'}$ being the (limit) distributions of the highest scores. As a consequence we obtain the following proposition.

Proposition 4.13 *For all scoring vectors s and s',*

$$E[Z_{(j)}^s] - E[Z_{(j+1)}^s] = (\sigma(s)/\sigma(s'))\left(E[Z_{(j)}^{s'}] - E[Z_{(j+1)}^{s'}]\right)$$

for all $j = 1, \ldots, m-1$, where E denotes the expectation operator.

[7]We thank Eric Beutner (Maastricht University) for helpful discussions on this topic.

Proposition 2.1 implies that the difference in expected value between any two consecutive scores is largest for rules based on scoring vectors with maximal standard deviation. Since the vectors of random variables Z^s have the same distributions up to these standard deviations of the score vectors, and taking into account that the probability of all alternatives having distinct scores converges to 1 if the number of voters goes to infinity[8], we have by the law of large numbers that scoring rules φ_s with maximal standard deviation $\sigma(s)$ have the smallest proportion of manipulable profiles. The following result, of which for completeness a proof is given in the Appendix, then implies that the k-approval rule with k around $m/2$ is least manipulable if the number of voters becomes large.

Proposition 4.14 *Among all scoring vectors s, $\sigma(s)$ is maximal if and only if $s_1 = \ldots = s_k = 1$ and $s_{k+1} = \ldots = s_m = 0$, where $k = m/2$ if m is even and $k \in \{(m-1)/2, (m+1)/2\}$ if m is odd.*

For ease of reference we formulate the main result of this subsection as a corollary.

Corollary 4.15 *Let k^* denote the value(s) of k in Proposition 2.2 and let s be an arbitrary scoring vector unequal to the scoring vector associated with k^*. Then for n sufficiently large the proportion of manipulable profiles under φ_{k^*} is smaller than the proportion of manipulable profiles under φ_s.*

This asymptotic result should be taken with some care, since the probability of being able to manipulate becomes very small if the number of voters grows, and so we are comparing small numbers. On the other hand there is some evidence that already for a relatively small number of voters the k-approval rule with k close to $m/2$ performs best, at least among the k-approval rules. See Table 3 in the next section.

[8]This is Proposition 3 in Pritchard and Wilson, 2009.

5. Some simulation results

Since general comparisons between the approval rule and k-approval rules are complex and hard to obtain, we present here some results of simulations.[9]

Table 1 gives the approximate percentages of non-manipulable profiles for the approval rule with 3–10 alternatives and 2, 3, 6, and 10 voters, based on 1,00,000 trials. While the number of trials is relatively low, we nevertheless think that the numbers in the table give reliable impressions.

	m	3	4	5	6	7	8	9	10
$n = 2$	worst comp	56	45	38	33	30	27	25	23
	best comp	65	59	55	52	49	48	46	45
	stoch comp	41	37	34	31	30	28	27	25
$n = 3$	worst comp	57	49	44	41	38	36	34	32
	best comp	56	47	41	38	34	32	30	28
	stoch comp	25	20	17	16	15	14	14	13
$n = 6$	worst comp	67	61	57	54	52	50	48	47
	best comp	51	40	33	29	25	22	20	18
	stoch comp	28	20	16	13	11	10	09	08
$n = 10$	worst comp	73	69	65	62	60	59	57	56
	best comp	54	43	37	32	28	25	23	21
	stoch comp	35	27	22	19	16	15	13	12

Table 1: Approximate percentages of non-manipulable preference profiles for the approval rule, based on 1,00,000 trials.

Some conclusions can be drawn from this table. Clearly, the possibility of manipulation increases with the number of alternatives. For more than two voters ma-

[9] We thank Bram Driesen for doing these simulations (with Matlab). They are based on randomly drawing profiles and checking for non-manipulability using the characterizations in Sections 3 and 4.

nipulability also increases from worst comparison to best comparison and from best comparison to stochastic dominance comparison. This is not entirely intuitive at first glance. One might expect that many profiles that are manipulable under stochastic comparison are also manipulable under worst and best comparison, since in order to improve under stochastic comparison a necessary condition is that the worst and best alternatives of a set should not decrease in preference. Thus, to explain the results in Table 1, it seems to be the case that manipulation under stochastic comparison is often performed by improving intermediate alternatives. Moreover, apparently this kind of manipulation can often lead to comparable sets, in spite of the fact that stochastic dominance preference is not complete.

As a final comment on Table 1, under worst comparison manipulability seems to decrease with the number of voters, but under the other two preference extensions manipulability first seems to increase and then to decrease again.

In Table 2 we present the results for k-approval rules for 6 alternatives; 2, 3, 6, or 10 voters; and again based on 1,00,000 trials. For comparison the corresponding results for the approval rule from Table 1 are copied in Table 2.

$(m = 6)$	k	1	2	3	4	5	Approval rule
$n = 2$	worst comp	100	54	45	40	83	33
	best comp	100	60	35	100	100	52
	stoch comp	100	93	50	40	83	31
$n = 3$	worst comp	44	57	61	32	56	41
	best comp	100	63	51	48	100	38
	stoch comp	100	30	54	35	56	16
$n = 6$	worst comp	60	64	60	51	25	54
	best comp	50	59	50	46	78	29
	stoch comp	34	41	24	16	16	13
$n = 10$	worst comp	69	70	68	64	56	62
	best comp	45	56	52	40	51	32
	stoch comp	29	36	31	27	07	19

Table 2: Approximate percentages of non-manipulable preference profiles for k-approval rules and the approval rule, $m = 6$, based on 1,00,000 trials. (The percentages equal to 100 are exact and reflect strategy-proofness in the involved cases.)

Again we see that for relatively high numbers of voters manipulability seems to increase from worst to best and from best to stochastic dominance comparison. Further, for more than two voters the approval rule is outperformed by (at least) the 3-approval rule as far as non-manipulability is concerned.

The final simulation results we present are collected in Table 3. This table gives the percentages of non-manipulable preference profiles for k-approval rules with k odd and the approval rule for 10 alternatives, 25 agents, based on 1,000,000 trials. We give more accurate numbers than in the other tables since some differences are very small. A prudent observation is that the 5-approval rule performs best with respect to (non-)manipulability among the odd values of k (except for the case $k = 9$

and best comparison) – in line with the asymptotic result in Section 4.6. Also, it performs better than the approval rule.

$m = 10$	k	1	3	5	7	9	Approval rule
$n = 25$	worst comp	67.1	74.4	74.5	70.2	48.8	69.5
	best comp	33.5	49.9	50.2	40.9	54.8	37.6
	stoch comp	26.6	37.6	37.8	29.7	03.7	29.1

Table 3: Percentages of non-manipulable preference profiles for k-approval rules and the approval rule, $m = 10$, $n = 25$, based on 1,000,000 trials.

6. Concluding remarks

We have characterized all (non-)manipulable preference profiles for the approval rule and for k-approval rules. Our simulation results indicate that k-approval voting may be a good substitute for approval voting. It preserves the simplicity of the voting procedure characteristic for approval voting but tends to be less manipulable. Asymptotically, the k-approval rule with k close to half of the number of alternatives is even best among all scoring rules in this respect. This result is in line with Pritchard and Wilson (2009), although their context is somewhat different.

Our characterizations are also useful for studying 'partial culture' voting, where uniform distribution of preference profiles is not assumed.

CHAPTER 3

Characterization of probabilistic rules on single peaked domain

1. Introduction

In modern age the study of probabilistic schemes gained a lot of interest in the literature. However, in the classical model of social choice with all preferences admissible, only strategy-proof rule is a so-called random dictatorship where each agent has an equal chance of being the dictator. Gibbard (1973), Gibbard (1977a) first studied strategy-proof probabilistic schemes and characterized random dictatorships as the only strategy-proof and unanimous rule. This arises the natural question as to whether in a more restricted context probabilistic schemes other than random dictatorships can be strategy-proof. For the deterministic setting the impossibility (i.e., dictatorship) results of Gibbard (1973) and Satterthwaite (1975) can be avoided by restricting the set of preferences and at the same time adapting the domain of alternatives. In particular, Moulin (1980) characterizes classes of schemes on the real line that are non-dictatorial and strategy-proof with respect to single-peaked preferences. These results have later been extended in several directions by many authors. In this chapter, we adopt the Moulin framework but consider probabilistic rules. Such a rule assigns to every profile of reported individual preferences a probability distribution over the real line. The main property that we impose is strategy-proofness. In order to formulate this condition the preferences of the agents must be extended to probability distributions. This will be done as follows. For a given single-peaked preference an agent (weakly) prefers one probability distribution over another if the former assigns at least as much probability to any upper contour set of the preference as the latter. Here, an upper contour set is an interval around the peak of the preference consisting of those points that are weakly preferred to a given outcome. A probabilistic rule is strategy-proof if honest reporting always results in a probability distribution that (weakly) dominates, in the sense just described, any probability distribution brought about by lying. Some references from the literature on probabilistic social choice mechanisms are Barbera et al. (1998), Dutta (1980), and Bandyopadhyay et al. (1982) - but this list is not exhaustive. Recently, probabilistic mechanisms in private goods contexts have been studied by Sasaki (1997), Abdulkadiroğlu and Sònmez (1998), Crès and Moulin (2001), and Bogomolnaia and Moulin (2001).

2. Model and main results

Let $N = \{1, \ldots, n\}$, $n \geq 2$, denote the set of *agents*, who collectively have to choose an element from a set A of *alternatives*. In this paper A is either the interval $[0, 1]$ or a finite subset of it containing both 0 and 1. A *single-peaked preference* of agent i on A is a complete, reflexive and transitive binary relation R_i on A for which there is a number $p(R_i) \in A$, the *peak* of R_i, such that for all $x, y \in A$: if $x < y \leq p(R_i)$ or $x > y \geq p(R_i)$ then yP_ix, where P_i denotes the asymmetric part of R_i. Let $\mathfrak{R}$ denote the set of all single-peaked preferences on A. Then $\mathfrak{R}^N$ is the set of all single-peaked preference profiles. For $R \in \mathfrak{R}^N$, $p(R) = (p(R_1), \ldots, p(R_n))$ is the vector of peaks, $\underline{p}(R) = \min\{p(R_i) \mid i \in N\}$, $\overline{p}(R) = \max\{p(R_i) \mid i \in N\}$. For $S \subseteq N$, $R_S = (R_i)_{i \in S}$ is the restriction of the profile R to S. For $i \in N$, profiles $R, R' \in \mathfrak{R}^N$ are *i-deviations* if $R_{N \setminus \{i\}} = R'_{N \setminus \{i\}}$.

A *deterministic rule* is a function $\varphi : \mathfrak{R}^N \to A$. A *probabilistic rule* is a function Φ that assigns to each profile $R \in \mathfrak{R}^N$ a (probability) distribution over A, i.e., a probability measure on the Borel σ-algebra $\mathcal{B}$ of subsets of A. (If A is finite then this is the set of all subsets of A.)

A deterministic rule can be seen as a probabilistic rule that selects for each $R \in \mathfrak{R}^N$ a distribution placing probability 1 on a single alternative.

For $x \in A$ and $R_i \in \mathfrak{R}$, the *upper contour set* of R_i at x is the set $B(x, R_i) = \{y \in A \mid yR_ix\}$. Single-peakedness of R_i implies that upper contour sets are closed intervals.

Preferences on A are extended to distributions on A as follows. For $R_i \in \mathfrak{R}$ and two distributions Q, Q' over A, Q is (weakly) preferred to Q' under R_i if Q assigns to each upper contour set of R_i at least the probability that is assigned by Q' to this set. Abusing notation we use the same symbols to denote preferences over distributions and preferences over alternatives. Formally we have:

Ordinal extension of preferences For all $R_i \in \mathfrak{R}$ and all distributions Q, Q' over A, QR_iQ' if and only if

$$\text{for all } x \in A,\ Q(B(x, R_i)) \geq Q'(B(x, R_i)). \tag{3.1}$$

Furthermore, QP_iQ' if and only if

$$QR_iQ' \text{ and for some } y \in A,\ Q(B(y, R_i)) > Q'(B(y, R_i)). \tag{3.2}$$

Inequality (5.1) says that Q stochastically dominates Q', given the ordering R_i on A. Clearly, this extension is not complete over the set of all distributions over A. Note, however, that for preferences over distributions completeness is a demanding requirement.

We now define the main properties of interest in this paper for a probabilistic rule Φ. Clearly, these properties extend to deterministic rules in a straightforward manner.

Strategy-proofness $\Phi(R) R_i \Phi\left(\overline{R}\right)$ for all $i \in N$ and all i-deviations $R, \overline{R} \in \mathfrak{R}^N$.

Strategy-proofness says that reporting a different preference results in a stochastically (weakly) dominated distribution. It implies that for any von Neumann-Morgenstern utility function representing an agent's preference relation, his expected utility is maximal when he reports his true preference relation.

Unanimity $\Phi(R)(\{p(R_1)\}) = 1$ for every $R \in \mathfrak{R}^N$ with $R_i = R_j$ for all $i, j \in N$.

For every $R \in \mathfrak{R}^N$ and every permutation π of N let R_π denote the profile $\left(R_{\pi(i)}\right)_{i \in N}$.

Anonymity $\Phi(R) = \Phi(R_\pi)$ for every $R \in \mathfrak{R}^N$ and every permutation π of N.

Peaks-onliness $\Phi(R) = \Phi\left(\overline{R}\right)$ for all $R, \overline{R} \in \mathfrak{R}^N$ with $p(R) = p(\overline{R})$.

Uncompromisingness $\Phi(R)(X) = \Phi\left(\overline{R}\right)(X)$ for all $X \in \mathcal{B}$, all $i \in N$ and all i-deviations $R, \overline{R} \in \mathfrak{R}^N$ such that $X \cap \left[p(R_i), p\left(\overline{R}_i\right)\right] = \emptyset$.

The following proposition summarizes some results from Ehlers et al. (2002). In that paper the set of alternatives is the real line but it is not difficult to adapt the results to our framework where A is either $[0,1]$ or a finite subset of it containing both 0 and 1.

Proposition 2.1 *Let Φ be a probabilistic rule. Then Φ is strategy-proof and peaks-only if and only if it is uncompromising. If Φ is unanimous and strategy-proof then it is peaks-only.*

In Ehlers et al. (2002) all uncompromising probabilistic rules are characterized. For $R \in \mathfrak{R}^N$, let $p_1(R), \ldots, p_k(R)$, $k \leq n$, denote the different peaks of R such that for all $\ell \in \{1, \ldots, k-1\}$, $p_\ell(R) < p_{\ell+1}(R)$. Thus, $p_1(R) = \underline{p}(R)$, $p_k(R) = \overline{p}(R)$, and $\{p_\ell(R) \mid \ell \in \{1, \ldots, k\}\} = \{p(R_i) \mid i \in N\}$. Let S_ℓ denote the set of agents whose peaks are smaller than or equal to $p_\ell(R)$. Thus, $S_1 \subsetneq S_2 \subsetneq \ldots \subsetneq S_k$ and $S_k = N$. Let $S_0 = \emptyset$.

For every $S \in 2^N$, let D_S be a probability distribution over A. We call $\Delta = (D_S)_{S \in 2^N}$ a collection of *fixed probabilistic ballots* if the following holds:

$$D_T([0,x]) \geq D_S([0,x[) \text{ for all } S \subseteq T \subseteq N \text{ and } x \in A\,. \tag{3.3}$$

For $X \subseteq A$ denote by 1_X the indicator function of X, i.e., $1_X(y) = 1$ if $y \in X$ and $1_X(y) = 0$ if $y \notin X$, for all $y \in A$.

Fixed probabilistic ballots rule The probabilistic rule Φ is a *fixed probabilistic ballots rule* if there is a collection $\Delta = (D_S)_{S \in 2^N}$ of fixed probabilistic ballots such that, for all $R \in \mathfrak{R}^N$ and all $X \in \mathcal{B}$

$$\begin{aligned} \Phi(R)(X) \;=\;& D_\emptyset\left(X \cap [0, \underline{p}(R)[\right) + D_N\left(X \cap]\overline{p}(R), 1]\right) \\ &+ \sum_{\ell=1}^{k-1} D_{S_\ell}\left(X \cap]p_\ell(R), p_{\ell+1}(R)[\right) \\ &+ \sum_{\ell=1}^{k} 1_X(p_\ell(R)) \left(D_{S_\ell}([0, p_\ell(R)]) - D_{S_{\ell-1}}([0, p_\ell(R)[)\right)\,. \end{aligned} \tag{3.4}$$

In this case we denote Φ by Φ^{Δ}.

This definition says that on the interval $[0, \underline{p}(R)[= [0, p_1(R)[$ the distribution $\Phi(R)$ coincides with $D_{S_0} = D_{\emptyset}$; on $]p_\ell(R), p_{\ell+1}(R)[$ it coincides with D_{S_ℓ}, for $1 \leq \ell \leq k-1$; and on $]\overline{p}(R), 1] =]p_k(R), 1]$ it coincides with $D_{S_k} = D_N$. To $p_\ell(R)$ it assigns the probability $D_{S_\ell}([0, p_\ell]) - D_{S_{\ell-1}}([0, p_\ell[)$ for $1 \leq \ell \leq k$; these numbers are nonnegative by (3.3). It is straightforward to check that $\Phi(R)(A) = 1$, so that $\Phi(R)$ is indeed a probability measure.

Also the following result is proved in Ehlers et al. (2002).

Proposition 2.2 *The probabilistic rule Φ is uncompromising if and only if $\Phi = \Phi^{\Delta}$ for some collection of fixed probabilistic ballots $\Delta = (D_S)_{S \in 2^N}$. In that case,*

- *Δ is uniquely determined;*
- *if Φ is anonymous, then $D_S = D_T$ whenever $|S| = |T|$;*
- *if Φ is unanimous, then $D_{\emptyset}(\{1\}) = D_N(\{0\}) = 1$.*

In the special case that $\Phi^{\Delta}(R)$ is degenerate for all $R \in \mathfrak{R}^N$, i.e., puts probability 1 on exactly one alternative, clearly all distributions D_S must be degenerate distributions. If the associated alternatives are denoted by a_S, $S \subseteq N$, then (3.3) implies $0 \leq a_T \leq a_S \leq 1$ for all $S \subseteq T \subseteq N$, and for $R \in \mathfrak{R}^N$ equation (5.3) implies $\Phi(R)(\{\varphi^{\Phi}(R)\}) = 1$ with

$$\begin{aligned} \varphi^{\Phi}(R) &= a_{\emptyset} 1_{[0, p_1(R)[}(a_{\emptyset}) + a_N 1_{]p_k(R), 1]}(a_N) + \sum_{\ell=1}^{k-1} a_{S_\ell} 1_{]p_\ell(R), p_{\ell+1}(R)[}(a_{S_\ell}) \\ &\quad + \sum_{\ell=1}^{k} p_\ell(R) 1_{[a_{S_\ell}, a_{S_{\ell-1}}]}(p_\ell(R)) \, . \end{aligned} \tag{3.5}$$

Thus, φ^{Φ} is the deterministic rule associated with Φ. More generally we have the following straightforward corollary. See also Moulin (1980) and Ching (1997).

Corollary 2.3 *Let φ be a deterministic rule. Then φ is uncompromising if and only if there are $a_S \in A$, $S \subseteq N$, satisfying $0 \leq a_T \leq a_S \leq 1$ for all $S \subseteq T \subseteq N$, such that $\varphi(R)$ is equal to the right-hand side of (3.5) for all $R \in \mathfrak{R}^N$. In that case,*

- *the numbers a_S, $S \subseteq N$, are uniquely determined;*
- *if φ is anonymous, then $a_S = a_T$ whenever $|S| = |T|$;*
- *if φ is unanimous, then $a_\emptyset = 1$ and $a_N = 0$.*

3. The finite case

In this section we consider the case where A is a finite subset of $[0, 1]$ including both 0 and 1. Specifically, let $A = \{x_1, \ldots, x_m\}$ with $0 = x_1 < x_2 < \ldots < x_m = 1$, where $m \geq 2$. We consider uncompromising probabilistic rules Φ and for ease of presentation assume anonymity. By Proposition 2.2, such a rule is characterized by fixed probabilistic ballots $D_0, \ldots, D_n$, where we write D_ℓ for the distributions D_L with $|L| = \ell$. For $\ell = 0, \ldots, n$ let $R^\ell \in \mathfrak{R}^N$ be a profile with $|\{i \in N \mid p_i(R^\ell) = 0\}| = \ell$ and $|\{i \in N \mid p_i(R^\ell) = 1\}| = n - \ell$. We use R^ℓ as the generic notation for such a boundary profile, which is without loss of generality as long as Φ is anonymous.

The following lemma follows easily from (5.3).

Lemma 3.1 *Let Φ be uncompromising and anonymous, with fixed probabilistic ballots $D_0, \ldots, D_n$. Then $\Phi(R^\ell)(x_j) = D_\ell(x_j)$ for all $\ell = 0, \ldots, n$ and $j = 1, \ldots, m$.*

The relevance of the boundary profiles R^ℓ stems from the following observation.

Lemma 3.2 *Let Φ^1 and Φ^2 be uncompromising and anonymous, and suppose that $\Phi^1(R^\ell) = \Phi^2(R^\ell)$ for all $\ell = 0, \ldots, n$. Then $\Phi^1 = \Phi^2$.*

Proof. Let $(D^1_\ell)_{\ell=0,\ldots,n}$ and $(D^2_\ell)_{\ell=0,\ldots,n}$ be the associated fixed probabilistic ballots. By Lemma 3.1 we have $D^1_\ell(x_j) = D^1_\ell(x_j)$ for all $\ell = 0, \ldots, n$ and all $j = 1, \ldots, m$. Hence $D^1_\ell = D^1_\ell$ for all $\ell = 0, \ldots, n$, so $\Phi^1 = \Phi^2$. □

By Corollary 2.3, any uncompromising and anonymous deterministic rule φ is characterized by a vector $\alpha = (\alpha_0, \ldots, \alpha_n) \in A^{n+1}$ such that $\alpha_k = a_K$ for any $K \subseteq N$ with $|K| = k$. In particular, $\alpha_0 \geq \ldots \geq \alpha_n$. We write φ_α for the probabilistic rule associated with φ. Let Ξ denote the (finite) set of all such vectors α. For each $\alpha \in \Xi$ let $\lambda_\alpha \geq 0$ such that $\sum_{\alpha\in\Xi} \lambda_\alpha = 1$. Then $\Phi = \sum_{\alpha\in\Xi} \lambda_\alpha\varphi_\alpha$ denotes the (uncompromising and anonymous) probabilistic rule such that $\Phi(R)(x_j) = \sum_{\alpha\in\Xi} \lambda_\alpha\varphi_\alpha(R)(x_j)$ for all $R \in \mathfrak{R}^N$ and $j = 1, \ldots, m$. The following theorem is the main result of this section. It says that every uncompromising and anonymous probabilistic rule is a convex combination of uncompromising and anonymous deterministic rules, where the coefficients in this convex combination are interpreted as probabilities.

Theorem 3.3 *Let $A = \{x_1, \ldots, x_m\}$ with $0 = x_1 \leq \ldots \leq x_n = 1$ and let Φ be a probabilistic rule. Then there are $\lambda_\alpha \geq 0$ with $\sum_{\alpha\in\Xi} \lambda_\alpha = 1$ such that $\Phi = \sum_{\alpha\in\Xi} \lambda_\alpha\varphi_\alpha$.*

Proof. The proof proceeds in four steps.

Step 1 Let $\Delta = D_0, \ldots, D_n$ be the fixed probabilistic ballots such that $\Phi = \Phi^\Delta$ (cf. Proposition 2.2). By Lemma 3.2 it is sufficient to show that there are $\lambda_\alpha \geq 0$ with $\sum_{\alpha\in\Xi} \lambda_\alpha = 1$ such that

$$\sum_{\alpha\in\Xi} \lambda_\alpha\varphi_\alpha(R^k)(x_j) = D_k(x_j) \text{ for all } k = 0, \ldots, n \text{ and } j = 1, \ldots, m. \tag{3.6}$$

We first note that if (3.6) holds then the condition $\sum\limits_{\alpha\in\Xi} \lambda_\alpha = 1$ is redundant, since:

$$1 = \sum_{j=1}^{m} \Phi^\Delta(R)(x_j) = \sum_{j=1}^{m}\sum_{\alpha\in\Xi} \lambda_\alpha\varphi_\alpha(R)(x_j) = \sum_{\alpha\in\Xi} \lambda_\alpha \sum_{j=1}^{m} \varphi_\alpha(R)(x_j) = \sum_{\alpha\in\Xi} \lambda_\alpha.$$

Thus, it is sufficient to show that (3.6) holds for some $\lambda_\alpha \geq 0$. We now write (3.6) in matrix form as $A\lambda = \mathbf{d}$ where

- A is an $(n+1)m \times |\Xi|$-matrix with rows indexed by pairs (k, j) and columns by $\alpha \in \Xi$, such that the entry in row (k, j) and column α is $\varphi_\alpha(R^k)(x_j)$;
- λ is a column vector of length $|\Xi|$ with λ_α at row α; and

• $\mathbf{d}$ is a column vector of length $(n+1)m$ with $D_k(x_j)$ at row (k,j).

Now, in order to show that the system of equations $A\lambda = \mathbf{d}$ has a nonnegative solution it is by Farkas' Lemma sufficient to show that $\mathbf{d}'\mathbf{y} \geq 0$ for any $\mathbf{y} \in \mathbb{R}^{(n+1)m}$ with $A'\mathbf{y} \geq \mathbf{0} \in \mathbb{R}^{|\Xi|}$. The system $A'\mathbf{y} \geq \mathbf{0}$ is equivalent to

$$\sum_{k=0}^{n}\sum_{j=1}^{m} \varphi_\alpha(R^k)(x_j) y_{k,j} \geq 0 \text{ for all } \alpha \in \Xi. \tag{3.7}$$

Consider an $\alpha \in \Xi$, and let $\varphi_\alpha(R^k)$ assign probability 1 to $x_{i_{n-k+1}} \in A$ for $k = 0, \ldots, n$. Then by strategy-proofness $1 \leq i_1 \leq \ldots \leq i_{n+1} \leq m$, and we have

$$\sum_{k=0}^{n}\sum_{j=1}^{m} \varphi_\alpha(R^k)(x_j) y_{k,j} = y_{n,i_1} + y_{n-1,i_2} + \ldots + y_{0,i_{n+1}} \ .$$

Therefore (3.7) and, thus, the system $A'\mathbf{y} \geq \mathbf{0}$ is equivalent to

$$y_{n,i_1} + y_{n-1,i_2} + \ldots + y_{0,i_{n+1}} \geq 0 \text{ for all } 1 \leq i_1 \leq \ldots \leq i_{n+1} \leq m. \tag{3.8}$$

So we have to prove that if $\mathbf{y} \in \mathbb{R}^{(n+1)m}$ satisfies (3.8) then

$$\sum_{k=0}^{n}\sum_{j=1}^{m} y_{k,j} D_k(x_j) \geq 0 \ . \tag{3.9}$$

Step 2 We will prove (3.9) by using a network formulation of the problem. In this step we collect some definitions and results on networks needed in the sequel of this proof.

A network is a directed graph $G = (V, E)$, with V the finite set of vertices and $E \subseteq V \times V$ the set of edges, and with $s, t \in V$ being the source (having only outgoing edges) and the sink (having only ingoing edges) of G, respectively. The capacity of an edge $(u, v) \in E$ is a nonnegative number $c(u, v)$: it represents the maximum amount of flow that can pass through it. The capacity of a vertex $v \in V$ is a nonnegative number $c(v)$. A flow is a mapping $f : E \to \mathbb{R}$ satisfying the following constraints:

(F1) $0 \leq f\,(u, v) \leq c\,(u, v)$ for each $(u, v) \in E$,

(F2) $\sum_{u:\,(u,v)\in E} f\,(u, v) = \sum_{u:\,(v,u)\in E} f\,(v, u)$, for each $v \in V \setminus \{s, t\}$, (F3) $\sum_{u \in V} f(u,v) \leq c(v)$ for each $v \in V \setminus \{s,t\}$.

The value of a flow f is defined by $|f| = \sum_{(s,v) \in E} f(s,v)$. It represents the amount of flow passing from the source to the sink.

A path from s to t is a sequence of vertices $s = v_1, \ldots, v_\ell = t$ such that $(v_i, v_{i+1}) \in E$ for all $i = 1, \ldots, \ell - 1$. A cut C is a subset of $V \cup E$ such that for any path from s to t there is a u on this path with $u \in C$ or $(u,v) \in C$ or $(v,u) \in C$ for some $v \in V$. If the cut C is a subset of V alone it is called a vertex cut. A vertex cut C is trivial if it contains s or t. The capacity of a cut C, denoted by $c(C)$, is the sum of the capacities of all edges and vertices in it. Hence, the value of any flow is smaller or equal to the capacity of any cut.

Generalized max-flow min-cut theorem *The maximum value of a flow is equal to the minimum capacity of a cut.*

For later reference we also record the following straightforward observation.

Lemma 3.4 *Suppose $c(u,v) = \min\{c(u), c(v)\}$ for all $(u,v) \in E$. Then for any cut C there is a vertex cut C' with $c(C') \leq c(C)$.*

Step 3 We now formulate our problem as a network problem and derive some results about this particular network. Let the set of vertices be the set

$$V = \{s,t\} \cup \{y_{k,j} \mid 0 \leq k \leq n, 1 \leq j \leq m\},$$

and let the set of edges be the set

$$\begin{aligned} E &= \{(s, y_{n,j}) \mid 1 \leq j \leq m\} \\ &\quad \cup \{(y_{k,j}, y_{k-1,j'}) \mid 1 \leq j \leq j' \leq m,\ 2 \leq k \leq n\} \\ &\quad \cup \{(y_{0,j}, t) \mid 1 \leq j \leq m\}\ . \end{aligned}$$

Let furthermore

$$c(s) = c(t) = 1;\ c(y_{k,j}) = D_k(x_j) \text{ for all } k = 0, \ldots, n,\ j = 1, \ldots, m\ ,$$

and

$$c(u,v) = \min(c(u), c(v)) \text{ for all } (u,v) \in E \, .$$

Then

$$\left\{ \left(s, y_{n,i_1}, y_{n-1,i_2}, ..., y_{n+1-l,i_l}, ..., y_{1,i_n}, y_{0,i_{n+1}}, t\right) \mid 1 \leq i_1 \leq ... \leq i_{n+1} \leq m \right\}$$

is the set of all paths from s to t. In this network the value of a maximal flow turns out to be equal to 1, as the following claim shows.

Claim 1 The maximal value of a flow is equal to 1.

In order to prove this claim it is by the generalized max-flow min-cut theorem it is sufficient to show that the minimal capacity of a cut is equal to 1. By Lemma 3.4 we only have to consider vertex cuts. If C is a trivial vertex cut then it contains s or t and therefore has capacity at least 1. Now let C be a nontrivial vertex cut. We claim that there are $n \geq k_1 \geq \ldots \geq k_m \geq 0$ such that $y_{k_i,i} \in C$ for all $i = 1, \ldots, m$. To see this first define $k_1 = \max\{k \in \{0, \ldots, n\} \mid y_{k,1} \in C\}$: this must exist since otherwise $y_{k,1} \notin C$ for all k and then the path $s, y_{n,1}, \ldots, y_{0,1}, t$ would not contain a vertex in C. Next, define $k_2 = \max\{k \in \{0, \ldots, k_1\} \mid y_{k,2} \in C\}$: again, this must exist since otherwise the path $s, y_{n,1}, \ldots, y_{k_1-1,1}, y_{k_1,2}, \ldots, y_{0,2}, t$ would not contain a vertex in C. And so on and so forth.

Now,

$$c(y_{k_1,1}) + c(y_{k_2,2}) = D_{k_1}(x_1) + D_{k_2}(x_2 2) \geq D_{k_2}(1) + D_{k_2}(2)$$

this inequality follows from the fact that $k_1 \geq k_2$ implies $D_{k_1}(x_1) \geq D_{k_2}(x_1)$ by (3.3). Suppose that

$$c(y_{k_1,1}) + c(y_{k_2,2}) + ... + c(y_{k_i,i}) \geq D_{k_i}(x_1) + \ldots + D_{k_i}(x_i) \, .$$

Then

$$\begin{aligned} & c(y_{k_1,1}) + \ldots + c(y_{k_i,i}) + c\left(y_{k_{i+1},i+1}\right) \\ \geq \; & D_{k_i}(x_1) + \ldots + D_{k_i}(x_i) + D_{k_{i+1}}(x_{i+1}) \\ \geq \; & D_{k_{i+1}}(x_1) + \ldots + D_{k_{i+1}}(x_i) + D_{k_{i+1}}(x_{i+1}) \end{aligned}$$

as $D_{k_i}(x_1) + \ldots + D_{k_i}(x_i) \geq D_{k_{i+1}}(x_1) + \ldots + D_{k_{i+1}}(x_i)$ for $k_i \geq k_{i+1}$. By induction we conclude

$$c(C) \geq c(y_{k_1,1}) + \ldots + c(y_{k_m,m}) \geq D_{k_m}(x_1) + \ldots + D_{k_m}(x_m) = 1.$$

Now consider the trivial cut $C = \{s\}$. Clearly $c(C) = 1$. So, the minimum capacity of a cut is 1. This completes the proof of Claim 1.

Claim 2 Let f be a flow with maximal value. Then $\sum_{u \in V:\ (u,v) \in E} f(u,v) = c(v)$ for any vertex $v \in V \setminus \{s,t\}$.

To prove this claim, consider a vertex $v = y_{k,j}$. Consider the cut $C = \{y_{k,j'} \mid 1 \leq j' \leq m\}$. Then, by repeated application of (F2) in the definition of a flow, we obtain

$$\begin{aligned} |f| &= \sum_{j=1}^{m} f(s, y_{n,j}) = \sum_{j=1}^{m} \sum_{\ell=j}^{m} f(y_{n,j}, y_{n-1,\ell}) = \ldots \\ &= \sum_{j=1}^{m} \sum_{\ell=j}^{m} f(y_{k-1,j}, y_{k,\ell}) = \sum_{v' \in C} \sum_{u \in V:\ (u,v') \in E} f(u,v') \leq \sum_{v' \in C} c(v') , \end{aligned}$$

where the inequality follows from (F3). Also,

$$\sum_{v' \in C} c(v') = \sum_{j'=1}^{m} c(y_{k,j'}) = \sum_{j'=1}^{m} D_k(x_{j'}) = 1 .$$

Hence, by Claim 1,

$$1 = |f| = \sum_{v' \in C} \sum_{u \in V:\ (u,v') \in E} f(u,v') \leq \sum_{v' \in C} c(v') = 1 .$$

So the inequality must be an equality, and by (F3) we have

$$\sum_{u \in V:\ (u,v') \in E} f(u,v') = c(v')$$

for all $v' \in C$, in particular for $v' = v$. This proves Claim 2.

Step 4 We return to our original problem and complete the proof of the theorem. First note that the left-hand sides of (3.8) correspond 1-1 with paths in the network by adding s at the beginning and t at the end.

Claim 3 For all $\mathbf{y} \in \mathbb{R}^{(n+1)m}$ we have

$$\mathbf{d}'\mathbf{y} = \sum_{1 \le i_1 \le \ldots \le i_{n+1} \le m} (y_{n,i_1} + \ldots + y_{0,i_{n+1}}) \cdot f(s, y_{n,i_1}, \ldots, y_{0,i_{n+1}}, t) \ , \tag{3.10}$$

where f is a maximal flow in the network.

To see this, we fix $y_{k,j}$ and show that its coefficients in the left and right-hand sides are equal. The coefficient of $y_{k,j}$ in the left-hand side is equal to $D_k(x_j)$, as follows from (3.9). The coefficient of $y_{k,j}$ in the right-hand side is equal to the total flow passing through $y_{k,j}$, which is equal to $c(y_{k,j})$ by Claim 2. In turn, $c(y_{k,j}) = D_k(x_j)$. This proves Claim 3.

The proof of the theorem is now complete by noting that if $\mathbf{y} \in \mathbb{R}^{(n+1)m}$ satisfies (3.8) then the right-hand side of (3.10) is nonnegative, hence $\mathbf{d}'\mathbf{y} \ge 0$. $\square$

CHAPTER 4

The structure of strategy-proof random choice functions over product domain and separable preferences: The case of two voters

1. Introduction

Randomization has been used as a method of resolving conflicts of interest since antiquity. It has been analyzed extensively in problems of aggregation, fairness and mechanism design in a variety of models including the pure voting model, matching, auctions and other allocation models. [1] From the perspective of mechanism design theory, allowing for randomization expands the set of incentive-compatible social choice functions because domain restrictions are inherent in the preference ranking of lotteries that satisfy the expected utility hypothesis. A classical result in this respect is that of Gibbard (1977b) which characterizes the class of strategy-proof random social choice functions over the complete domain of preferences.

In this chapter we investigate the class of strategy-proof random social choice rules over multi-dimensional (or multi-component) domains with separable preferences. This model is an important one with several applications and has been extensively studied in the deterministic setting, for example in Barberá et al. (1991), Barberà et al. (1993), LeBreton and Sen (1999), Barberà et al. (1997), Barberà et al. (2005) and Svensson and Torstensson (2008). For a survey see Sprumont (1995).

LeBreton and Sen (1999) show that strategy-proof deterministic social choice functions defined over a *rich* domain of preferences are *decomposable*, i.e. a strategy-proof social choice function is composed of strategy-proof social choice functions defined over each component domain. A corollary of this result is the following: if the domain comprises of *all* separable preferences and each component domain consists of at least three alternatives, then a social choice function is strategy-proof only if there is a *dictator* for each component.

In this chapter, we analyze the structure of random strategy-proof social choice functions over a sub-domain of separable preferences, the domain of *lexicographically*

[1] See for example, Barberá and Sonnenschein (1978), Myerson (1981), Bogomolnaia and Moulin (2001), Bogomolnaia and Moulin (2004), Bogomolnaia et al. (2005), Moulin and Stong (2002)

separable preferences or simply lexicographic preferences. If the decomposability property extended straightforwardly to random social choice functions, we would expect strategy-proof random social choice functions over product domains to be the *stochastic product* of component random dictatorships. This is false; for instance, random dictatorship itself is clearly strategy-proof but not the product of component random dictatorships. The latter would put non-zero probabilities on alternatives that are not first-ranked by any voter unlike a random dictatorship. Thus products of component random dictatorships are strategy-proof but do not describe all random strategy-proof choice functions.

Our main result is a complete characterization of random strategy-proof social choice functions in the case of two voters. The case of an arbitrary number of voters involves technical difficulties which we are unable to address at the moment. We call such random social choice functions, *generalized random dictatorships* and they include random dictatorships and products of component random dictatorships. A random dictatorship is a fixed probability distribution on the set of voters. At any preference profile, the probability of an alternative is the sum of the probability weights of voters for whom the alternative is the best. A generalized random dictatorship on the other hand, is a fixed probability distribution on the set of all *voter sequences* of length m where m is the number of components. For instance, if there are three voters and five components, there are 3^5 possible voter sequences. A generalized random dictatorship assigns a probability to each of these sequences. An alternative is consistent with a sequence at a profile if each component of the alternative is the best (amongst all component alternatives) for the voter specified in the sequence for that component. The total probability of the alternative at the profile is simply the sum of probabilities of voter sequences consistent with the alternative. A generalized random dictatorship thus induces a fixed probability distribution on the *product set* of the maximal alternatives of all voters. A critical feature of these social choice functions is that the induced *marginal* probability distribution on each component is a random dictatorship.

The chapter is organized as follows. In the next section, we introduce the model, the notation and the background results. The following section contains the main result and its proof while the final section concludes.

2. Background and Preliminaries

The set of alternatives is a finite set $A \equiv A_1 \times A_2 ... \times A_m$ where A_j, $j = 1, ...m$ is the j^{th} component set. The set of components will be written as $M = \{1, ..., m\}$. An element $a \in A$ is an m-tuple $a \equiv (a_1, ..., a_m)$. For any $Q \subset M$, we will let $A_Q = \prod_{j \in Q} A_j$. Abusing notation slightly, we will write A_{M-j} for the set $\prod_{k \neq j} A_k$. Typical elements of A_{M-j} will be denoted by a_{M-j}, b_{M-j} etc.

The set of voters is $I = \{1, \ldots, N\}$. Each voter i has an antisymmetric preference ordering P^i over the elements of A which is assumed to be *separable*.

Definition 2.1 *The ordering P^i is separable if for all $Q \subset M$, for all $a_Q, b_Q \in A_M$, for all $c_{M-Q}, d_{M-Q} \in A_{M-Q}$,*

$$[(a_Q, c_{M-Q})P^i(b_Q, c_{M-Q})] \Rightarrow [(a_Q, d_{M-Q})P^i(b_Q, d_{M-Q})].$$

If a preference ordering is separable, then choices over a subset of components do not affect ranking of alternatives over the remaining components. In other words, choices over components do not impose "externalities" over other components. A separable preference P^i induces a *marginal preference ordering* P^i_Q over A_Q, $Q \subset M$ in a natural way: for every $a_Q, b_Q \in A_Q$

$$[a_Q P^i_Q b_Q] \text{ if } [(a_Q, c_{M-Q})P^i(b_Q, c_{M-Q}) \text{ for all } c_{M-Q} \in A_{M-Q}].$$

The set of marginal preference orderings over components in the set $Q \subset M$ will be denoted by $[\mathbb{D}_Q]^S$.

A particular class of separable orderings is the class of *lexicographically separable* orderings.

Definition 2.2 *The ordering P^i is lexicographic if there exists an antisymmetric ordering $>$ on the set M and antisymmetric orderings P^i_j on each component set A_j, $j \in M$ such that, for all $a, b \in A$, aP^ib iff there exists a component j and*

1. $a_j P^i_j b_j$
2. $a_k = b_k$ *for all* $k \in M$ *such that* $k > j$.

Let P^i be a lexicographic ordering. We shall refer to the components that are maximal and minimal according to the ordering $>$ over M as the *lexicographically best* and *lexicographically worst* components respectively. In general if, components j and k are such that $k > j$, we shall say that component k is *lexicographically better* than component j.

Let $\mathbb{P}$, $\mathbb{D}^S$ and $\mathbb{D}^L$ denote respectively the set of all antisymmetric orderings, the set of separable orderings and the set of lexicographically orderings over A respectively. We note that $\mathbb{D}^L \subseteq \mathbb{D}^S \subset \mathbb{P}$. [2] We also let $\mathbb{P}_j$, $j \in M$ denote the set of all possible antisymmetric orderings over the elements of the set A_j.

Let $\mathbb{D} \subset \mathbb{P}$. A *preference profile* P is an N-tuple $(P^1, ..., P^N) \in \mathbb{D}^N$. For any voter i, ordering $\bar{P}^i$ and profile P, $(\bar{P}^i, P^{-i})$ will denote the profile where the i^{th} component of P has been replaced by $\bar{P}^i$. A *marginal preference profile* for components Q, $Q \in M$ is similarly an N-tuple, $P_Q \equiv (P^1_Q, ..., P^N_Q)$. We shall say that two profiles $P, \bar{P} \in [\mathbb{D}^S]^N$ are *marginally equivalent* if $P^i_j = \bar{P}^i_j$ for all voters $i \in I$ and $j \in M$.

We let $\mathcal{L}(A)$ denote the set of lotteries over the elements of the set A. If $\lambda \in \mathcal{L}(A)$, then λ_a will denote the probability that λ puts on $a \in A$. Clearly $\lambda_a \geq 0$ and $\sum_{a \in A} \lambda_a = 1$. For every $j \in M$ we can define $\mathcal{L}(A_j)$ accordingly.

Definition 2.3 *Let $\mathbb{D} \subset \mathbb{P}$. A Random Social Choice Function (RSCF) (for the domain $\mathbb{D}$) is a map $\varphi : \mathbb{D}^N \to \mathcal{L}(A)$.*

[2]The set of lexicographic orderings coincides with the separable orderings in the special case when there are two components and exactly two alternatives in every component set. In general $\mathbb{D}^L$ is a strict subset of $\mathbb{D}^S$.

Our focus is on RSCFs that are *strategy-proof*, i.e. which provide voters with dominant-strategy incentives to reveal their preference orderings (which are assumed to be private information), truthfully. In models such as ours where the outcome of voting is a probability distribution over outcomes, there are several ways to define strategy-proofness. Here we follow the approach of Gibbard (1977b).

Definition 2.4 *A utility function $u : A \to \Re$ represents the ordering P^i over A if for all $a, b \in A$,*

$$[aP^ib] \Leftrightarrow [u(a) > u(b)]$$

Definition 2.5 *A RSCF $\varphi : \mathbb{D}^N \to \mathcal{L}(A)$ is manipulable by voter i at profile $P \in \mathbb{D}^N$ via $\bar{P}^i \in \mathbb{D}$ if there exists a utility functions u representing P^i such that*

$$\textstyle\sum_{a\in A} u(a)\varphi_a(\bar{P}^i, P^{-i}) > \sum_{a\in A} u(a)\varphi_a(P^i, P^{-i}).$$

Definition 2.6 *A RSCF $\varphi : \mathbb{D}^N \to \mathcal{L}(A)$ is strategy-proof if it is not manipulable by any voter at any profile. Equivalently, φ is strategy-proof if, for all $i \in I$, for all $P \in \mathbb{D}^N$, for all $\bar{P}^i \in \mathbb{D}$ and all utility functions u representing P^i, we have*

$$\textstyle\sum_{a\in A} u(a)\varphi_a(P^i, P^{-i}) \geq \sum_{a\in A} u(a)\varphi_a(\bar{P}^i, P^{-i}).$$

A RSCF is strategy-proof if at every profile no voter can obtain a higher expected utility by deviating from her true preference ordering than she would if she announced her true preference ordering. Here, expected utility is computed with respect an arbitrary utility representation of her true preferences. It is well-known that this is equivalent to requiring that the probability distribution from truth-telling stochastically dominates the probability distribution from misrepresentation in terms of a voter's true preferences. This is stated formally below.

For any $i \in I$, $P^i \in \mathbb{D}$ and $a \in A$, we let $B(a, P^i) = \{b \in A : bP^ia\} \cup \{a\}$, i.e. $B(a, P^i)$ denotes the set of alternatives that are weakly preferred to a according to the ordering P^i.

Definition 2.7 *A RSCF $\varphi : \mathbb{D}^N \to \mathcal{L}(A)$ is manipulable by voter i at profile $P \in \mathbb{D}^N$ via $\bar{P}^i \in \mathbb{D}$ if there exists $a \in A$ such that*

$$\sum_{b \in B(a,P^i)} \varphi_b(\bar{P}_i, P_{-i}) > \sum_{b \in B(a,P^i)} \varphi_b(P^i, P^{-i}).$$

It is strategy-proof if for all $i \in I$, for all $P \in \mathbb{D}^N$, for all $\bar{P}_i \in \mathbb{D}$ and all $a \in A$, we have

$$\sum_{b \in B(a,P^i)} \varphi_b(P_i, P_{-i}) \geq \sum_{b \in B(a,P^i)} \varphi_b(\bar{P}^i, P^{-i}).$$

We also introduce the mild requirement of *unanimity* for RSCFs. This requires an alternative which is first-ranked by all voters in any profile to be selected with probability one in that profile. For any $P^i \in \mathbb{D}$, let $\tau(P^i, A)$ denote the maximal element in A according to P^i. Since the domain consists of antisymmetric orderings and A is finite, a maximal element always exists and is unique.

Definition 2.8 *A RSCF $\varphi : \mathbb{D}^N \to \mathcal{L}(A)$ satisfies unanimity if for all $P \in \mathbb{D}^N$ and $a \in A$,*

$$[a = \tau(P^i, A) \text{ for all } i \in I] \Rightarrow [\varphi_a(P) = 1].$$

A RSCF of particular significance is *random dictatorship.*

Definition 2.9 *The RSCF $\varphi^r : \mathbb{D}^N \to \mathcal{L}(A)$ is a random dictatorship if there exist non-negative real numbers β^i, $i \in I$ with $\sum_{i \in I} \beta^i = 1$ such that for all $P \in \mathbb{D}$ and $a \in A$,*

$$\varphi^r_a(P) = \sum_{\{i:\tau(P,A)=a\}} \beta^i$$

In a random dictatorship, each voter i gets weight β_i where the sum of these β_i's is one. At any profile, the probability assigned to an alternative a is simply the sum of the weights of the voters whose maximal element is a. A random dictatorship is clearly strategy-proof for any domain; by misrepresentation, a voter can only transfer weight from her most-preferred to a less-preferred alternative. A fundamental result in Gibbard (1977b) states that the converse is also true for the complete domain $\mathbb{P}$. [3]

Theorem 2.10 [Gibbard (1977b)] Assume $|A| \geq 3$. A RSCF $\varphi : \mathbb{P}^N \to \mathcal{L}(A)$ is strategy-proof and satisfies unanimity if and only if it is a random dictatorship.

In the paper, we investigate the structure of strategy-proof RSCFs satisfying unanimity for the domain $\mathbb{D}^S$. First, we introduce some concepts which we will refer to later. Let $\varphi : [\mathbb{D}^S]^N \to \mathcal{L}(A)$ be a RSCF. For every component j and every profile $P \in [\mathbb{D}^S]^N$, we can define the *marginal probability distribution* $\varphi_j(P)$ in an obvious way: for every $a_j \in A_j$, $\varphi_{j,a_j}(P) = \sum_{a_{M-j} \in A_{M-j}} \varphi_{(a_j, a_{M-j})}(P)$. A RSCF induces a *marginal random social choice function* MRSCF $\varphi : [\mathbb{D}^S]^N \to \mathcal{L}(A_j)$ by associating the marginal probability distribution $\varphi_j(P)$ over component j for every profile $P \in [\mathbb{D}^S]^N$.

We now recall results for *deterministic* social choice functions over the domain $\mathbb{D}^S$.

Definition 2.11 *A deterministic social choice function (DSCF) f is a map f : $[\mathbb{D}^S]^N \to A$.*

A DSCF is simply a RSCF whose image set is the set of degenerate probability distributions over A. We can similarly define component DSCFs. The definitions of strategy-proofness and unanimity for a DSCF are special cases of those of RSCFs and are omitted.

Definition 2.12 *A DSCF $f : [\mathbb{D}^S]^N \to A$ is a component dictatorship if there exists a map $\sigma : M \to N$ such that for all $P \in [\mathbb{D}^S]^N$,*

$$[f(P) = a] \Rightarrow [a_j = \tau(P_j^{\sigma(j)}, A_j)]$$

[3] Gibbard's result is actually more general than Theorem 2.10 below because it does not assume unanimity. However since unanimity will be a maintained hypothesis throughout the paper, we state only the version of the result with unanimity.

In a component dictatorship, the j^{th} component of the outcome at a profile is the maximal element of the j^{th} component of voter $\sigma(j)$. Since preferences are separable for all individuals, these maximal elements are well-defined. We can imagine the DSCF being decomposable into component DSCFs which are dictatorial. LeBreton and Sen (1999) establish a general decomposability result a special case of which is the result below.

Theorem 2.13 [LeBreton and Sen (1999)] Assume $|A_j| \geq 3$ for all $j \in M$. A DSCF $f : \mathbb{D}^N \to A$ is strategy-proof and satisfies unanimity if and only if it is a component dictatorship.

Does this result carry over to RSCFs? A particular generalization of component dictatorship to the random case is a RSCF where the probability distribution over the set A is the *product* of component random dictatorships. We define this below.

Definition 2.14 *A RSCF is $\varphi : [\mathbb{D}^S]^N \to \mathcal{L}(A)$ is an independent component random dictatorship if for each $j \in M$, there exists a random dictatorship $\varphi_j : \mathbb{P}^N \to \mathcal{L}(A_j)$ such that for all $P \in \mathbb{D}^N$*

$$\varphi(P) = \textstyle\prod_{j \in M} \varphi_j(P_j).$$

Consider the case where there are two voters and two components. Suppose the weight vectors for components 1 and 2 are (β_1^1, β_1^2) and (β_2^1, β_2^2) respectively. Then in any profile where voter 1 and 2's maximal elements are a_1a_2 and b_1b_2 respectively, the alternatives a_1a_2, a_1b_2, b_1a_2 and b_1b_2 get probability weights $\beta_1^1\beta_2^1$, $\beta_1^1\beta_2^2$, $\beta_2^1\beta_1^2$ and $\beta_1^2\beta_2^2$ respectively.

An independent component random dictatorship is strategy-proof (we shall verify this later) and clearly satisfies unanimity. Is every strategy-proof RSCF defined over the domain $\mathbb{D}^S$ with $|A_j| \geq 3$ an independent component random dictatorship? No, and this is established by the observation that a random dictatorship is strategy-proof but not an independent component random dictatorship unless it is deterministic, i.e. there exists a voter i such that $\beta^i = 1$. Showing that a random dictatorship is strategy-proof is routine. To demonstrate the other claim, consider for simplicity the case where there are two voters i and k and two components. Suppose also that $\beta^i, \beta^k > 0$. Consider a profile where i's maximal element is a_1a_2 and k's is b_1b_2 where $a_1 \neq b_1$ and $a_2 \neq b_2$. Observe that this RSCF would put probabilities β^i and β^k on a_1a_2 and b_1b_2 respectively and zero on all other alternatives. However every independent component random dictatorship which puts strictly positive probability on a_1a_2 and b_1b_2 also puts strictly positive probability on the alternatives a_1b_2 and b_1a_2.

Below, we formulate a generalization of both random dictatorship and independent component random dictatorship which coincides with the class of strategy-proof RSCFs satisfying unanimity in the case where each component set has at least three alternatives.

Let $\underline{i} \equiv (i_1, \ldots, i_m) \in I^m$ be an m-tuple of voters. We shall call such an m-tuple, a *voter sequence*. For all $a \in A$ and $P \in [\mathbb{D}^S]^N$, we shall let $\chi(a, P)$ denote the set of voter sequences *consistent with a and P* where $\chi(a, P) = \{\underline{i} \in I^m : a_j = \tau(P_j^{i_j}, A_j) \quad \text{for all } j = 1, \ldots m\}$.

Definition 2.15 *A RSCF $\varphi^g : [\mathbb{D}^S]^N \to \mathcal{L}(A)$ is a generalized random dictatorship if there exist non-negative real numbers $\gamma(\underline{i})$ for all $\underline{i} \in I^m$ with $\sum_{\underline{i} \in I^M} \gamma(\underline{i}) = 1$ such that for all $a \in A$ and $P \in [\mathbb{D}^S]^N$,*

$$\varphi_a^g(P) = \textstyle\sum_{\underline{i} \in \chi(a,P)} \gamma(\underline{i}).$$

Consider the following example. Suppose $I = \{1, 2\}$ and $A_j = \{a_j, b_j, c_j\}$ with $j = 1, 2$. Here $\underline{i}$ is one of four, two-tuples $(1, 1)$, $(1, 2)$, $(2, 1)$ and $(2, 2)$. The function γ specifies four non-negative real numbers $\gamma(1, 1)$, $\gamma(1, 2)$, $\gamma(2, 1)$ and $\gamma(2, 2)$ which sum to one. Consider a profile P where the maximal alternatives of voters 1 and 2 are (a_1a_2) and (b_1b_2) respectively. Observe that $\chi((a_1a_2), P) = \{(1, 1)\}$, $\chi((a_1b_2), P) = \{(1, 2)\}$, $\chi((b_1a_2), P) = \{(2, 1)\}$ and $\chi((b_1b_2), P) = \{(2, 2)\}$. Hence, a generalized random dictatorship puts probabilities of $\gamma(1,1)$, $\gamma(1,2)$, $\gamma(2,1)$ and $\gamma(2,2)$ on (a_1a_2), (a_1b_2), (b_1a_2) and (b_1b_2) respectively and zero on all other alternatives. Consider another profile $\bar{P}$ where voter 1 and 2's maximal alternatives are (a_1, c_2) and (a_1b_2) respectively. Here $\chi((a_1c_2), P) = \{(1,1),(2,1)\}$, and $\chi((a_1b_2), P) = \{(1,2),(2,1)\}$. Hence this RSCF will put probability $\gamma(1,1)+\gamma(2,1)$ on (a_1c_2) and $\gamma(1,2)+\gamma(2,2)$ on (a_1b_2) and zero on everything else.

In general, a generalized random dictatorship is specified by N^m non-negative real numbers adding up to one. For any $\underline{i} \equiv (i_1, \ldots, i_m)$, the probability of an alternative a in profile P is the sum of $\gamma(\underline{i})$'s over those $\underline{i}$'s which have the property that for every $j = 1, \ldots, m$, a_j is the maximal element in A_j for voter i_j, i.e. over all elements of the set $\chi(a, P_i)$.

We make several observations about generalized random dictatorships.

Observation 2.16 The value of a generalized random dictatorship at a profile depends only on the maximal alternatives (or "tops") of voter preferences at the profile. However it may assign positive probabilities to all alternatives in the *product set* of the top alternatives. In other words, a generalized random dictatorship is a probability distribution over the set $\prod_{j\in M}\{\tau(P_j^1, A_j), \ldots, \tau(P_j^N, A_j)\}$.

Observation 2.17 Let φ^g be a generalized random dictatorship with an associated map γ. Pick an arbitrary voter s and an arbitrary component j. Let $\beta^s = \sum_{\{\underline{i}\equiv(i_1,\ldots,i_m):i_j=s\}} \gamma(\underline{i})$. Clearly $0 \leq \beta^s \leq 1$ and $\sum_{s\in I}\beta^s = 1$. Pick a component j and a profile $P \in [\mathbb{D}^S]^N$. Observe that the probability of $a_j \in A_j$ in the marginal distribution $\varphi_j^g(P)$ is $\sum_{\{s:\tau(P_j^s)=a_j\}} \beta^s$. Hence, a generalized random dictatorship induces a MRSCF over each component that is a random dictatorship with respect to marginal preferences over that component. More formally, if φ^g is a generalized random dictatorship, there exist component random dictatorships $\varphi_j^r : \mathbb{P}^N \to \mathcal{L}(A)$, $j = 1, \ldots, m$ such that for all $P \in [\mathbb{D}^S]^N$ such that $\varphi_j^g(P) = \varphi_j^r(P_j)$ for each $j = 1, \ldots m$. As we shall remark at the end of the next section, this observation implies that our main result can be interpreted as a decomposability result for strategy-proof RSCFs for the domain $\mathbb{D}^S$.

Observation 2.18 A random dictatorship is a special case of a generalized random dictatorship when $\gamma(\underline{i}) = 0$ for all voter sequences $\underline{i}$ such that $i_j \neq i_{j'}$ for some $j \neq j'$. Equivalently, $\gamma(\underline{i}) > 0$ implies $\underline{i} = (i, i, \ldots, i)$ for some $i \in I$.

Observation 2.19 An independent component dictatorship is a special case of a generalized random dictatorship. Define the component random dictatorships as follows: for all $j = 1, \ldots, m$, let $\gamma_j(i)$, $i = 1, \ldots, N$ be a non-negative real numbers with $\sum_{i \in I} \gamma_j(i) = 1$. Now define a generalized random dictatorship as follows: for all voter sequences $\underline{i} \equiv (i_1, \ldots, i_m)$, $\gamma(\underline{i}) = \gamma_1(i_1) \times \gamma_2(i_2) \times \ldots \times \gamma_m(i_m)$.

Observation 2.20 In the special case where $m = 1$, a generalized random dictatorship is simply a random dictatorship.

In the next section we show that all strategy-proof RSCFs satisfying unanimity are generalized random dictatorships.

3. The Result

Our main result is the following.

Theorem 3.1 *Assume $|A_j| \geq 3$ for all $j = 1, \ldots, m$. A generalized random dictatorship $\varphi^g : [\mathbb{D}^S]^N \rightarrow \mathcal{L}(A)$ is strategy-proof and satisfies unanimity. If a RSCF $\varphi : [\mathbb{D}^L]^2 \rightarrow \mathcal{L}(A)$ is strategy-proof and satisfies unanimity then it is a generalized random dictatorship.*

Proof: (Sufficiency) Let φ^g be a generalized random dictatorship specified by the function γ in Definition 2.15. Let $P \in [\mathbb{D}^S]^N$ be an arbitrary separable profile and let i be an arbitrary voter. Consider a possible manipulation by i at P via $\bar{P}^i$. It follows from the definition of a generalized random dictatorship that the value of φ^g

at any profile depends only on the maximal alternatives of voters at the profile. Let $\tau(P^i) = a$ and $\tau(\bar{P}^i) = b$ where $b_Q \neq a_Q$ and $a_{M-Q} = b_{M-Q}$ for some non-empty subset Q of M.

Pick an arbitrary $\underline{i} \equiv (i_1, \ldots, i_m) \in N^m$. If $i_j \neq i$ for any $j \in Q$ then probability $\gamma(\underline{i})$ is assigned to the same alternative under profiles P and $(\bar{P}^i, P^{-i})$. If $i_j = i$ for all $j \in T$ for some $T \subseteq Q$, then probability $\gamma(\underline{i})$ is shifted from alternative (a_T, x_{M-T}) for some $x_{M-T} \in A_{M-T}$ in profile P to (b_T, x_{M-T}) in profile $(\bar{P}^i, P^{-i})$. However $a_j P^i_j b_j$ for all $j \in T$ by assumption so that $(a_T, x_{M-T}) P^i (b_T, x_{M-T})$ by separability. Therefore the distribution $\varphi^g(\bar{P}^i, P^{-i})$ is obtained by transferring probabilities from higher-ranked alternatives to lower-ranked alternatives according to P^i. from $\varphi^g(P)$. Clearly $\varphi^g(\bar{P}^i, P^{-i})$ stochastically dominates $\varphi^g(P)$ according to P^i and φ^g is strategy-proof.

(Necessity) We proceed as follows. In Step 1 we establish an important "conditional unanimity" property which holds for an arbitrary number of voters and in Step 2 we establish generalized random dictatorship in the case of two voters.

We begin with a Lemma which holds for arbitrary domains and is a straightforward adaptation of a result in Gibbard (1977b).

Let $\mathbb{D}$ be an arbitrary domain. Let $P^i \in \mathbb{D}$ and let $x, y \in A$ and assume that xP^iy. We say x and y are *contiguous* in P^i if there does not exist $z \in A$ distinct from x and y such that xP^izP^iy. We say that the ordering $\bar{P}_i$ is a *feasible local switch* of x and y in P^i if (i) x and y are contiguous (ii) xP^iy and $y\bar{P}^ix$ (iii)$B(x, P^i) = B(y, \bar{P}^i)$ (iv) $\bar{P}^i \in \mathbb{D}$.

Lemma 3.2 *Let $\varphi : \mathbb{D}^N \to \mathcal{L}(A)$ be strategy-proof. Let i be an arbitrary voter and let $\bar{P}^i$ be a feasible local switch of x and y in P^i (i.e xP^iy and $y\bar{P}^ix$). Then*

- $\varphi_y(\bar{P}^i, P^{-i}) \geq \varphi_y(P)$.
- $\varphi_x(\bar{P}^i, P^{-i}) + \varphi_y(\bar{P}^i, P^{-i}) = \varphi_x(P) + \varphi_y(P)$.

We omit the proof of this Lemma which is an implication of the definition of strategy-proofness.

Step 1. We consider an arbitrary strategy-proof RSCF $\varphi : [\mathbb{D}^L]^N \to \mathcal{L}(A)$ satisfying unanimity. Recall that every $P^i \in \mathbb{D}^L$ induces an ordering P^i_Q over A_Q for every $Q \subset M$.

The goal of this Step is to show the following. Pick an arbitrary non-empty subset $Q \subset M$. Then there exists a unanimous, strategy-proof RSCF $\varphi^Q : [\mathbb{D}^L_Q]^N \to \mathcal{L}(A_Q)$ such that for all profiles $P \in [\mathbb{D}^L]^N$ satisfying $\tau(P^i_{M-Q}, A_{M-Q}) = a_{M-Q}$ for all $i \in I$, we have

1. $[\varphi_x(P) > 0] \Rightarrow [x_{M-Q} = a_{M-Q}]$ and
2. $\varphi(P) = \varphi^Q(P_Q)$.

Thus, there exists a strategy-proof RSCF φ_Q defined for every arbitrary non-empty set of components Q with the property that whenever all voters are unanimous with respect to say $a_{M-Q} \in A_{M-Q}$, then φ (i) puts strictly positive probability only on those alternatives whose $M - Q$ are given by a_{M-Q} and (ii) the probability of an alternative (a_Q, a_{M-Q}) in the profile P is the probability given to a_Q in the RSCF φ^Q in the component Q induced profile P_Q. Moreover φ_Q satisfies unanimity.

We first establish some preliminary lemmas. Let $\succ$ be an ordering over the set M and let $j \in M$. Then $E(\succ, j) = \{i \in M : i \succ j\}$. Thus $E(\succ, j)$ is the set of components which lexicographically dominate j.

Lemma 3.3 *Let $P \in [\mathbb{D}^L]^N$ and $i \in I$. Let P^i be lexicographic with respect to $\succ$. Let j be a component and let $\succ'$ be another ordering over M such that (i) $E(\succ, j) = E(\succ', j) = Q$ and (ii) $\succ$ and $\succ'$ agree on Q. Let $\bar{P}^i$ be lexicographic with respect to $\succ'$ such that $P^i_Q = \bar{P}^i_Q$. Then $\varphi_k(P) = \varphi_k(\bar{P}^i, P_{-i})$ for all $k \in Q$. Also let $a_Q \in A_Q$. Then $\varphi_{Q,a_Q}(P) = \sum_{\{b: b_Q = a_Q\}} \varphi_b(P) = \sum_{\{b: b_Q = a_Q\}} \varphi_b(\bar{P}^i, P^{-i}) = \varphi_{Q,a_Q}(\bar{P}^i, P^{-i})$.*

Proof: We first prove the first part of the Lemma. Suppose it is false. Let k be $\succ$ maximal element in Q such that $\varphi_k(P) \neq \varphi_k(\bar{P}^i, P^{-i})$. Let $b_k \in A_k$ be the P^i_k-maximal element in A_k such that $\varphi_{k,b_k}(P) \neq \varphi_{k,b_k}(\bar{P}^i, P^{-i})$; in other words, b_k is the maximal element in P^i_k such that the marginal probability of b_k changes as i switches from P^i to $\bar{P}^i$. Note that

$$\varphi_{k,b_k}(P) = \sum_{x_{Q-k}} \sum_{x_{M-Q}} \varphi_{(x_{Q-k}, b_k, x_{M-Q})}(P)$$

and

$$\varphi_{k,b_k}(\bar{P}^i, P^{-i}) = \sum_{x_{Q-k}} \sum_{x_{M-Q}} \varphi_{(x_{Q-k}, b_k, x_{M-Q})}(\bar{P}^i, P^{-i})$$

The LHS of the expressions above are not equal to each other. Let c_{Q-k} be the maximal alternative in A_{Q-k} such that

$$\sum_{x_{M-Q}} \varphi_{(c_{Q-k}, b_k, x_{M-Q})}(P) \neq \sum_{x_{M-Q}} \varphi_{(c_{Q-k}, b_k, x_{M-Q})}(\bar{P}^i, P^{-i})$$

Suppose

$$\sum_{x_{M-Q}} \varphi_{(c_{Q-k}, b_k, x_{M-Q})}(P) < \sum_{x_{M-Q}} \varphi_{(c_{Q-k}, b_k, x_{M-Q})}(\bar{P}^i, P^{-i})$$

Let w_{M-Q} be the worst alternatives in A_{M-Q} according to P^i_{M-Q}. Since P^i is lexicographic with respect to $\succ$ it must be the case that $\bar{B} = B((c_{Q-k}, b_k, w_{M-Q}), P^i)$ is the set of alternatives (x_{Q-k}, x_k, w_{M-Q}) where either $x_{Q-k} P^i_{Q-k} c_{Q-k}$ or $x_{Q-k} = c_{Q-k}$ and $x_k P^i_k b_k$.

Then,

$$\begin{aligned}
& \sum_{x\in\bar{B}} \varphi_x(\bar{P}^i, \bar{P}^{-i})) \\
= & \sum_{\{x: x_{Q-k} P^i_{Q-k} c_{Q-k}\}} \varphi_x(\bar{P}^i, P^{-i}) + \sum_{\{(x_k, x_{M-Q}): x_k P^i_k b_k\}} \varphi_{(c_{Q-k}, x_k, x_{M-Q})}(\bar{P}^i, P^{-i}) \\
+ & \sum_{x_{M-Q}} \varphi_{(c_{Q-k}, b_k, x_{M-Q})}(\bar{P}^i, P^{-i}) \\
> & \sum_{\{x: x_{Q-k} P^i_{Q-k} c_{Q-k}\}} \varphi_x(P) + \sum_{\{(x_k, x_{M-Q}): x_k P^i_k b_k\}} \varphi_{(c_{Q-k}, x_k, x_{M-Q})}(P) \\
+ & \sum_{x_{M-Q}} \varphi_{(c_{Q-k}, b_k, x_{M-Q})}(P) \\
= & \sum_{x\in\bar{B}} \varphi_x(P)
\end{aligned}$$

contradicting strategy-proofness. Note that the strict inequality from the observations that

$$\sum_{\{x: x_{Q-k} P^i_{Q-k} c_{Q-k}\}} \varphi_x(\bar{P}^i, P^{-i}) = \sum_{\{x: x_{Q-k} P^i_{Q-k} c_{Q-k}\}} \varphi_x(P)$$

$$\sum_{\{(x_k, x_{M-Q}): x_k P^i_k b_k\}} \varphi_{(c_{Q-k}, x_k, x_{M-Q})}(\bar{P}^i, P^{-i}) = \sum_{\{(x_k, x_{M-Q}): x_k P^i_k b_k\}} \varphi_{(c_{Q-k}, x_k, x_{M-Q})}(P)$$

and

$$\sum_{x_{M-Q}} \varphi_{(c_{Q-k}, b_k, x_{M-Q})}(\bar{P}^i, P^{-i}) > \sum_{x_{M-Q}} \varphi_{(c_{Q-k}, b_k, x_{M-Q})}(P)$$

which follow from our definitions of c_{Q-k} and b_k.

The remaining case is when

$$\sum_{x_{M-Q}} \varphi_{(c_{Q-k}, b_k, x_{M-Q})}(P) > \sum_{x_{M-Q}} \varphi_{(c_{Q-k}, b_k, x_{M-Q})}(\bar{P}^i, P^{-i})$$

Using the fact that $\bar{B} = B((c_{Q-k}, b_k, w_{M-Q}), \bar{P}^i)$ (since $P^i_Q = \bar{P}^i_Q$) we can use arguments analogous to the ones above to show that voter i manipulates from $(\bar{P}^i, P^{-i})$ via P^i.

To prove the second part of the Lemma, we replace k and b_k above with Q and a_Q respectively and replicate the arguments above.

Lemma 3.4 *Let $P \in [\mathbb{D}^L]^N$, $i \in I$ and $\bar{P}^i \in \mathbb{D}^L$ be such that (i) P^i and $\bar{P}^i$ are marginally equivalent, i.e. $P^i_j = \bar{P}^i_j$ for all $j \in M$ and (ii) if P^i and $\bar{P}^i$ are lexicographic with respect to the orderings $\succ$ and $\bar{\succ}$ over M respectively, then $\succ$ and $\bar{\succ}$ agree over all components except j and k where j and k are contiguous in $\succ$. If $\varphi(P) \neq \varphi(\bar{P}^i, P^{-i})$ then $\varphi_j(P) \neq \varphi_j(\bar{P}^i, P^{-i})$ and $\varphi_k(P) \neq \varphi_k(\bar{P}^i, P^{-i})$.*

Proof: Suppose the Lemma is false. In view of Lemma 3.3, we can assume that j and k are the lexicographic best and second best components respectively in P^i and the lexicographic second and best components in $\bar{P}^i$ respectively. We have therefore assumed that $\varphi(P^i, P^{-i}) \neq \varphi(\bar{P}^i, P^{-i})$ but $\varphi_l(P^i, P^{-i}) = \varphi_l(\bar{P}^i, P^{-i})$ for all components l. Let a be the highest ranked alternative in P^i such that $\varphi_a(P^i, P^{-i}) \neq \varphi_a(\bar{P}^i, P^{-i})$. Since φ is strategy-proof, it must be the case that $\varphi_a(P^i, P^{-i}) > \varphi_a(\bar{P}^i, P^{-i})$. Let $Y = \{x \in A : x_k \bar{P}^i_k a_k\} = \{x \in A : x_k P^i a_k\}$ (since $\bar{P}^i_k = P^i_k$). Let $Z = \{x \in A : x_k = a_k \text{ and } x\bar{P}^i a\}$. Note that $Z = \{x \in A : x_k = a_k \text{ and } xP^i a\}$ since P^i and $\bar{P}^i$ are marginally equivalent orderings and the lexicographic ordering of components in $M \setminus k$ in the two orderings is also the same. Note the following

(i) $B(a, \bar{P}^i) = Y \cup Z \cup \{a\}$

(ii) $\varphi_z(\bar{P}^i, P^{-i}) = \varphi_z(P^i, P^{-i})$ for all $z \in Z$ since $z \in Z$ implies that $zP^i a$ and a is the highest-ranked alternative P^i such that $\varphi(P^i, P^{-i}) \neq \varphi(\bar{P}^i, P^{-i})$ and

(iii) $\sum_{\{x_k : x_k \bar{P}^i_k a_k\}} \varphi_{k,x_k}(\bar{P}^i, P^{-i}) = \sum_{\{x_k : x_k \bar{P}^i_k a_k\}} \varphi_{k,x_k}(P^i, P^{-i})$ by virtue of our assumption that the φ yields the same marginal probability distribution all over components.

Hence,

$$\begin{aligned}
\sum_{x\in B(a,\bar{P}^i)} \varphi_x(P^i,P^{-i}) &= \sum_{x\in Y}\varphi_x(P^i,P^{-i}) + \sum_{x\in Z}\varphi_x(P^i,P^{-i}) + \varphi_a(P^i,P^{-i}) \\
&= \sum_{\{x_k: x_k P^i_k a_k\}} \varphi_{k,x_k}(P^i,P^{-i}) + \sum_{x\in Z}\varphi_x(P^i,P^{-i}) + \varphi_a(P^i,P^{-i}) \\
&> \sum_{\{x_k: x_k \bar{P}^i_k a_k\}} \varphi_{k,x_k}(\bar{P}^i,P^{-i}) + \sum_{x\in Z}\varphi_x(\bar{P}^i,P^{-i}) + \varphi_a(\bar{P}^i,P^{-i}) \\
&= \sum_{x\in Y}\varphi_{k,x_k}(\bar{P}^i,P^{-i}) + \sum_{x\in Z}\varphi_x(\bar{P}^i,P^{-i}) + \varphi_a(\bar{P}^i,P^{-i}) \\
&= \sum_{x\in B(a,\bar{P}^i)} \varphi_x(\bar{P}^i,P^{-i})
\end{aligned}$$

Consequently voter i manipulates at $(\bar{P}^i, P^{-i})$ via P^i contradicting the strategy-proofness of φ.

We now return to Step 1. The first lemma asserts that φ satisfies a conditional unanimity property.

Lemma 3.5 *Let $Q \subset M$, $P \in [\mathbb{D}^L]^N$ and $a \in A$ be such that $\tau(P^i_{M-Q}, A_{M-Q}) = a_{M-Q}$ for all $i \in I$. Then $[\varphi_b(P) > 0] \Rightarrow [b_{M-Q} = a_{M-Q}]$.*

Proof: Suppose that the Lemma is false. Assume that $\tau(P^i_{M-Q}, A_{M-Q}) = a_{M-Q}$ for all $i \in I$ but $\varphi_b(P) > 0$ where $b_{M-Q} \neq a_{M-Q}$. For all $i \in I$, let $\bar{P}^i \in \mathbb{D}^L$ be such that (i) $\bar{P}^i_k = P^i_k$ for all $k \in M - Q$ (ii) $\tau(\bar{P}^i_Q, A_j) = b_Q$ and (iii) all components in Q lexicographically dominate all components in $M - Q$.

Pick an arbitrary voter i and suppose $\varphi_x(\bar{P}^i, P^{-i}) = 0$ whenever $x_Q = b_Q$. For any $k \neq j$ let $d_k \in A_k$ be the worst ranked element in A_k according to $\bar{P}^1_k$ (and P^1_k). Let $\bar{B} = B((b_Q, d_{M-Q}), \bar{P}^i)$. Since components in Q lexicographically dominate those in $M - Q$ in $\bar{P}^i$, it follows that $c \in \bar{B} \Rightarrow [c_Q = b_Q]$. Therefore

$$\sum_{x\in\bar{B}} \varphi_x(P^i,P^{-i}) \geq \varphi_b(P) > 0 = \sum_{x\in\bar{B}} \varphi_x(\bar{P}^i,P^{-i}).$$

Hence i manipulates φ at $(\bar{P}^i, P_{-i})$ via P^i. Therefore, $\varphi_{(b_Q, c_{M-Q})}(\bar{P}^i, P^{-i}) > 0$ for some $c_{M-Q} \in A_{M-Q}$.

Now suppose $\varphi_{(b_Q, a_{M-Q})}(\bar{P}^i, P^{-i}) = 1$. Let $\hat{B} = B((b_Q, a_{M-Q}), P^i)$. Note that $(b_Q, a_{M-Q})P^i b$ since $a_{M-Q} = \tau(P^i, A_{M-Q})$. Since

$$\sum_{x \in \hat{B}} \varphi_x(P^i, P^{-i}) < \sum_{x \in \hat{B}} \varphi_x(\bar{P}^i, P^{-i}) = 1$$

voter i will manipulate at P via $\bar{P}^i$. Therefore $\varphi_{(b_Q, a_{M-Q})}(\bar{P}^i, P^{-i}) < 1$.

We can conclude from the arguments in the two previous paragraphs that there exists $c_{M-Q} \in A_{M-Q} \setminus \{a_{M-Q}\}$ such that $\varphi_{(b_Q, c_{M-Q})}(\bar{P}^i, P^{-i}) > 0$. Now pick a voter $i' \neq i$ and replace $P^{i'}$ in the profile $(\bar{P}^i, P^{-i})$ by $\bar{P}^{i'}$. Replicating the arguments above, we can conclude that there exists $d_{M-Q} \in A_{M-Q} \setminus \{a_{M-Q}\}$ such that $\varphi_{(b_Q, d_{M-Q})}(\bar{P}^i, \bar{P}^{i'}, \bar{P}^{-i,i'}) > 0$. Proceeding in this manner, it follows that $\varphi_{(b_Q, x_{M-Q})}(\bar{P}) > 0$ where $x_{M-Q} \in A_{M-Q} \setminus \{a_{M-Q}\}$. But all voters have (b_Q, a_{M-Q}) as their first-ranked alternative in the profile $\bar{P}$. Hence φ violates unanimity completing the proof of the Lemma.

For every $Q \subset M$ and $a \in A$, Let $[\mathbb{D}^L(a, Q)]^N \subset [\mathbb{D}^L]^N$ be the set of lexicographic profiles P with the property that $\tau(P^i_Q, A_{M-Q}) = a_{M-Q}$.

Lemma 3.6 *Let $Q \subset M$ and $a \in A$. Let $\hat{P}, \bar{P} \in [\mathbb{D}^L(a, Q)]^N$ be such that $\hat{P}_Q = \bar{P}_Q$. Then $\varphi(\hat{P}) = \varphi(\bar{P})$.*

Proof: It follows from Lemma 3.5 that $[\varphi_b(\hat{P}) > 0] \Rightarrow [b_Q = a_Q]$ and $[\varphi_b(\bar{P}) > 0] \Rightarrow [b_Q = a_Q]$. We first claim that $\varphi(\bar{P}^i, \hat{P}^{-i}) = \varphi(\hat{P})$ for an arbitrary voter i. Suppose this is false. Let c_Q be the best-alternative in A_Q according to $\hat{P}^i_Q = \bar{P}^i_Q$ such that

$$\varphi_{(c_Q, a_{M-Q})}(\bar{P}^i, \hat{P}^{-i}) \neq \varphi_{(c_Q, a_{M-Q})}(\bar{P}^i, \hat{P}^{-i}).$$

Such an alternative c_Q must exist. If

$$\varphi_{(c_Q, a_{M-Q})}(\bar{P}^i, \hat{P}^{-i}) > \varphi_{(c_Q, a_{M-Q})}(\hat{P})$$

then

$$\sum_{x \in B((c_Q, a_{M-Q}), \hat{P}^i)} \varphi_x(\bar{P}^i, \hat{P}^{-i}) > \sum_{x \in B((c_Q, a_{M-Q}), \hat{P}^i)} \varphi_x(\hat{P})$$

contradicting the strategy-proofness of φ. If

$$\varphi_{(c_Q, a_{M-Q})}(\bar{P}^i, \hat{P}^{-i}) < \varphi_{(c_Q, a_{M-Q})}(\hat{P})$$

then

$$\sum_{x \in B((b_Q, a_{M-Q}), \bar{P}^i)} \varphi_x(\hat{P}) > \sum_{x \in B((c_Q, a_{M-Q}), \bar{P}^i)} \varphi_x(\bar{P}^i, \hat{P}^{-i})$$

again contradicting the strategy-proofness of φ. Therefore $\varphi(\bar{P}^i, \hat{P}^{-i}) = \varphi(\hat{P})$. Progressively switching preferences of voters from $\hat{P}^i$ to $\bar{P}^i$ and repeatedly applying these arguments above yields $\varphi(\hat{P}) = \varphi(\bar{P})$ as required.

Let $a \in A$ and $Q \subset M$. We define the function $\varphi^{a,Q} : [\mathbb{D}^L(a, Q)]^N \to \mathcal{L}(A_Q)$ as follows: for all $P_Q \in [\mathbb{D}^L(a, j)]^N$, $\varphi^{a,Q}(P_Q) = \varphi(P)$. Thus we obtain $\varphi^{a,Q}$ by considering a profile $P \in [\mathbb{D}^L(a, Q)]^N$ and equating the probability that $\varphi^{a,Q}(P_Q)$ gives to every $b_Q \in A_Q$ with $\varphi_{(b_j, a_{M-Q})}(P)$. A critical observation is that Lemma 3.6 implies that $\varphi^{a,Q}$ is well-defined. The next Lemma demonstrates that it is strategy-proof.

Lemma 3.7 $\varphi^{a,Q}$ *is strategy-proof and satisfies unanimity.*

Proof: Suppose $\varphi^{a,Q}$ is not strategy-proof. Then there must exist $i \in I$, $P_Q \in [\mathbb{D}^L(a, Q)]^N$, $\bar{P}^i_Q \in [\mathbb{D}^L(a, Q)]$ and $b_Q \in A_Q$ such that

$$\sum_{x_Q \in B(b_Q, P^i_Q)} \varphi^{a,Q}_{x_Q}(\bar{P}^i_Q, P^{-i}_Q) > \sum_{x_Q \in B(b_Q, P^i_Q)} \varphi^{a,Q}_{x_Q}(P_Q).$$

Let $\hat{P} \in [\mathbb{D}^L]^N$ be a profile and $\tilde{P}^i \in \mathbb{D}^L$ be an ordering such that (i) $\hat{P}_Q = P_Q$ (ii) $\tau(\hat{P}^t_{M-Q}, A_{M-Q}) = a_{M-Q}$ for all voters $t \in I$ (iii) all components in Q are lexicographically dominated by those in $M - Q$ in $\hat{P}^t$ for all $t \in I$ (iv) $\tilde{P}^i_Q = \bar{P}^i_Q$ (v)

$\tau(\tilde{P}^i_{M-Q}, A_{M-Q}) = a_{M-Q}$ and (vi) all components in Q are lexicographically dominated by those in $M-Q$ in $\tilde{P}^i$.

By construction, the profiles $\hat{P}, (\tilde{P}^i, \hat{P}^{-i}) \in [\mathbb{D}^S(a,Q)]^N$. Hence Lemma 3.5 implies that $[\varphi_b(\hat{P}) > 0] \Rightarrow [b_{M-Q} = a_{M-Q}]$ and $[\varphi_b(\tilde{P}^i, \hat{P}^{-i}) > 0] \Rightarrow [b_{M-Q} = a_{M-Q}]$. Since components in Q are lexicographically dominated by those in $M-Q$ in both $\hat{P}^i$ and $\tilde{P}^i$ and $\tau(\hat{P}^i_{M-Q}, A_{M-Q}) = a_{M-Q}$ we must have $B((b_Q, a_{M-Q}), \hat{P}^i) = \{(x_Q, a_{M-Q}) : x_Q \in B(b_Q, \hat{P}^i_Q)\}$.

Consequently

$$\begin{aligned}
\sum_{x \in B((b_Q, a_{M-Q}), \hat{P}^i)} \varphi_x(\tilde{P}^i, \hat{P}^{-i}) &= \sum_{x_Q \in B(b_Q, \hat{P}^i_Q)} \varphi^{a,Q}_{x_Q}(\tilde{P}^i_Q, \hat{P}^{-i}_Q) \\
&= \sum_{x_Q \in B(b_Q, P^i_Q)} \varphi^{a,Q}_{x_Q}(\tilde{P}^i_Q, \hat{P}^{-i}_Q) \\
&> \sum_{x_Q \in B(b_Q, P^i_Q)} \varphi^{a,Q}_{x_Q}(P_Q) \\
&= \sum_{x \in B((b_Q, a_{M-Q}), \hat{P}^i)} \varphi_x(\hat{P})
\end{aligned}$$

contradicting the strategy-proofness of φ. Therefore $\varphi^{a,Q}$ is strategy-proof.

Now let P be a profile such that $P_Q \in [\mathbb{D}^L(a,Q)]^N$ be a profile such that all voters are unanimous with respect to components in Q, i.e. suppose $\tau(P^i_Q, A_Q) = b_Q$ for some $b_Q \in A_Q$. Clearly $\tau(P^i, A) = (b_Q, a_{M-Q})$. Since φ satisfies unanimity, $\varphi(P) = (b_Q, a_{M-Q})$ which implies that $\varphi^{a,Q}(P_Q) = b_Q$. Therefore $\varphi^{a,Q}$ satisfies unanimity.

Lemma 3.8 *$\varphi^{a,Q}$ does not depend on a i.e. $\varphi^{a,Q} = \varphi^{b,Q}$ for all $b \in A$.*

Proof: Suppose not, i.e. $\varphi^{a,Q}(P_Q) \neq \varphi^{b,Q}(P_Q)$ for some $a, b \in A$ and $P_Q \in [\mathbb{D}^L_Q]^N$. Assume without loss of generality that $\varphi^{a,Q}_{x_Q}(P_Q) > \varphi^{b,Q}_{x_Q}(P_Q)$ for some $x_Q \in A_Q$. Let $\bar{P} \in [\mathbb{D}^L(a,Q)]^N$ and $\hat{P} \in [\mathbb{D}^L(b,Q)]^N$ be such that (i) $\hat{P}_Q = \bar{P}_Q = P_Q$, (ii) all components in Q lexicographically dominate all components in $M-Q$ in $\bar{P}^i$ for all

$i \in I$ and (iii) all components in Q lexicographically dominate all components in $M - Q$ in $\hat{P}^i$ for all $i \in I$. It follows that

$$\varphi_{(x_Q, a_{M-Q})}(\bar{P}) = \varphi^{a,Q}_{x_Q}(P_Q) > \varphi^{b,Q}_{x_Q}(P_Q) = \varphi_{(x_Q, b_{M-Q})}(\hat{P}).$$

Moreover, since $\varphi_{(x_Q, y_{M-Q})}(\bar{P}) > 0$ only if $y_{M-Q} = a_{M-Q}$ and $\varphi_{(x_Q, y_{M-Q})}(\hat{P}) > 0$ only if $y_{M-Q} = b_{M-Q}$, it follows that

$$\sum_{c_{M-Q}} \varphi_{(x_Q, c_{M-Q})}(\bar{P}) > \sum_{c_{M-Q}} \varphi_{(x_Q, c_{M-Q})}(\hat{P}).$$

Let voter i switch from $\bar{P}^i$ to $\hat{P}^i$ in the profile $\bar{P}$. We claim that

$$\sum_{c_{M-Q}} \varphi_{(x_j, c_{M-j})}(\bar{P}) = \sum_{c_{M-Q}} \varphi_{(x_Q, c_{M-Q})}(\hat{P}^i, \bar{P}^{-i}).$$

Suppose this is false. Suppose that

$$\sum_{\{c_{M-Q} \in A_{M-Q}\}} \varphi_{(x_Q, c_{M-Q})}(\bar{P}) < \sum_{\{c_{M-Q} \in A_{M-Q}\}} \varphi_{(x_Q, c_{M-Q})}(\hat{P}^i, \bar{P}^{-i}).$$

Let d_{M-Q} be the worst alternative in A_{M-Q} according to P^i_Q, i.e. $y_{M-Q} P^i_Q d_{M-Q}$ for all $y_{M-Q} \neq d_{M-Q}$. Since components in Q lexicographically dominate components in $M - Q$ in $\bar{P}^i$, we have

$$\begin{aligned}
\sum_{x \in B((x_Q, d_{M-Q}), \bar{P}^i)} \varphi_x(\hat{P}^i, \bar{P}^{-i})) &= \sum_{c_{M-Q} \in A_{M-Q}} \varphi_{(x_Q, c_{M-Q})}(\hat{P}^i, \bar{P}^{-i}) \\
&> \sum_{c_{M-Q} \in A_{M-Q}} \varphi_{(x_Q, c_{M-Q})}(\bar{P}) \\
&= \sum_{x \in B((x_Q, d_{M-Q}), \bar{P}^i)} \varphi_x(\bar{P})
\end{aligned}$$

contradicting the strategy-proofness of φ. An analogous argument shows that

$$\sum_{\{c_{M-Q} \in A_{M-Q}\}} \varphi_{(x_Q, c_{M-Q})}(\bar{P}) > \sum_{\{c_{M-Q} \in A_{M-Q}\}} \varphi_{(x_Q, c_{M-Q})}(\hat{P}^i, \bar{P}^{-i})$$

cannot hold. Progressively switching preferences of voters from $\bar{P}$ to $\hat{P}$, we obtain

$$\sum_{c_{M-Q}} \varphi_{(x_Q,d_{M-Q})}(\bar{P}) = \sum_{c_{M-Q}} \varphi_{(x_Q,c_{M-Q})}(\hat{P}).$$

However this contradicts our earlier conclusion that

$$\sum_{c_{M-Q}} \varphi_{(x_Q,c_{M-Q})}(\bar{P}) > \sum_{c_{M-Q}} \varphi_{(x_Q,c_{M-Q})}(\hat{P}).$$

This concludes Step 1.

Step 2: The goal of this step is to show the following: Let $I = \{1, 2\}$ and let $\varphi : [\mathbb{D}^L]^2 \to \mathcal{L}(A)$ be a strategy-proof RSCF satisfying unanimity. Then φ is a generalized random dictatorship. Throughout Step 2, we assume that φ is a two-voter RSCF defined on the domain of lexicographic preferences that is strategy-proof and satisfies unanimity.

For any $P \in [\mathbb{D}^L]^2$, the *Top Product Set* at P or $TPS(P)$ is defined as follows:

$$TPS(P) \equiv \{\tau(P_1^1, A_1) \cup \{\tau(P_1^2, A_1)\} \times \ldots \times \{\tau(P_m^1, A_m)\} \cup \{\tau(P_m^2, A_m)\}.$$

We say that φ satisfies the TPS Property if

$$\sum_{a \in TPS(P)} \varphi_a(P) = 1 \quad \text{for all } P \in [\mathbb{D}^L]^2.$$

Lemma 3.9 *φ satisfies the TPS Property.*

Proof: Suppose the Lemma is false. Then there exists $P \in [\mathbb{D}^L]^2$, $a, b, c \in A$ and $j \in M$ such that $\varphi_c(P) > 0$ and $\tau(P^1, A) = a$, $\tau(P^2, A) = b$ and $c_j \notin \{a_j, b_j\}$. We consider two mutually exhaustive cases.

Case 1: Component j is the lexicographically worst component in P^1 and P^2.

Claim 1: If $\varphi_{j,x_j}(P) > 0$ then $a_j P_j^1 x_j P_j^1 b_j$ and $b_j P_j^2 x_j P_j^2 a_j$.

Suppose that the Claim is false. Assume without loss of generality that $b_j P_j^1 x_j$ and $\varphi_{j,x_j}(P) > 0$. Consider $\bar{P}^1 \in \mathbb{D}^L$ such that (i) $\bar{P}^1_{M-j} = P^1_{M-j}$ and (ii) $\tau(\bar{P}^1_j, A_j) = b_j$. From Lemma 3.3, we deduce that $\sum_{z_j} \varphi_{(y_{M-j},z_j)}(P) = \sum_{z_j} \varphi_{(y_{M-j},z_j)}(\bar{P}^1, P^2)$ for all $y_{M-j} \in A_{M-j}$. Also $\varphi_{j,b_j}(\bar{P}^1, P^2) = 1$ from Step 1, i.e. $\varphi_{j,x_j}(\bar{P}^1, P^2) = 0$. There must therefore exist y_{M-j} such that

$$\varphi_{(y_{M-j},x_j)}(P) > \varphi_{(y_{M-j},x_j)}(\bar{P}^1, P^2) = 0$$

In fact, assume without loss of generality that y_{M-j} is the P^1_{M-j}-maximal alternative in A_{M-j} with this property. Let $\bar{B} = B((y_{M-j}, x_j), P^1) \setminus (y_{M-j}, x_j)$. Then

$$\sum_{z \in \bar{B}} \varphi_z(\bar{P}^1, P^2) > \sum_{z \in \bar{B}} \varphi_z(P).$$

Therefore 1 manipulates at P via $\bar{P}^1$. This proves Claim 1.

In view of Claim 1, we can assume that $a_j P_j^1 c_j P_j^1 b_j$ and $b_j P_j^2 c_j P_j^2 a_j$. Let $\bar{P}^1 \in \mathbb{D}^L$ be such that (i) $\bar{P}^1_{M-j} = P^1_{M-j}$ and (ii) $\tau(\bar{P}^1_j, A_j) = a_j$ and b_j is ranked second in A_j according to $\bar{P}^1_j$. We claim that $\varphi_{j,a_j}(P) = \varphi_{j,a_j}(\bar{P}^1, P^2)$. If $\varphi_{j,a_j}(P) < \varphi_{j,a_j}(\bar{P}^1, P^2)$, then we can construct an argument analogous to the one above to show that 1 manipulates at P via $\bar{P}^1$. If the reverse is true, 1 manipulates at $(\bar{P}^1, P^2)$ via P^1. It follows from our earlier arguments that $\varphi_{j,z_j}(\bar{P}^1, P_2) = 0$ for all $z_j \neq a_j, b_j$. Since $\varphi_{j,c_j}(P) > 0$ and $\varphi_{j,a_j}(P) = \varphi_{j,a_j}(\bar{P}^1, P^2)$, we must have $\varphi_{j,b_j}(P) < \varphi_{j,b_j}(\bar{P}^1, P^2)$.

Now construct $\bar{P}^2 \in \mathbb{D}^L$ be such that (i) $\bar{P}^2_{M-j} = P^2_{M-j}$ and (ii) $\tau(\bar{P}^2_j, A_j) = b_j$ and a_j is ranked second in A_j according to $\bar{P}^2_j$. From our earlier arguments $\varphi_{j,z_j}(\bar{P}) = 0$ for all $z_j \neq a_j, b_j$ and $\varphi_{j,b_j}(\bar{P}) = \varphi_{j,b_j}(\bar{P}^1, P^2)$. Therefore $\varphi_{j,a_j}(\bar{P}) = \varphi_{j,a_j}(\bar{P}^1, P^2)$. Hence $\varphi_{j,b_j}(\bar{P}) > \varphi_{j,b_j}(P)$.

Now consider $\varphi_j(P^1, \bar{P}^2)$. Using the same arguments as before, we can deduce that $\varphi_{j,b_j}(P) = \varphi_{j,b_j}(P^1, \bar{P}^2)$, $\varphi_{j,a_j}(P) < \varphi_{j,a_j}(P^1, \bar{P}^2)$ and $\varphi_{j,z_j}(P^1, \bar{P}^2) = 0$ for all $z_j \neq a_j, b_j$. Furthermore $\varphi_{j,b_j}(\bar{P}) = \varphi_{j,b_j}(P^1, \bar{P}^2)$, $\varphi_{j,a_j}(\bar{P}) = \varphi_{j,a_j}(P^1, \bar{P}^2)$ and

$\varphi_{j,z_j}(\bar{P}) = 0$ for all $z_j \neq a_j, b_j$. Hence $\varphi_{j,b_j}(\bar{P}) = \varphi_{j,b_j}(P)$ contradicting our earlier conclusion that $\varphi_{j,b_j}(\bar{P}) > \varphi_{j,b_j}(P)$. This completes Case 1.

Case 2: Case 1 does not hold. Assume without loss of generality that j is not the lexicographically worst component in P^1. Let S and T denote the set of components lexicographically worse than j and lexicographically better than j respectively. Let $\bar{P}^1 \in \mathbb{D}^L$ such that (i) $\bar{P}^1_T = P^1_T$ (ii) the set of components lexicographically better than j in $\bar{P}^i$ is T and (iii) $\tau(\bar{P}^1_S, A_S) = b_S$. From Lemma 3.3 $\varphi_j(P) = \varphi_j(\bar{P}^1, P^2)$. Let $\hat{P}^1 \in \mathbb{D}^L$ be such that (i) the ordering of all components other than j is the same as in $\bar{P}^i$ and j is the lexicographically worst (ii) $\hat{P}^1_k = \bar{P}^1_k$ for all $k \in M$. We claim that $\varphi_j(\bar{P}^1, P^2) = \varphi_j(\hat{P}^1, P^2) = \varphi_j(P)$. Consider $k \in T$. It follows from Lemma 3.3 that $\varphi_k(\bar{P}^1, P^2) = \varphi_k(\hat{P}^1, P^2)$. For components $k \in S$, $\varphi_{k,b_k}(\hat{P}^1, P^2) = \varphi_{k,b_k}(\bar{P}^1, P^2) = 1$ from Step 1. Therefore $\varphi_k(\bar{P}^1, P^2) = \varphi_k(\hat{P}^1, P^2)$ for all $k \neq j$. If $\varphi_j(\bar{P}^1, P^2) \neq \varphi_j(\hat{P}^1, P^2)$, we can use arguments analogous to ones used earlier to show that either 1 manipulates at $(\bar{P}^1, P^2)$ via $\hat{P}^1$ or at $(\hat{P}^1, P^2)$ via $\bar{P}^1$. This establishes the claim. Similarly we can find $\hat{P}^2 \in \mathbb{D}^L$ where j is lexicographically worst and $\varphi_j(\hat{P}) = \varphi_j(P)$. Therefore $\varphi_{j,c_j}(\hat{P}) > 0$. Note that $\hat{P}^i_j, A_j) \neq c_j$ for $i = 1, 2$. Hence we are in the situation described in Case 1 and we can use the same arguments to show that $\varphi_{j,c_j}(\hat{P}) > 0$ is not possible.

Recall that profiles $P, \bar{P} \in [\mathbb{D}^L]^2$ are said to be marginally equivalent if, $P^i_Q = \bar{P}^i_Q$ for all $i = \{1, 2\}$ and $Q \subset M$. According to our next lemma, the outcome of a strategy-proof RSCF in the two-voter case is identical across marginally equivalent profiles.

Lemma 3.10 *Let $P, \bar{P} \in [\mathbb{D}^L]^2$ be marginally equivalent profiles. Then $\varphi(P) = \varphi(\bar{P})$.*

Proof: Let $P, \bar{P} \in [\mathbb{D}^S]^2$ be marginally equivalent profiles. Suppose $\tau(P^1) = a$ and $\tau(P^2) = b$. Since P and $\bar{P}$ are marginally equivalent it follows that $\tau(\bar{P}^1) = a$ and $\tau(\bar{P}^2) = b$. Moreover $a_j P^1_j b_j$, $a_j \bar{P}^1_j b_j$ and $b_j P^2_j a_j$, $b_j \bar{P}^2_j a_j$ for all $j \in M$ whenever a_j and b_j are distinct. According to Lemma 3.9, the support of the lotteries $\varphi(P)$ and

$\varphi(\bar{P})$ are the same and equal to the set $\{a_1, b_1\} \times \ldots \times \{a_m, b_m\}$. We will show that these lotteries are in fact, equal to each other. We prove this by induction on the number of components.

The result from the Gibbard random dictatorship result in the case where $m = 1$. Assume now that the following is true.

Induction Hypothesis (IH): Suppose there are $t \geq 2$ components. Let $P, \bar{P} \in [\mathbb{D}^L]^2$ be marginally equivalent profiles. Then $\varphi(P) = \varphi(\bar{P})$.

We will show that Lemma 3.10 holds in the case of $t + 1$ components. We will prove this in two steps.

Claim 2: Suppose that there are $t + 1$ components. Let $P, \bar{P} \in [\mathbb{D}^L]^2$ be two profiles such that there exists a component assumed without loss of generality to be component $t + 1$ and

1. $\tau(P^1_{t+1}) = \tau(P^2_{t+1}) = x_{t+1}$ and $\tau(\bar{P}^1_{t+1}) = \tau(\bar{P}^2_{t+1}) = y_{t+1}$
2. $P^i_k = \bar{P}^i_k$ for $i = 1, 2$ and all components $k = 1, \ldots t$.

Then $\varphi_{(a,x_{t+1})}(P) = \varphi_{(a,y_{t+1})}(\bar{P})$ for all t-component alternatives a.

Let $P_{-(t+1)}$ and $\bar{P}_{-(t+1)}$ denote the profiles of preferences induced over all components other than $t + 1$ by the profiles P and $\bar{P}$ respectively. Observe that $P_{-(t+1)}$ and $\bar{P}_{-(t+1)}$ are marginally equivalent over all components other than $t+1$ by 2 above. Applying Lemma 3.8, we know that there exists a t component strategy-proof RSCF φ' such that

- $[\varphi_{(a,a_{t+1})}(P) > 0] \Rightarrow [a_{t+1} = x_{t+1}]$
- $[\varphi_{(a,a_{t+1})}(\bar{P}) > 0] \Rightarrow [a_{t+1} = y_{t+1}]$
- $\varphi_{(a,x_{t+1})}(P) = \varphi'_a(P_{-(t+1)})$
- $\varphi_{(a,y_{t+1})}(\bar{P}) = \varphi'_a(\bar{P}_{-(t+1)})$

The Induction Hypothesis implies that $\varphi'_a(P_{-(t+1)}) = \varphi'_a(\bar{P}_{-(t+1)})$. Therefore $\varphi_{(a,x_{t+1})}(P) = \varphi_{(a,y_{t+1})}(\bar{P})$. This completes Claim 2.

We now complete the proof of the induction step. In view of Claim 2 the only case that needs to be considered is the one where $\tau(P^1) = a$ and $\tau(P^2) = b$ and $a_j \neq b_j$ for all $j = 1, \ldots t+1$. Suppose that $\varphi(P^1, P^2) \neq \varphi(\bar{P}^1, P^2)$. (Recall that P^1 and $\bar{P}^1$ are marginally equivalent.) There must exist $x, y \in TPS(P) = TPS(\bar{P})$ such that xP^1y and $y\bar{P}^2x$. We claim that there must exist at least two components say j and k such that $x_j \neq y_j$ and $x_k \neq y_k$. Of course, at least one such component is required; otherwise $x = y$. Suppose there exists exactly one such component , say j. Then separability of preference orderings would imply that the marginal preferences over component j have switched between P^1 and $\bar{P}^1$ contradicting our hypothesis that P^1 and $\bar{P}^1$ are marginally equivalent.

From Lemma 3.4 we know that there exist components j and k such that $\varphi_{a_j}(\bar{P}^1, P^2) > \varphi_{a_j}(P)$ and $\varphi_{a_k}(\bar{P}^1, P^2) < \varphi_{a_k}(P)$.

Consider the second ranked alternative x in P^2. There must exist a unique component say l such that $x_l = a_l$ and $x_j = b_j$ for all $j \neq l$. We consider two cases.

Case 1: $j \neq l$. Let $\tilde{P}^1$ be a lexicographic ordering where component j is lexicographically best. If $\varphi_{a_j}(\tilde{P}^1, P^2) < \varphi_{a_j}(\bar{P}^1, P^2)$, then voter 1 will manipulate at $(\tilde{P}^1, P^2)$ via $\bar{P}^1$. Therefore $\varphi_{a_j}(\tilde{P}^1, P^2) \geq \varphi_{a_j}(\bar{P}^1, P^2)$. Let $\hat{P}^1$ be a lexicographic ordering where j is the best and $\tau(\hat{P}^1_l) = b_l$. By strategy-proofness, $\varphi_{a_j}(\hat{P}^1, P_2) = \varphi_{a_j}(\tilde{P}^1, P_2)$. Hence $\varphi_{a_j}(\hat{P}^1, P_2) > \varphi_{a_j}(P)$. Observe that at the profile $(\hat{P}^1, P^2)$, both voters have a common maximal alternative b_l for component l.

Let $\tilde{P}^2$ be a lexicographic ordering where j and l are the best and the worst components respectively. Using the same argument as before, $\varphi_{b_j}(P^1, \tilde{P}^2) \geq \varphi_{b_j}(P)$, i.e. $\varphi_{a_j}(P^1, \tilde{P}^2) \leq \varphi_{a_j}(P)$. Let $\hat{P}^2$ be a lexicographic ordering such that component j and l are the best and worst respectively and $\tau(\hat{P}^2_l) = a_l$. As before, $\varphi_{a_j}(P^1, \tilde{P}^2) = \varphi_{a_j}(P^1, \hat{P}^2)$ so that $\varphi_{a_j}(P^1, \hat{P}^2) \leq \varphi_{a_j}(P)$.

Observe that at the profile $(P^1, \hat{P}^2)$, both voters have a common maximal alternative a_l for component l. By Claim 5 we must have $\varphi_{a_j}(\hat{P}^1, P^2) = \varphi_{a_j}(P^1, \hat{P}^2)$. However, we have shown that $\varphi_{a_j}(\hat{P}^1, P_2) > \varphi_{a_j}(P) \geq \varphi_{a_j}(P^1, \hat{P}^2)$. We have a contradiction.

Case 2: $j = l$. Let $\tilde{P}^2$ be a lexicographic ordering where component k is lexicographically best. Using a similar argument as before we have $\varphi_{a_k}(\bar{P}^1, \tilde{P}^2) \leq \varphi_{a_k}(\bar{P}^1, P^2)$. Let $\hat{P}^2$ be a lexicographic ordering where k is the best and $\tau(\hat{P}^2_l) = a_l$. By strategy-proofness, $\varphi_{a_k}(P^1, \hat{P}^2) = \varphi_{a_k}(P^1, \tilde{P}^2)$. Hence $\varphi_{a_k}(P^1, \hat{P}^2) < \varphi_{a_k}(P)$. Observe that at the profile $(P^1, \hat{P}^2)$, both voters have a common maximal alternative a_l for component l.

Let $\tilde{P}^1$ be a lexicographic ordering where k and l are the best and the worst components respectively. Using the same argument as before, $\varphi_{a_k}(\tilde{P}^1, P^2) \geq \varphi_{a_k}(P)$. Let $\hat{P}^1$ be a lexicographic ordering such that component k and l are the best and worst respectively and $\tau(\hat{P}^1_l) = b_l$. As before, $\varphi_{a_k}(\tilde{P}^1, P^2) = \varphi_{a_k}(\hat{P}^1, P^2)$ so that $\varphi_{a_k}(\hat{P}^1, P^2) \geq \varphi_{a_k}(P)$.

Observe that at the profile $(\hat{P}^1, P^2)$, both voters have a common maximal alternative b_l for component l. By Claim 5 we must have $\varphi_{a_k}(\hat{P}^1, P^2) = \varphi_{a_k}(P^1, \hat{P}^2)$. However, we have shown that $\varphi_{a_k}(\hat{P}^1, P^2) > \varphi_{a_k}(P) \geq \varphi_{a_k}(P^1, \hat{P}^2)$. We have a contradiction.

Lemma 3.11 *Let $j \in M$, $P \in [\mathbb{D}^L]^2$, $\bar{P}^i \in \mathbb{D}^L$ and $(x_j, z_{M-j}), (y_j, z_{M-j}) \in A$ be such that (i) $\tau(P^i_k, A_k) = \tau(\bar{P}^i_k, A_k)$ for all $k \neq j$ (ii) $\tau(P^i_j, A_j) = x_j$ and $\tau(\bar{P}^i_j, A_j) = y_j$ and (iii) $(x_j, z_{M-j}) \in TPS(P)$. Then*

$$\varphi_{(x_j,z_{M-j})}(\bar{P}^i, P^{-i}) + \varphi_{(y_j,z_{M-j})}(\bar{P}^i, P^{-i}) = \varphi_{(x_j,z_{M-j})}(P) + \varphi_{(y_j,z_{M-j})}(P).$$

Moreover $\varphi_{(d_j,z_{M-j})}(\bar{P}^i, P^{-i}) = \varphi_{(d_j,z_{M-j})}(P)$ for all $d_j \in A_j$.

Proof: In view of Lemmas 3.9 and 3.10 we can assume without loss of generality that (i) j is the lexicographically worst component in P^i and $\bar{P}^i$ and (ii) $a_{M-j} P^i_{M-j} b_{M-j} \Leftrightarrow$

$a_{M-j}\bar{P}^i_{M-j}b_{M-j}$. In other words, the lexicographic ordering of all components in P^i and $\bar{P}^i$ are the same and the marginal preferences for each component other than j is the same in P^i and $\bar{P}^i$. It follows then that (x_j, z_{M-j}) and (y_j, z_{M-j}) are contiguous in P^i. Suppose $(a_j, z_{M-j})P^i(x_j, z_{M-j})$ where $a_j \neq y_j$. Since j is the lexicographically worst component and (x_j, z_{M-j}) and (y_j, z_{M-j}) are contiguous it follows that $(a_j, z_{M-j})\bar{P}^i(y_j, z_{M-j})$. Similarly $(a_j, z_{M-j})\bar{P}^i(y_j, z_{M-j}) \Rightarrow (a_j, z_{M-j})P^i(x_j, z_{M-j})$. Now suppose $(a_j, b_{M-j})P^i(x_j, z_{M-j})$ where $b_{M-j} \neq z_{M-j}$. From our assumptions, $b_{M-j}P^i_{M-j}z_{M-j}$. Hence $b_{M-j}\bar{P}^i_{M-j}z_{M-j}$ and $(a_j, b_{M-j})\bar{P}^i(y_j, z_{M-j})$. Similarly, $(a_j, b_{M-j})\bar{P}^i(y_j, z_{M-j})$ implies $(a_j, b_{M-j})P^i(x_j, z_{M-j})$. Hence $\bar{P}^i$ is a feasible local switch of (x_j, z_{M-j}) and (y_j, z_{M-j}). The result now follows from Lemma 3.2.

To show the second part of the Lemma, note that $\varphi_{(d_j, z_{M-j})}(\bar{P}^i, P^{-i}) = \varphi_{(d_j, z_{M-j})}(P) = 0$ if $d_j \neq \tau(P^{-i}, A_j)$. Suppose $d_j = \tau(P^{-i}, A_j)$. Again, using Lemmas 3.9 and 3.10, we can assume that d_j is ranked third in both P^i_j and $\bar{P}^i_j$. This implies that $B((d_j, z_{M-j}), P^i) = B((d_j, z_{M-j}), \bar{P}^i)$. Now strategy-proofness implies that $\varphi_{(d_j, z_{M-j})}(\bar{P}^i, P^{-i}) = \varphi_{(d_j, z_{M-j})}(P)$.

We now complete the proof of Step 2. Let $P \in [\mathbb{D}^L]^2$ be such that $\tau(P^1) = a$ and $\tau(P^2) = b$ where $a_j \neq b_j$ for all $j \in M$. Pick an arbitrary $\underline{i} \in I^m$ and let $\gamma(\underline{i}) = \varphi_x(P)$ where $\chi(x, P) = \underline{i}$. Since the maximal alternatives of the two voters for each component are distinct, there exists a unique $\underline{i} \in I^m$ for every $x \in TPS(P)$ such that $\chi(x, P) = \underline{i}$. Therefore

$$\sum_{\underline{i} \in I^m} \gamma(\underline{i}) = \sum_{x \in TPS(P)} \varphi_x(P) = 1$$

where the second equality follows from the fact that φ satisfies the TPS property (Lemma 3.9).

Now consider $j \in M$ and $\bar{P}^1 \in \mathbb{D}$ such that $\tau(\bar{P}^1_j, A_j) = c_j \neq b_j$ and $\tau(\bar{P}^1_k, A_k) = a_k$ for all $k \neq j$. Let $x \in TPS(P)$. Observe that $[\underline{i} \in \chi((a_j, x_{M-j}), P) \Leftrightarrow [\underline{i} \in \chi((c_j, x_{M-j}), (\bar{P}^i), P^2)]$. Now applying Lemma 3.11 and the fact that $(c_j, x_{M-j}) \notin TPS(P)$ and $(a_j, x_{M-j}) \notin TPS(\bar{P}^1, P^2)$, we conclude that $\varphi_{(a_j,x_{M-j})}(P) = \varphi_{(c_j,x_{M-j})}(\bar{P}^1, P^2)$. Using this and the second part of Lemma 3.11, it follows that $\varphi_x(\bar{P}^1, P^2) = \sum_{\underline{i} \in \chi(x,(\bar{P}^1,P^2))} \gamma(\underline{i})$.

Now consider the case where $\tau(\bar{P}^1_j, A_j) = b_j$. Let $x \in TPS(P)$. Observe that $[\underline{i} \in \chi((a_j, x_{M-j}), P) \cup \chi((b_j, x_{M-j}), P) \Leftrightarrow [\underline{i} \in \chi((b_j, x_{M-j}), (\bar{P}^1, P^2)]$. Now applying Lemma 3.11 and noting the fact that $(a_j, x_{M-j}) \notin TPS(\bar{P}^1, P^2)$, we have

$$\varphi_{(b_j,x_{M-j})}(\bar{P}^1, P^2) = \varphi_{(a_j,x_{M-j})}(P) + \varphi_{(b_j,x_{M-j})}(P).$$

Once again, we have $\varphi_x(\bar{P}^1, P^2) = \sum_{\underline{i} \in \chi(x,(\bar{P}^1,P^2))} \gamma(\underline{i})$ for all $x \in A$. Progressively replacing the maximal alternative of each component in voter 1 and voter 2's preferences and noting that the previous expression holds at all profiles along the sequence, we conclude that the expression holds for all profiles P. This establishes the result.

4. Conclusion

We have generalized the random dictatorship result of Gibbard (1977b) to a multi-dimensional setting where there are two voters and preferences are lexicographically separable. In particular we have shown that strategy-proof random social choice functions satisfying unanimity are generalized random dictatorships. These are induced by a fixed probability distribution on voter sequences of length equal to the number of components. Although the joint distribution on outcomes is not the product of strategy-proof component random social functions, we have shown that the marginal probability distribution on each component at a preference profile depends only on component preferences. Moreover the marginal random social choice functions are in fact, strategy-proof and therefore random dictatorships. An important question for future research is whether the decomposability of the marginal random social choice functions holds more generally, for instance, for "rich domains" as defined in cite LeBreton and Sen (1999).

CHAPTER 5

A foundation for proper rationalizability from an incomplete information perspective

1. Introduction

Epistemic game theory deals with the ways the players may reason about their opponents before making a decision. More precisely, in epistemic game theory players base their choices on the beliefs about the opponents' behavior, which in turn depend on their beliefs about the opponents' beliefs about others' behavior, and so on. A major goal of epistemic game theory is to study such infinite belief hierarchies, to impose reasonable conditions on these, and to investigate their behavioral implications.

A central idea in epistemic game theory is *common belief in rationality* (Tan and da Costa Werlang (1988)), stating that a player believes that his opponents choose rationally, believes that his opponents believe that their opponents choose rationally, and so on. In our view, one of its most natural refinements is the concept of *proper rationalizability* (Schuhmacher (1999) and Asheim (2002), which is based on Myerson's (1978) (Myerson (1978)) notion of proper equilibrium, but without imposing any equilibrium assumption. Proper rationalizability is based on the following two conditions: The first states that players are cautious, meaning that they do not exclude any opponents' choice from consideration. The second condition is an extension of Myerson's ϵ-proper trembling condition, which states that whenever you believe that a choice a is better than another choice b for your opponent, then the probability you assign to b must be at most ϵ times the probability you assign to a. Under ϵ-proper rationalizibility there is common belief in the event that every player is cautious and satisfies the ϵ-proper trembling condition. A choice is called properly rationalizable if it can be chosen under ϵ-proper rationalizability for every $\epsilon > 0$.

We will now explain this concept by means of an example. Consider the game in Figure 1, where player 1 chooses between a, b and c and player 2 chooses between d, e and f. Note that for player 2, choice d is better than choice e, and choice e is better than choice f. Hence, under proper rationalizability player 1 deems d for player 2 much more likely than e, and e much more likely than f. Consequently, only choice c will be optimal for player 1. So, if $\epsilon > 0$ is small enough, then only choices

	d	e	f
a	$0,2$	$1,1$	$1,0$
b	$1,2$	$0,1$	$1,0$
c	$1,2$	$1,1$	$0,0$

Figure 1: An example for proper rationalizability

c and d can rationally be made under ϵ-proper rationalizability. As such, only the choices c for player 1 and d for player 2 are properly rationalizable.

The usual interpretation of proper rationalizability is that you assume that your opponent makes mistakes, but that you deem more costly mistakes much less likely than less costly mistakes. In this chapter we offer a rather different foundation for proper rationalizability. Instead of assuming that you believe your opponent to make mistakes, we rather suppose that you have uncertainty about his utility function, while believing that he chooses rationally. We thus consider a game with incomplete information. Our main result states that, if we let the uncertainty about the opponent's utility go to zero in some regular manner, then every choice that can rationally be made under common belief in rationality in the game with incomplete information, will be properly rationalizable in the original game, in which there is no uncertainty about the opponent's utilities.

In the game with incomplete information, we impose some regularity conditions on the players' beliefs about the opponent's utility functions which can be summarized as follows: First, for every outcome in the game, the belief that player i has about player j's utility from this outcome, is always normally distributed with its mean at the "original" utility in the original game. As a consequence, player i deems any utility function possible for player j, and hence every choice for player j can be optimal for some utility function deemed possible by i. Together with the condition that i believes in j's rationality, this actually makes sure that player i deems every choice possible for player j, thus mimicking the cautiousness condition described above. Secondly,

i's belief about j's utility function should be independent from his belief about j's belief hierarchy. This makes intuitive sense since j's belief hierarchy is an epistemic property of this player, whereas his utility function is not. So there is no obvious reason to expect any correlation between these two characteristics. Thirdly, i's belief about j's utilities from different outcomes in the game should be independent from each other. Possibly some of these conditions can be relaxed for the proof of our main result, but we leave this issue for future research.

Our game with incomplete information is related to the one used in Dekel and Fudenberg (1990). They also consider games with incomplete information where the player's uncertainty about the opponent's utilities goes to zero. An important difference with our approach is that Dekel and Fudenberg apply the concept of *iterated elimination of weakly dominated choices* to the games with incomplete information. They show that if the uncertainty about the opponent's utilities vanishes, then we obtain one round of deletion of weakly dominated strategies, followed by iterated deletion of strongly dominated strategies, in the original game. The latter procedure is also called the Dekel-Fudenberg procedure in the literature. In contrast, we apply common belief in rationality to our games with incomplete information. We then show that if the uncertainty about the opponent's utilities vanishes, we obtain a subselection (that is some, but in general not all) of the properly rationalizable choices in the original game, which is fundamentally different from the Dekel-Fudenberg procedure. Another fundamental difference between our work and Dekel and Fudenberg lies in the way the uncertainty about the opponent's utilities is modeled. Their model assumes that players only deem possible finitely many utility functions for the opponent, and that a large probability must be assigned to the opponent's "original" utility function. In contrast, we assume that the uncertainty about the opponent's utilities is given by a normal distribution. In particular, players deem *every* utility function possible for the opponent.

The chapter is organized as follows: In Section 2 we introduce our epistemic model for games with incomplete information, we formalize the idea of common belief in rationality for these games, and show that common belief in rationality is always possible. In Section 3 we introduce our epistemic model for games with complete information, and present the concept of proper rationalizability for these games. In Section 4 we state our main result, establishing the connection between common belief in rationality in the game with incomplete information (in the presence of small uncertainty about the opponent's utility function), and proper rationalizability in the original game. In Section 5 we provide some concluding remarks. All proofs are collected in Section 6.

2. Rationalizability in Games with Incomplete Information

2.1. *Epistemic Model*

Throughout this paper we restrict attention to games with two players. Let $\Gamma = (C_i, u_i)_{i\in I}$ be a finite, static game where $I = \{1, 2\}$ is the set of players, C_i is the finite set of choices of player i, and u_i is player i 's utility function. The function u_i assigns to every pair of choices $(c_{1,}c_2) \in C_1 \times C_2$ a utility $u_i\,(c_1, c_2) \in \mathbb{R}$.

In a game with incomplete information players do not only have uncertainty about the opponent's choices, they also have uncertainty about the opponent's utility function. Hence a belief hierarchy should not only specify what the player believes about the opponent's choice but also what he believes about the opponent's utility function. Not only this, it should also specify what the player believes about the opponent's belief about his own choice and utility function, and so on. A possible way of modeling such belief hierarchies is by means of the following definition.

Epistemic model An epistemic model for Γ with incomplete information is a tuple $M = (T_i, b_i, v_i)_{i\in I}$ where (1) T_i is the set of types for player i, (2) $b_i : T_i \longrightarrow \triangle\,(C_j \times T_j)$ is the belief assignment taking only finitely many different probability distributions on $\triangle\,(C_j \times T_j)$, and (3) v_i is the utility assignment that assigns to every $t_i \in T_i$ a utility function $v_i\,(t_i) : C_1 \times C_2 \longrightarrow \mathbb{R}$.

By $\triangle(X)$ we denote the set of probability distributions on X. So, in an epistemic model, each type t_i has a belief about player j's choice-type combinations. And hence, in particular, it has a belief about j's choice. But, as player j's type also specifies his utility function and his belief about player i's choice, player i also has some belief about player j's utility function, and about player j's belief about his own choice, and so on. In this way one can derive a complete belief hierarchy for every given type.

Note that each type t_i can be indentified with a pair $(v_i(t_i), b_i(t_i))$ where $v_i(t_i)$ is its utility function and $b_i(t_i)$ is its belief hierarchy. Since we required the belief assignment to take only finitely many different probability distributions, the epistemic model contains only finitely many different belief hierarchies.

2.2. *Restrictions on the Epistemic Model*

Our goal will be to model the situation where the players have uncertainty about the opponent's utility function, but where this uncertainty "vanishes in the limit". In order to formalise this we need to impose additional restrictions on the epistemic model.

Recall that every type t_i can be identified with a pair $(v_i(t_i), b_i(t_i))$, where $v_i(t_i)$ is t_i's utility function and $b_i(t_i)$ is its belief hierarchy. Denote by V_i the set of all possible utility functions, and by B_i the set of all belief hierarchies in the epistemic model $M = (T_i, v_i, b_i)_{i \in I}$. The first condition we impose is that $T_i = V_i \times B_i$, that is, for every possible utility function we can think of, and every belief hierarchy in the model, there exists a type in the model with exactly this combination of utility function and belief hierarchy. So in a sense we assume that the type space is rich enough.

Secondly we assume that t_i's belief about j's utility from (c_1, c_2) is statistically independent from its belief about j's utility from (c_1', c_2') whenever $(c_1, c_2) \neq (c_1', c_2')$, and that this belief is also statistically independent from its belief about j's belief hierarchy.

Finally we assume that t_i's beliefs about j's utilities from the various outcomes in the game are all induced by a unique normal distribution. More formally, t_i's belief about j's utility from (c_1, c_2) is given by a normal distribution with its mean at $u_j(c_1, c_2)$ – the "true" utility of player j in the original game. So, all these beliefs are distributed identically around the mean. By collecting all these conditions we arrive at the following definition.

σ-regular epistemic model Let P be the normal distribution on $\mathbb{R}$ with mean 0 and variance $\sigma^2 > 0$. Then an epistemic model $M = (T_i, b_i, v_i)_{i \in I}$ is σ-regular if for both players i, (1) $T_i = V_i \times B_i$, (2) for every type $t_i \in T_i$, his belief about j's utility from (c_1, c_2) is statistically independent from his belief about j's utility from (c_1', c_2') whenever $(c_1, c_2) \neq (c_1', c_2')$, and his belief about j's utilities is statistically independent from his belief about j's belief hierarchy, and (3) for every type $t_i \in T_i$, and every choice-pair (c_1, c_2), the belief of t_i about j's utility from (c_1, c_2) is given by P, up to a shift of the mean to $u_j(c_1, c_2)$.

2.3. *σ-Rationalizability*

In this subsection we will define common belief in rationality inside an epistemic model with incomplete information. In addition, if we require the epistemic model to be σ-regular for a given normal distribution with mean 0 and variance σ^2, then we obtain the concept of σ-rationalizability.

We first need some more notation. For given type t_i and choice c_i, let $v_i(t_i)(c_i)$ be the expected utility for type t_i from choosing c_i, given his belief $b_i(t_i)$ about the opponent's choice, and given his utility function $v_i(t_i)$.

Rational choice A choice c_i is rational for t_i if $v_i(t_i)(c_i) \geq v_i(t_i)(c_i')$ for all $c_i' \in C_i$.

We will now define common belief in rationality. In words it says that a player believes that his opponent makes rational choices, and believes that his opponent believes that he makes rational choices, and so on.

Formally, for every $\widetilde{T}_i \subseteq T_i$, let

$$\left(C_i \times \widetilde{T}_i\right)^{rat} = \{(c_i, t_i) \in C_i \times \widetilde{T}_i : c_i \text{ is rational for } t_i\}.$$

Common belief in rationality For both players i we define subsets of types $T_i^1, T_i^2, ...$ in a recursive way as follows:

$$\begin{aligned}
T_i^1 &:= \{t_i \in T_i : b_i(t_i)[(C_j \times T_j)^{rat}] = 1\}, \\
T_i^2 &:= \{t_i \in T_i : b_i(t_i)[(C_j \times T_j^1)^{rat}] = 1\}, \\
&\vdots \\
T_i^l &:= \{t_i \in T_i : b_i(t_i)[(C_j \times T_j^{l-1})^{rat}] = 1\}, \\
&\vdots
\end{aligned}$$

Type t_i expresses common belief in rationality if $t_i \in \cap_{l \in \mathbb{N}} T_i^l$.

A type t_i is σ-rationalizable if it expresses common belief in rationality within a σ-regular epistemic model.

σ-rationalizable type Let $M = (T_i, b_i, v_i)_{i \in I}$ be a σ-regular epistemic model. Every type $t_i \in T_i$ that expresses common belief in rationality is called σ-rationalizable.

Now we show that σ-rationalizable types always exist.

Theorem 2.1 (σ-rationalizable types always exist) *Consider a finite static game $\Gamma = (C_i, u_i)_{i \in I}$, and some $\sigma > 0$. Then there is a σ-regular epistemic model $M = (T_i, b_i, v_i)_{i \in I}$ for Γ where all types are σ-rationalizable.*

The proof can be found in Section 6.

2.4. *Limit Rationalizability*

In this subsection we focus on those choices which can rationally be made under common belief in rationality when the uncertainty about the opponent's utility vanishes. This will lead to the concept of limit rationalizability. We first need an additional definition.

Constant type spaces and utility assignments A sequence of epistemic models $((T_i^n, b_i^n, v_i^n)_{i\in I})_{n\in\mathbb{N}}$ has constant type spaces and utility assignments if $T_i^n = T_i^m$ and $v_i^n = v_i^m$ for all n and m, and for both players i.

We are now ready to define the concept of limit rationalizable choice.

Limit rationalizable choice Consider a finite static game $\Gamma = (C_i, u_i)_{i\in I}$ with two players. A choice c_i is limit rationalizable if there is a sequence $(\sigma_n)_{n\in\mathbb{N}} \to 0$, and a sequence $(M^n)_{n\in\mathbb{N}}$ of σ_n-regular epistemic models with constant type spaces and utility assignments, such that in every M^n there is a σ_n-rationalizable type t_i^n with utility function u_i, for which choice c_i is optimal.

3. Proper Rationalizability in Games with Complete Information

3.1. *Epistemic Model*

Let $\Gamma = (C_i, u_i)_{i\in I}$ be a finite, static game with two players. In a game with complete information players do not have uncertainty about the opponent's utility function. Therefore a belief hierarchy only needs to specify what a player believes about the opponent's choice, what he believes about the opponent's belief about his own choice, and so on. Therefore the epistemic model will be simpler compared to the case of incomplete information.

Epistemic model An epistemic model for Γ with complete information is a tuple $M = (\Theta_i, \beta_i)_{i\in I}$ where (1) Θ_i is the finite set of types for player i, and (2) $\beta_i : \Theta_i \longrightarrow \triangle(C_j \times \Theta_j)$ is the belief assignment.

So, in an epistemic model, each type θ_i has a belief about player j's choice-type combinations. And hence, in particular, it has a belief about j's choice. But, as player j's type also specifies his belief about player i's choice, player i also has some belief about player j's belief about his own choice, and so on. In this way one can derive a complete belief hierarchy for every given type.

For given type θ_i and choice c_i we define $u_i(c_i, \theta_i)$ as the expected utility for type θ_i from choosing c_i given his belief $\beta_i(\theta_i)$ about his opponent's choice (and given his "fixed" utility function u_i). Type θ_i is said to *prefer* choice c_i to choice c_i' when $u_i(c_i, \theta_i) > u_i(c_i', \theta_i)$. We say that a type θ_i *considers possible* some opponent's type θ_j if $\beta_i(\theta_i)(c_j, \theta_j) > 0$ for some $c_j \in C_j$. Now we introduce the key condition in proper rationalizability, which is the ϵ-proper trembling condition. Intuitively it says that (1) a player should deem possible all opponent's choices, and (2) if a player believes choice a is better than choice b for the other player, then he should deem choice a much more likely than choice b.

ϵ-proper trembling condition Let $\epsilon > 0$. A type θ_i satifies the ϵ-proper trembling condition if

(1) for each θ_j that θ_i deems possible, $\beta_i(\theta_i)(c_j, \theta_j) > 0$ for all $c_j \in C_j$, and

(2) for every θ_j that θ_i deems possible, whenever θ_j prefers c_j to c_j', then $\beta_i(\theta_i)\left(c_j', \theta_j\right) \leq \epsilon \cdot \beta_i(\theta_i)(c_j, \theta_j)$.

So, the first condition says that whenever θ_i deems some type θ_j possible, θ_i also assumes every choice is possible for θ_j.

Proper rationalizability is based on the event that the types should not only satisfy the ϵ-proper trembling condition themselves, but also express common belief in the event that types satisfy the ϵ-proper trembling condition.

ϵ-properly rationalizable type A type θ_i is ϵ-properly rationalizable if:

θ_i satisfies the ϵ-proper trembling condition,

θ_i only deems possible opponent's types θ_j which satisfy the ϵ-proper trembling condition,

θ_i only deems possible opponent's types θ_j which only deem possible player i's types θ_i' which satisfy the ϵ-proper trembling condition, and so on.

Properly rationalizable choices are those choices which can rationally be made by ϵ-properly rationalizable types, for all ϵ.

Properly rationalizable choice A choice c_i is ϵ-properly rationalizable if there is an epistemic model and an ϵ-properly rationalizable type θ_i within it for which c_i is optimal. A choice c_i is properly rationalizable if it is ϵ-properly rationalizable for all $\epsilon > 0$.

3.2. *Example*

Consider again the game in Figure 1. Let the type sets of player 1 and player 2 be $\Theta_1 = \{\theta_1, \theta_1'\}$ and $\Theta_2 = \{\theta_2, \theta_2'\}$. For $\epsilon > 0$ (small), let the beliefs for the types be given by

$$\begin{aligned}
\beta_1(\theta_1) &= \left(1-\epsilon^2-\epsilon^3\right)(d,\theta_2) + \epsilon^2 (e,\theta_2) + \epsilon^3 (f,\theta_2), \\
\beta_1(\theta_1') &= \frac{1}{6}(d,\theta_2) + \frac{1}{6}(e,\theta_2) + \frac{1}{6}(f,\theta_2) + \frac{1}{6}(d,\theta_2') + \frac{1}{6}(e,\theta_2') + \frac{1}{6}(f,\theta_2'), \\
\beta_2(\theta_2) &= \left(1-\epsilon^2-\epsilon^3\right)(c,\theta_1) + \epsilon^2 (b,\theta_1) + \epsilon^3 (a,\theta_1), \text{ and} \\
\beta_2(\theta_2') &= \frac{1}{6}(a,\theta_1) + \frac{1}{6}(b,\theta_1) + \frac{1}{6}(c,\theta_1) + \frac{1}{6}(a,\theta_1') + \frac{1}{6}(b,\theta_1') + \frac{1}{6}(c,\theta_1').
\end{aligned}$$

It may be verified that the types θ_1 and θ_2 both satisfy the ϵ-proper trembling condition. Also, type θ_1 only deems possible the opponent's type θ_2, and θ_2 only deems possible the opponent's type θ_1. This implies that both θ_1 and θ_2 are ϵ-properly rationalizable. So, choice c for player 1, and d for player 2 are ϵ-properly rationalizable for any $\epsilon > 0$ small enough. Hence, choice c for player 1, and d for player 2 are properly rationalizable.

On the other hand, we see that the type θ_1' of player 1 believes that the choices d, e and f are equally likely to be taken by type θ_2 of player 2 while for type θ_2, d is better than e, and e is better than f. So, type θ_1' of player 1 does not satisfy the ϵ-proper trembling condition. Similarly, type θ_2' also does not satisfies the ϵ-proper trembling condition.

4. Main Result

4.1. *Statement of the Main Result*

For a static game we analysed two contexts, one with incomplete information and another with complete information. In the context with incomplete information, where players have uncertainty about the opponent's utility, we introduced the concept of a limit rationalizable choice. In the context with complete information, where players have no uncertainty about the opponent's utility, we discussed the concept of a properly rationalizable choice. In our main result we connect these two concepts.

Theorem 4.1 (Limit rationalizability implies proper rationalizability) *Consider a finite static game with two players. Every limit rationalizable choice for the context with incomplete information is a properly rationalizable choice for the context with complete information.*

4.2. *Illustration of the Main Result*

By means of an example we provide some intuition for our main result. More precisely we show how a σ-rationalizable type in the context of incomplete information can be transformed into an ϵ-properly rationalizable type in the context of complete information. Also we show that when σ goes to zero then ϵ goes to zero as well.

Consider again the game from Figure 1. Let us start with the context of incomplete information. Let P be the normal distribution with mean 0 and variance σ^2. From the proof of Theorem 2.1 we know that there exists a σ-regular epistemic model $M = (T_i, b_i, v_i)_{i \in I}$ where every type is σ-rationalizable and all the types have the same belief hierarchy. So, types only differ by their utility function. For each of the types t_1 of player 1 we denote by β_1 the belief about player 2's choice, and for each type t_2 let β_2 be the belief about player 1's choice. As we assume that all the types have the same belief hierarchy, β_1 and β_2 are unique.

For both players i let Q_i be the probability distribution on player i's utility functions generated by P. Since the epistemic model is σ-regular every type t_j has the belief Q_i about i's utility function. Let $V_i(c_i, \beta_i)$ be the set of utility functions for player i such that choice c_i is optimal under the belief β_i about the opponent's choice. Since every type t_i expresses common belief in rationality, the probability it assigns to an opponent's choice c_j is exactly the probability it assigns to the event that j's utility function is in $V_j(c_j, \beta_j)$, which is $Q_j(V_j(c_j, \beta_j))$. So, we can derive the following six equations:

$$\begin{aligned}
\beta_1(d) &= Q_2(V_2(d, \beta_2)) \\
\beta_1(e) &= Q_2(V_2(e, \beta_2)) \\
\beta_1(f) &= Q_2(V_2(f, \beta_2)) \\
\beta_2(a) &= Q_1(V_1(a, \beta_1)) \\
\beta_2(b) &= Q_1(V_1(b, \beta_1)) \\
\beta_2(c) &= Q_1(V_1(c, \beta_1)).
\end{aligned}$$

Since P has full support on $\mathbb{R}$, it follows that all these probabilities are positive.

Now we turn to the context of complete information. We construct an epistemic model with a single type θ_1 for player 1 and a single type θ_2 for player 2. Let the belief of θ_1 about player 2's choice be given by the β_1 constructed above, and similarly for the belief of θ_2. So, the belief about the opponent's choice has not changed by moving from the context with incomplete information to the context with complete information.

Since in the original game d is better than e and e is better than f for player 2, for small σ we will have that $Q_2(V_2(d, \beta_2))$ is much bigger than $Q_2(V_2(e, \beta_2))$, and $Q_2(V_2(e, \beta_2))$ is much bigger than $Q_2(V_2(f, \beta_2))$. So, by our equations above we have that $\beta_1(d)$ is much bigger than $\beta_1(e)$, and $\beta_1(e)$ is much bigger than $\beta_1(f)$. Given such a β_1, in the original game c will be better than b and b will be better than a. So, similarly, for small σ we will have that $Q_1(V_1(c, \beta_1))$ is much bigger than $Q_1(V_1(b, \beta_1))$, and $Q_1(V_1(b, \beta_1))$ is much bigger than $Q_1(V_1(a, \beta_1))$. And hence, from the equations above, we have that $\beta_2(c)$ is much bigger than $\beta_2(b)$, and $\beta_2(b)$ is much bigger than $\beta_2(a)$. Now define

$$\epsilon = \max\{\frac{\beta_2(a)}{\beta_2(b)}, \frac{\beta_2(b)}{\beta_2(c)}, \frac{\beta_1(e)}{\beta_1(d)}, \frac{\beta_1(f)}{\beta_1(e)}\}.$$

Then, by construction, θ_1 and θ_2 are ϵ-properly rationalizable. Moreover, if σ goes to zero then the associated ϵ would go to zero as well.

If the variance of P is small then choice c is optimal for the σ-rationalizable type t_1 in the model with incomplete information that has the original utility function. Similarly, d is optimal for the σ-rationalizable type t_2 that has the original utility function in the model with incomplete information. As a consequence, c and d are limit rationalizable in the context with incomplete information. On the other hand, in the associated epistemic model with complete information c is optimal for the ϵ-properly rationalizable type θ_1 and d is optimal for the ϵ-properly rationalizable type θ_2. As ϵ goes to zero when σ goes to zero, we conclude that c and d are properly rationalizable. So, in this example the limit rationalizable choices are also properly rationalizable.

5. Concluding remarks

We believe that proper rationalizability is a very natural concept in game theory, but it has not yet received the attention it deserves. In this paper we have established a new foundation for proper rationalizability from the viewpoint of games with incomplete information. In games with incomplete information we define a choice as limit rationalizable if it can rationally be made under common belief of rationality when the uncertainty vanishes gradually in some regular way. We show the existence of such choices. We then prove that each limit rationalizable choice in the game with incomplete information is properly rationalizable for the context with complete information.

Throughout this paper it is assumed that the players' uncertainty about the opponent's utilities are described by a normal distribution. We have used the normal distribution as it is a very natural candidate to describe the uncertainty. We believe, however, that we can extend our framework to wider classes of probability distributions here, as long as this class is closed under taking convex combinations, and Lemma 6.4 is satisfied.

In this paper we restricted our attention to two players for the sake of simplicity. However, we believe our result can be extended to more than two players in a natural way.

6. Proofs

6.1. *Existence of σ-Rationalizable Types*

We prove Theorem 2.1, which guarantees the existence of σ-rationalizable types. Consider a finite static game $\Gamma = (C_i, u_i)_{i\in I}$, and some $\sigma > 0$. Let P be the normal distribution with mean 0 and variance σ^2. In fact we will construct a σ-regular epistemic model where all types of player 1 have the same belief β_2 about player 2's choice and all types of player 2 have the same belief β_1 about player 1's choice. We construct β_1 and β_2 by means of the fixed point of some correspondence.

For every belief $\beta_j \in \Delta(C_j)$ and every utility function w_i, we define

$$C_i(\beta_j, w_i) := \{c_i \in C_i : w_i(c_i, \beta_j) \geq w_i(c_i', \beta_j) \text{ for all } c_i'\}.$$

We also define Q_i as the probability distribution on the set of utility functions of player i induced by P. For every $\beta_j \in \Delta(C_j)$ we define

$$\begin{aligned} F_i(\beta_j) \quad &: \quad = \{\beta_i \in \Delta(C_i) : \beta_i = \int_{w_i \in V_i} \gamma_i(w_i)\, dQ_i \,, \\ \text{where } \gamma_i(w_i) \quad &\in \quad \Delta(C_i(\beta_j, w_i)) \text{ for every } w_i \in V_i\}. \end{aligned}$$

Here V_i denotes the set of all possible utility functions for player i. So every $\beta_i \in F_i(\beta_j)$ is obtained by taking for every utility function w_i a randomization over optimal choices against β_j and then taking the expected randomization with respect to Q_i. Now we define a correspondence F from $\Delta(C_1) \times \Delta(C_2)$ to $\Delta(C_1) \times \Delta(C_2)$ by

$$F(\beta_1, \beta_2) := F_1(\beta_2) \times F_2(\beta_1).$$

Now we use Kakutani's fixed point theorem to prove that F has a fixed point. Clearly F is upper hemi-continuous and compact valued. We show that F is convex valued. For this it is sufficient to show that F_1 and F_2 are convex valued. For a given β_2, take β_1', β_1'' in $F_1(\beta_2)$. We show that $\lambda\beta_1' + (1-\lambda)\beta_1''$ is also in $F_1(\beta_2)$. By definition

$$\beta_1' = \int_{w_1} \gamma_1'(w_1)\, dQ_1 \text{ and } \beta_1'' = \int_{w_1} \gamma_1''(w_1)\, dQ_1$$

where $\gamma_1'(w_1), \gamma_1''(w_1) \in \Delta(C_1(\beta_2, w_1))$ for every w_1. So we have

$$\lambda\beta_1' + (1-\lambda)\beta_1'' = \int_{w_1} (\lambda\gamma_1'(w_1) + (1-\lambda)\gamma_1''(w_1))dQ_1$$

where $\lambda\gamma_1'(w_1) + (1-\lambda)\gamma_1''(w_1) \in \Delta(C_1(\beta_2, w_1))$ for every w_1. Hence by definition $\lambda\beta_1' + (1-\lambda)\beta_1'' \in F_1(\beta_2)$. This implies that F_1 is convex valued. The same applies to F_2 and hence we can conclude that F is convex valued. Now using Kakutani's fixed point theorem F has a fixed point (β_1^*, β_2^*).

Since $\beta_1^* \in F_1(\beta_2^*)$ it follows that

$$\beta_1^* = \int_{w_1} \gamma_1^*(w_1)\, dQ_1$$

where $\gamma_1^*(w_1) \in \Delta(C_1(\beta_2^*, w_1))$ for every w_1. Similarly

$$\beta_2^* = \int_{w_2} \gamma_2^*(w_2)\, dQ_2$$

where $\gamma_2^*(w_2) \in \Delta(C_2(\beta_1^*, w_2))$ for every w_2.

We will now construct an epistemic model $M = (T_i, b_i, v_i)_{i\in I}$. For both players i, define

$$T_i = \{t_i^{w_i} \ : \ w_i \in V_i\}.$$

Let the utility assignment v_i be given by

$$v_i(t_i^{w_i}) = w_i$$

for every $t_i^{w_i} \in T_i$. In order to define the belief assignment b_i we first define for every type $t_i^{w_i}$ a density function $\tilde{b}_i(t_i^{w_i})$ on $C_j \times T_j$ as follows:

$$\tilde{b}_i(t_i^{w_i})\left(c_j, t_j^{w_j}\right) := \gamma_j^*(w_j)(c_j),$$

where $\gamma_j^*(w_j)(c_j)$ is the probability that probability distribution $\gamma_j^*(w_j)$ assigns to c_j. For every type $t_i^{w_i}$ let $b_i(t_i^{w_i}) \in \Delta(C_j \times T_j)$ be the probability distribution induced by density function $\tilde{b}_i(t_i^{w_i})\left(c_j, t_j^{w_j}\right)$ and the probability distribution Q_j on V_j. That is, for every set of types $E \subseteq T_j$ given by

$$E := \{t_j^{w_j} : w_j \in F\}$$

we have that

$$b_i\left(t_i^{w_i}\right)\left(\{c_j\}\times E\right) := \int_{w_j\in F} \tilde{b}_i\left(t_i^{w_i}\right)\left(c_j, t_j^{w_j}\right) dQ_j.$$

It follows that the belief of type $t_i^{w_i}$ about player j's choice is given by β_j^*. Namely, the probability that type $t_i^{w_i}$ assigns to choice c_j is equal to

$$\begin{aligned} b_i\left(t_i^{w_i}\right)\left(\{c_j\}\times V_j\right) &= \int_{w_j\in V_j} \tilde{b}_i\left(t_i^{w_i}\right)\left(c_j, t_j^{w_j}\right) dQ_j \\ &= \int_{w_j\in V_j} \gamma_j^*\left(w_j\right)\left(c_j\right) dQ_j \\ &= \beta_j^*\left(c_j\right). \end{aligned}$$

So all types of player i have the same belief β_j^* about player j's choice. This completes the construction of the epistemic model. It follows directly from the construction that the epistemic model is σ-regular.

We now show that every type in this model expresses common belief in rationality. For this it is sufficient to show that every type $t_i^{w_i}$ believes in the opponent's rationality. So, we must show for both players i and every $t_i^{w_i}\in T_i$ that $b_i\left(t_i^{w_i}\right)\left[\left(C_j\times T_j\right)^{rat}\right]=1$. In order to prove so we show that $\tilde{b}_i\left(t_i^{w_i}\right)\left(c_j, t_j^{w_j}\right)>0$ only if c_j is rational for $t_j^{w_j}$.

Suppose that $\tilde{b}_i\left(t_i^{w_i}\right)\left(c_j, t_j^{w_j}\right) > 0$. Since $\tilde{b}_i\left(t_i^{w_i}\right)\left(c_j, t_j^{w_j}\right) := \gamma_j^*\left(w_j\right)\left(c_j\right)$, it follows that $\gamma_j^*\left(w_j\right)\left(c_j\right) > 0$. As by definition $\gamma_j^*\left(w_j\right) \in \Delta\left(C_j\left(\beta_i^*, w_j\right)\right)$ it follows that $c_j \in C_j\left(\beta_i^*, w_j\right)$. Remember that the belief of type $t_j^{w_j}$ about player i's choice is exactly β_i^*. Since $c_j \in C_j\left(\beta_i^*, w_j\right)$ it follows that c_j is rational for type $t_j^{w_j}$. So we have shown that $\tilde{b}_i\left(t_i^{w_i}\right)\left(c_j, t_j^{w_j}\right)>0$ only if c_j is rational for $t_j^{w_j}$. This implies that type $t_i^{w_i}$ believes in the opponent's rationality. Since this holds for every type in the model it follows that every type in the epistemic model expresses common belief in rationality. So every type in the model is σ-rationalizable because the model is σ-regular. This completes the proof. ■

6.2. *Some Technical Lemmas*

In this subsection we state some technical lemmas which we need for the proof of the main result.

Lemma 6.1 *If X, Y and Z are real valued, independent random variables then* $\Pr\left(X \geq \max\{Y, Z\}\right) \geq \Pr\left(X \geq Y\right) \cdot \Pr\left(X \geq Z\right).$

Proof. Let f_Y and f_Z be the probability density functions of the random variables Y and Z. Now,

$$\begin{aligned}
& \Pr\left(X \geq \max\{Y, Z\}\right) \\
= & \int_y \int_z \Pr\left(X \geq \max\left\{y, z\right\}\right) df_Y\left(y\right) df_Z\left(z\right) \\
\geq & \int_y \int_z \Pr\left(X \geq \max\left\{y, z\right\}\right) \cdot \Pr(X \geq \min\left\{y, z\right\}) df_Y\left(y\right) df_Z\left(z\right) \\
= & \int_y \int_z \Pr\left(X \geq y\right) \cdot \Pr\left(X \geq z\right) df_Y\left(y\right) df_Z\left(z\right) \\
= & \int_y \Pr\left(X \geq y\right) df_Y\left(y\right) \cdot \int_z \Pr\left(X \geq z\right) df_Z\left(z\right) \\
= & \Pr\left(X \geq Y\right) \cdot \Pr\left(X \geq Z\right).
\end{aligned}$$

Note that the first and third equality follow from the fact that Y and Z are independent, and the inequality holds because $\Pr(X \geq \min\left\{y, z\right\}) \leq 1$. ■

We now state the well-known Chebyshev's inequality, which we use in the proof of Lemma 6.3.

Lemma 6.2 (Chebyshev's inequality) *Let X be a random variable with $E\left(X\right) = \mu$. Then for any number $k > 0$,*

$$\Pr\left(\left|X - \mu\right| \geq k\right) \leq \frac{Var\left(X\right)}{k^2}.$$

Lemma 6.3 *For every $n \in \mathbb{N}$, let $X_n^1, X_n^2, ..., X_n^m$ be independent random variables with $E\left(X_n^i\right) = \mu^i$ for all n and i, $\mu^1 > \mu^2 > ... > \mu^m$, and $\lim_{n\to\infty} Var\left(X_n^i\right) = 0$ for all i. Then,*

$$\lim_{n\to\infty} \Pr\left(X_n^1 \geq X_n^2 \geq ... \geq X_n^m\right) = 1.$$

Proof. For a given n,

$$\Pr\left(X_n^1 \geq X_n^2 \geq ... \geq X_n^m\right) \geq 1 - \Pr\left(X_n^i < X_n^j \text{ for some } i < j\right).$$

For fixed $i < j$ we have,

$$\begin{aligned}
\Pr\left(X_n^i < X_n^j\right) &= \Pr\left(X_n^j - X_n^i > 0\right) = \Pr\left(\left(X_n^j - X_n^i\right) - \left(\mu^j - \mu^i\right) > \mu^i - \mu^j\right) \\
&\leq \Pr\left(\left|\left(X_n^j - X_n^i\right) - \left(\mu^j - \mu^i\right)\right| > \mu^i - \mu^j\right) \\
&\leq \frac{Var\left(X_n^j - X_n^i\right)}{\left(\mu^i - \mu^j\right)^2} \\
&= \frac{Var\left(X_n^j\right) + Var\left(X_n^i\right)}{\left(\mu^i - \mu^j\right)^2}.
\end{aligned}$$

Here, the inequality comes from Chebyshev's inequality and the last equality follows from the fact that X_n^j and X_n^i are independent. Now, note that $\lim_{n\to\infty} Var\left(X_n^i\right) = 0$ and $\lim_{n\to\infty} Var\left(X_n^j\right) = 0$, which implies $\lim_{n\to\infty} \Pr\left(X_n^i < X_n^j\right) = 0$. Then, from above it follows that

$$\lim_{n\to\infty} \Pr\left(X_n^1 \geq X_n^2 \geq ... \geq X_n^m\right) = 1. \blacksquare$$

Consider a sequence $(P_n)_{n\in\mathbb{N}}$ of normal distributions with mean 0 and variance σ_n^2 such that $\sigma_n \to 0$ as $n \to \infty$. The density function f_n of P_n is given by

$$f_n(x) = \frac{1}{\sigma_n\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma_n^2}} \text{ for all } x.$$

We show that for large n the right tail of P_n becomes arbitrarily steep everywhere.

Lemma 6.4 *Consider a sequence $(P_n)_{n\in\mathbb{N}}$ of normal distributions with mean 0 and variance σ_n^2, such that $\sigma_n \to 0$ as $n \to \infty$. Let f_n be the density functions of these distributions. Then for all $c > 0$ and $\epsilon > 0$ there is $N \in \mathbb{N}$ such that $\frac{f_n(x+c)}{f_n(x)} \leq \epsilon$ for all $n \geq N$ and all $x > 0$.*

Proof. Take $c > 0$ and $\epsilon > 0$. Then

$$\frac{f_n(x+c)}{f_n(x)} = \frac{e^{-\frac{(x+c)^2}{2\sigma_n^2}}}{e^{-\frac{x^2}{2\sigma_n^2}}} = e^{-\frac{1}{2\sigma_n^2}\left((x+c)^2 - x^2\right)} = e^{-\frac{1}{2\sigma_n^2}\left(2cx + c^2\right)} \leq e^{-\frac{c^2}{2\sigma_n^2}}.$$

Now as $c > 0$ is fixed and $\sigma_n \to 0$ as $n \to \infty$, we can find N large enough such that $e^{-\frac{c^2}{2\sigma_n^2}} \leq \epsilon$ for $n \geq N$. $\blacksquare$

Lemma 6.5 *Consider a sequence* $(X_n)_{n\in\mathbb{N}}$ *of normally distributed random variables such that* $E(X_n) = 0$ *for all* n*, and* $var(X_n) \to 0$ *as* $n \to \infty$*. Let* f_n *be the density functions of these random variables. Then, for every* $0 < x < y$ *it holds that*

$$\lim_{n\to\infty} \frac{\Pr(X_n \geq y)}{\Pr(X_n \geq x)} = 0.$$

Proof. Fix $0 < x < y$, and fix an $\epsilon > 0$. Then, by Lemma 6.4 there is an N such that $\frac{f_n(z+(y-x))}{f_n(z)} \leq \epsilon$ for all $n \geq N$ and all $z > 0$. Take some $n \geq N$. Then,

$$\begin{aligned} \Pr(X_n \geq y) &= \int_y^\infty f_n(z)\,dz = \int_x^\infty f_n(z+(y-x))\,dz \\ &\leq \epsilon \cdot \int_x^\infty f_n(z)\,dz = \epsilon \cdot \Pr(X_n \geq x). \end{aligned}$$

This implies that $\lim_{n\to\infty} \frac{\Pr(X_n\geq y)}{\Pr(X_n \geq x)} = 0$. ■

6.3. *Proof of the Main Result*

We finally prove or main theorem, which is Theorem 4.1. We proceed by three steps.

In step 1, we show how a σ-regular epistemic model M with incomplete information can be transformed into an epistemic model $\hat{M}$ with complete information. More precisely, we transform every type t_i in M into a type $\theta_i(t_i)$ in $\hat{M}$ which has the same belief about the opponent's choice as t_i.

In step 2, we take a choice c_i^* that is limit rationalizable. So we can find a sequence $(P_n)_{n\in\mathbb{N}}$ of normal distributions with mean 0 and variance σ_n^2, with $\sigma_n^2 \to 0$ as $n \to \infty$, and a sequence $(M^n)_{n\in\mathbb{N}}$ of σ_n-regular epistemic models with constant type spaces and utility assignments, such that in every M^n there is a σ_n-rationalizable type t_i^n with utility function u_i for which choice c_i^* is optimal. We show that the type t_i^n is transformed into a type $\theta_i(t_i^n)$ which is ϵ_n-properly rationalizable for some ϵ_n. Since, for all n, c_i^* is rational for t_i^n, and $\theta_i(t_i^n)$ has the same belief about the opponent's choice and the same utility function as t_i^n, it follows that c_i^* is rational for $\theta_i(t_i^n)$ for all

n. As $\theta_i(t_i^n)$ is ϵ_n-properly rationalizable for every n, it follows that c_i^* is ϵ_n-properly rationalizable for all n.

In step 3, we prove that $\lim_{n\to\infty}\epsilon_n = 0$. Hence, c_i^* is ϵ-properly rationalizable for every $\epsilon > 0$ and therefore properly rationalizable.

Step 1. Take some $\sigma > 0$. Let $M = (T_i, b_i, v_i)_{i\in I}$ be a σ-regular epistemic model for Γ with incomplete information. Now we transform this epistemic model M into an epistemic model $\hat{M} = (\Theta_i, \beta_i)_{i\in I}$ with complete information. Using the fact that M is σ-regular we can write

$$T_i = V_i \times B_i,$$

where V_i is the set of all possible utility functions and B_i is the finite set of belief hierarchies in T_i. Then, for $t_i \in T_i$,

$$b_i(t_i) \in \triangle(C_j \times V_j \times B_j).$$

Now take $\Theta_i = B_i$ and $\Theta_j = B_j$. Clearly, Θ_i and Θ_j are finite sets as B_i and B_j are finite. For every $t_i \in T_i$ define the type $\theta_i(t_i) \in \Theta_i$ by

$$\beta_i(\theta_i(t_i)) := \mathrm{marg}_{C_j\times B_j} b_i(t_i).$$

So,

$$\beta_i(\theta_i(t_i))(c_j, b_j) = b_i(t_i)(V_j \times \{(c_j, b_j)\})$$

for all $(c_j, b_j) \in C_j \times B_j$. Hence,

$$\beta_i(\theta_i(t_i)) \in \triangle(C_j \times B_j) = \triangle(C_j \times \Theta_j).$$

By construction $\theta_i(t_i)$ has the same belief about j's choice as t_i. This completes the construction of the epistemic model $\hat{M} = (\Theta_i, \beta_i)_{i\in I}$.

Step 2. Take a choice c_i^* that is limit rationalizable. Hence, there exists a sequence $(P_n)_{n\in\mathbb{N}}$ of normal distributions with mean 0 and variance σ_n^2, with $\sigma_n^2 \to 0$ as $n \to \infty$, and a sequence $(M^n)_{n\in\mathbb{N}}$ of σ_n-regular epistemic models with constant type spaces and utility assignments, such that in every M^n there is a σ_n-rationalizable type t_i^n with utility function u_i for which choice c_i^* is optimal. Let the constant type spaces in the sequence $(M^n)_{n\in\mathbb{N}}$ of epistemic models be T_i and T_j, and the constant utility assignments be v_i and v_j.

Fix an n. Then, within the epistemic model $M^n = (T_i, b_i^n, v_i)_{i\in I}$ there is a σ_n-rationalizable type $t_i^n \in T_i$ with utility function u_i for which c_i^* is optimal. Since type t_i^n only deems possible j's types which are σ_n-rationalizable, and only deems possible j's types which only deem possible i's types which are σ_n-rationalizable, and so on, we may assume without loss of generality that all the types in M^n are σ_n-rationalizable. Let $\hat{M}^n = (\Theta_i^n, \beta_i^n)_{i\in I}$ be the corresponding epistemic model with complete information, as constructed in step 1.

For every $\theta_i \in \Theta_i^n$, we define a number $\epsilon_n(\theta_i)$ as follows: Let Poss(θ_i) be the set of types in Θ_j that θ_i deems possible. For a given type $\theta_j \in$ Poss(θ_i), suppose that θ_j prefers choice c_j^1 to c_j^2, c_j^2 to c_j^3, and so on. So, we obtain an ordering $\left(c_j^1, c_j^2, c_j^3, ..., c_j^m\right)$ of j's choices. Then define

$$\epsilon_n(\theta_i, \theta_j) = \max_{k\in\{2,3,\ldots,m\}} \frac{\beta_i^n(\theta_i)\left(c_j^k, \theta_j\right)}{\beta_i^n(\theta_i)\left(c_j^{k-1}, \theta_j\right)}.$$

Next we define

$$\epsilon_{i,n} = \max_{\theta_i\in\Theta_i^n, \theta_j\in \text{ Poss}(\theta_i)} \epsilon_n(\theta_i, \theta_j).$$

Finally let

$$\epsilon_n = \max\{\epsilon_{i,n},\ \epsilon_{j,n}\}.$$

Note that by construction every type in $\hat{M}^n$ satisfies the ϵ_n-proper trembling condition, hence every type in $\hat{M}^n$ is ϵ_n-properly rationalizable. In particular $\theta_i(t_i^n)$ is ϵ_n-properly rationalizable.

Step 3. Now we show that $\lim_{n\to\infty} \epsilon_n = 0$. It is sufficient to show that

$$\lim_{n\to\infty} \frac{\beta_i^n(\theta_i)\left(c_j^k, \theta_j\right)}{\beta_i^n(\theta_i)\left(c_j^{k-1}, \theta_j\right)} = 0 \tag{5.1}$$

for every $\theta_i \in \Theta_i^n$, and every $\theta_j \in \mathrm{Poss}(\theta_i)$ and every k. As before, player j's choices are ordered $c_j^1, ..., c_j^m$ such that θ_j prefers choice c_j^1 to c_j^2, c_j^2 to c_j^3, and so on. We assume, without loss of generality, that all preferences are strict.

Fix some $\theta_i \in \Theta_i^n$ and $\theta_j \in \mathrm{Poss}(\theta_i)$. Suppose that $\theta_i = \theta_i(t_i)$ for some $t_i \in T_i$, and that $\theta_j = \theta_j(t_j)$ for some $t_j \in T_j$. Let $\gamma_j \in \Delta(C_i)$ be θ_j's belief about i's choice. As before, let V_j be the set of utility functions for player j. For every $k \in \{1, ..., m\}$, let $X^k : V_j \to \mathbb{R}$ be given by

$$X^k(v_j) := v_j(c_j^k, \gamma_j) = \sum_{c_i \in C_i} \gamma_j(c_i) \cdot v_j(c_j^k, c_i)$$

for every $v_j \in V_j$. So, $X^k(v_j)$ denotes the expected utility for player j induced by choice c_j^k, under the belief γ_j and the utility function v_j. Note that X^k is a random variable, as player i holds a probability distribution on V_j, induced by P_n. The probability distribution of X^k depends on n, and is denoted by $\varphi^{nk}(X^k)$. Note that X^k has a normal distribution with mean

$$E(X^k) = u_j(c_j^k, \gamma_j),$$

and variance

$$Var^n(X^k) = \sum_{c_i \in C_i} (\gamma_j(c_i))^2 \cdot \sigma_n^2. \tag{5.2}$$

In particular, it follows that $\lim_{n\to\infty} Var^n(X^k) = 0$, as $\lim_{n\to\infty} \sigma_n^2 = 0$. Since, by assumption, θ_j strictly prefers c_j^1 to c_j^2, strictly prefers c_j^2 to c_j^3, and so on, we have that $E(X^1) > E(X^2) > ... > E(X^m)$.

Let φ^n be the probability distribution of the random vector $(X^1, ..., X^m)$. Recall that all types in M^n are σ_n-rationalizable, which implies that all types in M^n express common belief in rationality. As such, type $t_i \in T_i$ (which generates θ_i) expresses common belief in rationality. In particular, t_i only assigns positive probability to those choice-type combinations (c_j, t_j) where c_j is optimal for t_j. Now, as $\theta_i = \theta_i(t_i)$ and $\theta_j = \theta_j(t_j)$, we have that $\beta_i^n(\theta_i)\left(c_j^k, \theta_j\right)$ is the probability that c_j^k is optimal for

t_j, and that is $\varphi^n(X^k \geq X^l \text{ for all } l)$. Then,

$$\frac{\beta_i^n(\theta_i)\left(c_j^k, \theta_j\right)}{\beta_i^n(\theta_i)\left(c_j^{k-1}, \theta_j\right)} = \frac{\varphi^n(X^k \geq X^l \text{ for all } l)}{\varphi^n(X^{k-1} \geq X^l \text{ for all } l)}. \tag{5.3}$$

Hence, in order to prove (5.1), we must show that

$$\lim_{n\to\infty} \frac{\varphi^n(X^k \geq X^l \text{ for all } l)}{\varphi^n(X^{k-1} \geq X^l \text{ for all } l)} = 0$$

for all $k \in \{2, ..., m\}$. We distinguish two cases.

Case 1. First we consider the case where $k = 2$. Then we have,

$$\frac{\varphi^n(X^k \geq X^l \text{ for all } l)}{\varphi^n(X^{k-1} \geq X^l \text{ for all } l)} \leq \frac{\varphi^n(X^2 \geq X^1)}{\varphi^n\left(X^1 \geq X^2 \geq X^3 \geq ... \geq X^m\right)}.$$

Recall that $E(X^1) > E(X^2) > ... > E(X^m)$. But then, by Lemma 6.3, $\varphi^n(X^2 \geq X^1) \to 0$, and $\varphi^n\left(X^1 \geq X^2 \geq X^3 \geq ... \geq X^m\right) \to 1$, and hence

$$\frac{\varphi^n(X^2 \geq X^1)}{\varphi^n\left(X^1 \geq X^2 \geq X^3 \geq ... \geq X^m\right)} \to 0,$$

which implies that

$$\frac{\varphi^n(X^k \geq X^l \text{ for all } l)}{\varphi^n(X^{k-1} \geq X^l \text{ for all } l)} \to 0$$

as $n \to \infty$.

Case 2. Now we consider the case where $k > 2$. Let $X^{\max}$ be the random variable given by $X^{\max} := \max_{j \neq k, k-1} X_j$. We have

$$\begin{aligned}
&\frac{\varphi^n(X^k \geq X^l \text{ for all } l)}{\varphi^n(X^{k-1} \geq X^l \text{ for all } l)} \\
= \; & \frac{\varphi^n(\left(X^k \geq X^{k-1}\right) \text{ and } \left(X^k \geq X^{\max}\right))}{\varphi^n((X^{k-1} \geq X^k) \text{ and } (X^{k-1} \geq X^{\max}))} \\
\leq \; & \frac{\varphi^n\left(X^k \geq X^{\max}\right)}{\varphi^n((X^{k-1} \geq X^k) \text{ and } (X^{k-1} \geq X^{\max}))} \\
\leq \; & \text{(by Lemma 6.1)} \frac{\varphi^n\left(X^k \geq X^{\max}\right)}{\varphi^n\left(X^{k-1} \geq X^k\right) \cdot \varphi^n\left(X^{k-1} \geq X^{\max}\right)} \\
= \; & \frac{\varphi^n\left(X^k \geq X^{\max}\right)}{\varphi^n\left(X^{k-1} \geq X^{\max}\right)} \cdot \frac{1}{\varphi^n\left(X^{k-1} \geq X^k\right)} \\
= \; & \frac{\varphi^n\left(X^k \geq X^{\max}\right)}{\varphi^n\left(X^k \geq X^{\max} - (E(X^{k-1}) - E(X^k)\right)} \cdot \frac{1}{\varphi^n\left(X^{k-1} \geq X^k\right)},
\end{aligned}$$

where the last equality follows from the observation that $X^{k-1} - E(X^{k-1})$ and $X^k - E(X^k)$ have the same distribution.

Now, from Lemma 6.3 it follows that $\varphi^n\left(X^{k-1} \geq X^k\right) \to 1$ as $n \to \infty$. We show that

$$\frac{\varphi^n\left(X^k \geq X^{\max}\right)}{\varphi^n\left(X^k \geq X^{\max} - (E(X^{k-1}) - E(X^k)\right)} \to 0$$

as $n \to \infty$.

Let us define $c := E(X^{k-1}) - E(X^k)$. So, we have to show that

$$\frac{\varphi^n\left(X^k \geq X^{\max}\right)}{\varphi^n\left(X^k \geq X^{\max} - c\right)} \to 0 \tag{5.4}$$

as $n \to \infty$. Note that $\varphi^n\left(X^k \geq X^{\max}\right) \leq \varphi^n\left(X^k \geq X^1\right)$. We first show that there exists $N \in \mathbb{N}$ such that for all $n \geq N$,

$$\varphi^n\left(X^k \geq X^{\max} - c\right) \geq \varphi^n\left(X^k \geq X^1 - c/2\right). \tag{5.5}$$

Now,

$$\begin{aligned}
& \varphi^n\left(X^k \geq X^{\max} - c\right) \\
= \; & \varphi^n\left(X^k \geq X^{\max} - c \mid X^{\max} = X^1\right) \cdot \varphi^n\left(X^{\max} = X^1\right) \\
& + \varphi^n\left(X^k \geq X^{\max} - c \mid X^{\max} \neq X^1\right) \cdot \varphi^n\left(X^{\max} \neq X^1\right) \\
\geq \; & \varphi^n\left(X^k \geq X^{\max} - c \mid X^{\max} = X^1\right) \cdot \varphi^n\left(X^{\max} = X^1\right) \\
= \; & \varphi^n\left(X^k \geq X^1 - c\right) \cdot \varphi^n\left(X^{\max} = X^1\right).
\end{aligned}$$

So, to show (5.5) it is sufficient to show that there exists $N \in \mathbb{N}$ such that for all $n \geq N$,

$$\varphi^n\left(X^k \geq X^1 - c\right) \cdot \varphi^n\left(X^{\max} = X^1\right) \geq \varphi^n\left(X^k \geq X^1 - c/2\right). \tag{5.6}$$

Using Lemma 6.3, $\varphi^n\left(X^{\max} = X^1\right) \to 1$ as $n \to \infty$. We have,

$$\begin{aligned}
& \frac{\varphi^n\left(X^k \geq X^1 - c/2\right)}{\varphi^n\left(X^k \geq X^1 - c\right)} \\
= \; & \frac{\varphi^n\left(\left(X^k - X^1\right) - \left(E\left(X^k\right) - E\left(X^1\right)\right) \geq -c/2 - \left(E\left(X^k\right) - E\left(X^1\right)\right)\right)}{\varphi^n\left(\left(X^k - X^1\right) - \left(E\left(X^k\right) - E\left(X^1\right)\right) \geq -c - \left(E\left(X^k\right) - E\left(X^1\right)\right)\right)}.
\end{aligned}$$

Note that $\varphi^n\left(\left(X^k - X^1\right) - \left(E\left(X^k\right) - E\left(X^1\right)\right)\right)$ has a normal distribution with mean 0, and where the variance of $\varphi^n\left(X^k - X^1\right)$ tends to 0 as $n \to \infty$. Moreover, $-c - \left(E\left(X^k\right) - E\left(X^1\right)\right) > 0$ as $E\left(X^k\right) - E\left(X^1\right) < E\left(X^k\right) - E\left(X^{k-1}\right) = -c$. Hence, using Lemma 6.5,

$$\frac{\varphi^n\left(\left(X^k - X^1\right) - \left(E\left(X^k\right) - E\left(X^1\right)\right) \geq -c/2 - \left(E\left(X^k\right) - E\left(X^1\right)\right)\right)}{\varphi^n\left(\left(X^k - X^1\right) - \left(E\left(X^k\right) - E\left(X^1\right)\right) \geq -c - \left(E\left(X^k\right) - E\left(X^1\right)\right)\right)} \to 0$$

as $n \to \infty$. Then, we have,

$$\frac{\varphi^n\left(X^k \geq X^1 - c/2\right)}{\varphi^n\left(X^k \geq X^1 - c\right)} \to 0.$$

So, there exists $N \in \mathbb{N}$ such that for all $n \geq N$,

$$\varphi^n\left(X^{\max} = X^1\right) \geq \frac{\varphi^n\left(X^k \geq X^1 - c/2\right)}{\varphi^n\left(X^k \geq X^1 - c\right)}.$$

This proves (5.6), which, as we have shown, implies (5.5).

Now, by (5.5) we have

$$\begin{aligned} & \frac{\varphi^n\left(X^k \geq X^{\max}\right)}{\varphi^n\left(X^k \geq X^{\max} - c\right)} \\ \leq \; & \frac{\varphi^n\left(X^k \geq X^1\right)}{\varphi^n\left(X^k \geq X^1 - c/2\right)} \\ = \; & \frac{\varphi^n\left(\left(X^k - X^1\right) - \left(E\left(X^k\right) - E\left(X^1\right)\right) \geq -\left(E\left(X^k\right) - E\left(X^1\right)\right)\right)}{\varphi^n\left(\left(X^k - X^1\right) - \left(E\left(X^k\right) - E\left(X^1\right)\right) \geq -c/2 - \left(E\left(X^k\right) - E\left(X^1\right)\right)\right)} \\ = \; & \frac{\varphi^n\left(\left(X^k - X^1\right) - \left(E\left(X^k\right) - E\left(X^1\right)\right) \geq \left(E\left(X^1\right) - E\left(X^k\right)\right)\right)}{\varphi^n\left(\left(X^k - X^1\right) - \left(E\left(X^k\right) - E\left(X^1\right)\right) \geq \left(E\left(X^1\right) - E\left(X^k\right)\right) - c/2\right)} \\ \to \; & 0 \end{aligned}$$

as n goes to infinity. Here the convergence follows from Lemma 6.5 as $\left(E\left(X^1\right) - E\left(X^k\right)\right) - c/2 > 0$. So, we have shown (5.4), which completes case 2. Hence, we have shown that (5.1) holds for all k. Therefore, $\lim_{n\to\infty} \epsilon_n = 0$ and hence the proof is complete. ■

REFERENCES

ABDULKADIROĞLU, A. AND T. SÒNMEZ (1998): "Random serial dictatorship and the core from random endowments in house allocation problems," *Econometrica*, 66, 689–701.

ASHEIM, G. (2002): "Proper rationalizability in lexicographic beliefs," *International Journal of Game Theory*, 30, 453–478.

BANDYOPADHYAY, T., R. DEB, AND P. PATTANAIK (1982): "The structure of coalitional power under probabilistic group decision rules* 1," *Journal of Economic Theory*, 27, 366–375.

BARBERA, S., A. BOGOMOLNAIA, AND H. VAN DER STEL (1998): "Strategy-proof probabilistic rules for expected utility maximizers," *Mathematical Social Sciences*, 35, 89–103.

BARBERÀ, S., F. GUL, AND E. STACHETTI (1993): "Generalized Median Voter Schemes and Committees," *Journal of Economic Theory*, 61, 262–289.

BARBERÀ, S., J. J. MASSÓ, AND A. NEME (1997): "Voting under Constraints," *Journal of Economic Theory*, 76, 298–321.

——— (2005): "Voting by Committees under Constraints," *Journal of Economic Theory*, 122, 185–205.

BARBERÁ, S. AND H. SONNENSCHEIN (1978): "Strategy-proof Voting Schemes with Continuous Preferences," *Journal of Economic Theory*, 18, 244–254.

BARBERÁ, S., H. SONNENSCHEIN, AND L. ZHOU (1991): "Voting by Committees," *Econometrica*, 59, 595–609.

BOGOMOLNAIA, A. AND H. MOULIN (2001): "A New Solution to the Random Assignment Problem," *Journal of Economic Theory*, 100, 295–328.

——— (2004): "Random Matching under Dichotomous Preferences," *Econometrica*, 72, 2957–279.

BOGOMOLNAIA, A., H. MOULIN, AND R. STONG (2005): "Collective choice under dichotomous preferences," *Journal of Economic Theory*, 122, 165–184.

CRÈS, H. AND H. MOULIN (2001): "Scheduling with opting out: improving upon random priority," *Operations Research*, 49, 565–577.

DEKEL, E. AND D. FUDENBERG (1990): "Rational behavior with payoff uncertainty," *Journal of Economic Theory*, 52, 243–267.

Dutta, B. (1980): "Strategic voting in a probabilistic framework," *Econometrica: Journal of the Econometric Society*, 447–456.

Gibbard, A. (1973): "Manipulation of voting schemes: a general result," *Econometrica: journal of the Econometric Society*, 41, 587–601.

——— (1977a): "Manipulation of schemes that mix voting with chance," *Econometrica: Journal of the Econometric Society*, 665–681.

——— (1977b): "Manipulation of Voting Schemes that Mix Voting With Chance," *Econometrica*, 45, 665–681.

LeBreton, M. and A. Sen (1999): "Separable Preferences, Strategy-proofness and Decomposability," *Econometrica*, 67, 605–628.

Maus, S., H. Peters, and T. Storcken (2007): "Minimal manipulability: unanimity and nondictatorship," *Journal of mathematical economics*, 43, 675–691.

Moulin, H. (1980): "On strategy-proofness and single peakedness," *Public Choice*, 35, 437–455.

Moulin, H. and R. Stong (2002): "Fair Queuning and other Probabilistic Allocation Methods," *Mathematics of Operations Research*, 27, 1–31.

Myerson, R. (1978): "Refinements of the Nash equilibrium concept," *International Journal of Game Theory*, 7, 73–80.

——— (1981): "Optimal Auction Design," *Mathematics of Operations Research*, 6, 1767–1797.

Saari, D. (1990): "Susceptibility to manipulation," *Public Choice*, 64, 21–41.

Sasaki, H. (1997): "Randomized uniform allocation mechanism and single-peaked preferences of indivisible good," *Waseda University*.

Satterthwaite, M. (1975): "Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions* 1," *Journal of Economic Theory*, 10, 187–217.

Schuhmacher, F. (1999): "Proper rationalizability and backward induction," *International Journal of Game Theory*, 28, 599–615.

Sprumont, Y. (1995): "Strategy-proof Collective Choice in Economic and Political Environments," *Canadian Economic Review*, 28, 68–107.

Svensson, L.-G. and Torstensson (2008): "Strategy-proof Allocation of Multiple Public Goods," *Social Choice and Welfare*, 30, 181–196.

Tan, T. and S. da Costa Werlang (1988): "The Bayesian foundations of solution concepts of games* 1," *Journal of Economic Theory*, 45, 370–391.

Curriculum Vitae

Souvik Roy

Contact Information

Department of Quantitative Economics
University of Maastricht
P.O. Box 616
6200 MD Maastricht
The Netherlands
Email: souvik_roy2004@yahoo.com

Education

Ph.D.	University of Maastricht (expected December 2010)
M.Stat.	Indian Statistical Institute, Delhi, 2004-2006.
B.Sc.	Narendrapur, R.K. Mission Residential College, 2001-2004.

Research Interests

social choice theory, game theory, epistemic game theory.

Working Papers

- "On the manipulability of approval voting and related scoring rules", with Hans Peters and Tom Storcken, 2008 (submitted).
- "Proper Rationalizability in Epistemic Game Theory", with Andres Perea, 2009 (submitted).
- "Characterization of probabilistic rules on single peaked domain", with Hans Peters, Arunave Sen and Tom Storcken, 2010.

- “The structure of strategy-proof random choice functions over product domain and separable preferences: The case of two voters”, with Shurojit Chatterji and Arunava Sen, 2010 (submitted).
- “Strategy-proof voting rules on a multidimensional policy space for a continuum of voters with elliptic preferences ”, with Hans Peters and Tom Storcken, 2010 (submitted).

Recent Talks

Social Choice and Welfare Meeting (2008, 2010); Indian Statistical Institute (2009); Jawaharlal Nehru University (2009); Bielefeld University (2009); Conference on Economic Design 2009; Tilburg University (2009).

Teaching

At University of Maastricht: Quantitative Economics-1; Quantitative Economics-2; Man. Of operations and New Product Development.

Personal Information

Birthdate: September 2, 1984; Citizenship: India.