Discussion Papers in Economics

ORDINALLY BAYSIAN INCENTIVE-COMPATIBLE VOTING

Dipjyoti Majumdar
and
Arunava Sen

February 2003

Discussion Paper 03-01

Indian Statistical Institute, Delhi
Planning Unit
7 S.J.S. Sansanwal Marg, New Delhi 110 016, India

ORDINALLY BAYESIAN INCENTIVE-COMPATIBLE VOTING SCHEMES *

Dipjyoti Majumdar[†]

and

Arunava Sen[‡]

Current Version: February, 2003

*We wish to thank Partha Sarathi Chakraborty, John Ledyard, Arup Pal, Rahul Roy, two anonymous referees, a Co-Editor and numerous conference and seminar participants for valuable comments.

[†]Indian Statistical Institute, New Delhi, India and Laboratoire d' Économétrie, École Polytechnique, Paris, France

[‡]Indian Statistical Institute, New Delhi, India

Abstract

We study strategic voting after weakening the notion of strategy-proofness to *Ordinal Bayesian Incentive Compatibility (OBIC)*. Under OBIC, truthtelling is required to maximize the expected utility of every voter, expected utility being computed with respect to the voter's prior beliefs and under the assumption that everybody else is also telling the truth. We show that for a special type of priors i.e., the uniform priors there exists a large class of social choice functions that are OBIC. However, for priors which are generic in the set of independent beliefs a social choice function is OBIC only if it is dictatorial. This result underlines the robustness of the Gibbard-Satterthwaite Theorem.

1 Introduction

In the classical model of strategic voting, each voter knows his own preferences but is ignorant of the preferences of other voters. The objectives of the social planner are represented by a *social choice function* which associates a feasible alternative with every profile of voter preferences. Voters are fully aware of their strategic opportunities; by making different announcements of their preferences, they can influence the alternative that is selected. The goal of the planner is to select a social choice function which gives voters appropriate incentives to reveal their private information truthfully. It is clear the choice of equilibrium concept is critical. The concept which has been preponderant in the literature is *strategy-proofness*. This requires truth-telling for each voter to be a dominant strategy. In other words, each voter cannot do better by deviating from the truth irrespective of what he believes the other voters will announce. This is clearly a demanding requirement. And this intuition is confirmed by the celebrated Gibbard-Satterthwaite Theorem which states that under mild assumptions, the only social choice functions which are strategy-proof are *dictatorial*. A dictatorial social choice function is one which always selects the maximal element of a particular voter (who is the dictator). It is quite clear that this is a powerful negative result.

Our objective in this paper is to analyse the implications of weakening the truth-telling requirement from strategy-proofness to *ordinal Bayesian incentive-compatibility*. This notion was introduced in d'Aspremont and Peleg (1988) in the context of a different problem, that of the representation of committees. It is the obvious adaptation to voting theory of the notion of incentive-compatibility which is widely used in standard incentive theory (for instance, in the theory of auctions). Truth-telling is required to maximize the *expected utility* of each voter. This expected utility is computed with reference to the voter's *prior beliefs* about the (possible) preferences of the other voters and based on the assumption that other voters follow the truth-telling strategy. More formally, truth-telling is required to be a Bayes-Nash equilibrium in the direct revelation game, modeled as a game of incomplete information. Since social choice functions depend only on voters' ranking of various alternatives, truth-telling is required to maximize expected utility for *every* representation of the voter's true ranking.

Ordinal Bayesian incentive-compatibility is a significant weakening of the truth-telling requirement. Note that whether or not a social choice function satisfies ordinal Bayesian incentive-compatibility depends on the beliefs of each voter. It satisfies strategy-proofness only if it satisfies ordinal Bayesian incentive-compatibility with respect to *all* beliefs of each voter. However, we are able to prove the following. Assume that voters have have a common prior that is independently distributed. There is a set of beliefs $\mathcal{C}$ which is generic in the set of all independently distributed beliefs such that a social choice function is ordinally Bayesian incentive-compatible with respect to *any* belief in $\mathcal{C}$, only if it is dictatorial. Of course, we assume that there are at least three alternatives and that all social choice functions under consideration satisfy the mild requirement of unanimity.

Our result underlines the extraordinary robustness of the Gibbard-Satterthwaite Theorem. For "almost all" beliefs, the weaker requirement of ordinal Bayesian incentive-compatibility is sufficient to force dictatorship. The Gibbard-Satterthwaite Theorem is, of course, a corollary of our result but the latter also provides a precise picture (in the space of beliefs), of how pervasive the dictatorship problem is.

The negative generic result requires a very important qualification. A significant non-generic case is the one where each voters' beliefs about the preferences of the others is a uniform distribution. This is an important case in decision theory and is the so-called case of "complete ignorance". A dramatically different picture emerges here. We provide a weak sufficient condition for a social choice function to be ordinally Bayesian incentive-compatible and show that a variety of well-behaved social choice functions do satisfy this condition (for instance, selections from scoring correspondences). The overall picture is therefore complex and nuanced. Generically, ordinal Bayesian incentive-compatibility implies dictatorship but in non-generic cases which are of considerable interest, significant possibility results exist.

The paper is organized as follows. In Section 2 we set out the basic notation and definitions. In Sections 3 and 4, we consider respectively the case of uniform priors and the generic case. We discuss our results in Section 5 while Section 6 concludes. The proof of the main result is contained in the Appendix.

2 Notation and Definitions

The set $N = \{1, \cdots, N\}$ is the set of voters or individuals. The set of outcomes is the set A with $|A| = m$. Elements of A will be denoted by a, b, c, d etc. Let $\mathbb{P}$ denote the set of strict orderings[1] of the elements of A. A typical preference ordering will be denoted by P_i where aP_ib will signify that a is preferred (strictly) to b under P_i. A preference profile is an element of the set $\mathbb{P}^N$. Preference profiles will be denoted by $P, \bar{P}, P'$ etc and their i-th components as $P_i, \bar{P}_i, P'_i$ respectively with $i = 1, \cdots, N$. Let $(\bar{P}_i, P_{-i})$ denote the preference profile where the i-th component of the profile P is replaced by $\bar{P}_i$.

For all $P_i \in \mathbb{P}$ and $k = 1, \cdots, M$, let $r_k(P_i)$ denote the k th ranked alternative in P_i, i.e., $r_k(P_i) = a$ implies that $|\{b \neq a | bP_ia\}| = k - 1$.

Definition 2.1 *A Social Choice Function or (SCF) f is a mapping $f : \mathbb{P}^N \to A$.*

A SCF can be thought of as representing the objectives of a planner, or equivalently, that of society as a whole. An important observation in the context of our paper is that we assume SCFs to be *ordinal*. In other words, the only information

[1]A strict ordering is a complete, transitive and antisymmetric binary relation

used for determining the value of an SCF are the rankings of each individual over feasible alternatives. This is a standard assumption in voting theory.

Throughout the paper, we assume that SCFs under consideration satisfy the axiom of *unaninimty*. This is an extremely weak assumption which states that in any situation where all individuals agree on some alternative as the best, then the SCF must respect this consensus. More formally,

DEFINITION 2.2 *A SCF f is unanimous if $f(P) = a_j$ whenever $a_j = r_1(P_i)$ for all individuals $i \in N$.*

We assume that an individual's preference ordering is private information. Therefore SCFs have to be designed in a manner such that all individuals have the "correct" incentives to reveal their private information. It has been standard in the strategic voting literature to require that SCFs are *strategy-proof*, i.e. they provide incentives for truth-telling behaviour in dominant strategies. A strategy-proof SCF has the property that no individual can strictly gain by misrepresenting his preferences, no matter what preferences are announced by other individuals.

DEFINITION 2.3 *A SCF f is strategy-proof if there does not exist $i \in N$, $P_i, P_i' \in \mathbb{P}$, and $P_{-i} \in \mathbb{P}^{N-1}$, such that*

$$f(P_i', P_{-i}) P_i f(P_i, P_{-i})$$

The Gibbard-Satterthwaite Theorem characterizes the class of SCFs which are strategy-proof and unanimous. This is the class of dictatorial SCFs.

DEFINITION 2.4 *A SCF f is dictatorial if there exists an individual i such that, for all profiles P we have $f(P) = r_1(P_i)$.*

THEOREM **2.1** *Gibbard (1973), Satterthwaite (1975)*

Assume $m \geq 3$. A SCF is unanimous and strategy-proof if and only if it is dictatorial.

In this paper, we explore the consequences of weakening the incentive requirement for SCFs from strategy-proofness to *ordinal Bayesian incentive compatibility.* This concept originally appeared in d'Aspremont and Peleg (1988) and we describe it formally below.

DEFINITION 2.5 *A belief for an individual i is a probability distribution on the set $\mathbb{P}^N$, i.e. it is a map $\mu_i : \mathbb{P}^N \to [0,1]$ such that $\sum_{P \in \mathbb{P}^N} \mu_i(P) = 1$.*

We assume that all individuals have a common prior belief μ. Clearly μ belongs to the unit simplex of dimension $m!^N - 1$. For all μ, for all P_{-i} and P_i, we shall let $\mu(P_{-i}|P_i)$ denote the conditional probability of P_{-i} given P_i. The conditional probability $\mu(P_{-i}|P_i)$ belongs to the unit simplex of dimension $m!^{N-1} - 1$

DEFINITION 2.6 *The utility function $u : A \to \Re$ represents $P_i \in \mathbb{P}$, if and only if for all $a, b \in A$,*

$$aP_ib \Leftrightarrow u(a) > u(b)$$

We will denote the set of utility functions representing P_i by $\mathcal{U}(P_i)$.

We can now define the notion of incentive compatibility that we use in the paper.

DEFINITION 2.7 *A SCF f is Ordinally Bayesian Incentice Compatible (OBIC) with respect to the belief μ if for all $i \in N$, for all $P_i, P_i' \in \mathbb{P}$, for all $u \in \mathcal{U}(P_i)$, we have*

$$\sum_{P_{-i} \in \mathbb{P}^{N-1}} u\left(f(P_i, P_{-i})\right) \mu(P_{-i}|P_i) \geq \sum_{P_{-i} \in \mathbb{P}^{N-1}} u\left(f(P_i', P_{-i})\right) \mu(P_{-i}|P_i) \qquad (1)$$

Let f be a SCF and consider the following game of incomplete information as formulated originally in Harsanyi (1967). The set of players is the set N. The set of types for a player is the set $\mathbb{P}$ which is also the set from which a player chooses an action. If player i's type is P_i, and if the action-tuple chosen by the players is P', then player i's payoff is $u(f(P'))$ where u is a utility function which represents P_i. Player i's beliefs are given by the probability distribution μ. The SCF f is OBIC if truth-telling is a Bayes-Nash equilibrium of this game. Since SCF's under consideration are ordinal by assumption, there is no "natural" utility function for expected uility calculations. Under these circumstances, OBIC requires that a player cannot gain in expected utility (conditional on type) by unilaterally misrepresenting his preferences no matter what utility function is used to represent his true preferences.

It is clear that strategy-proofness is a more stringent requirement than OBIC with respect to a particular belief. We record without proof the precise relationship between the two concepts below.

Remark 2.1 A SCF is strategy-proof if and only it is OBIC with respect to all beliefs μ.

It is possible to provide an alternative definition of OBIC in terms of stochastic dominance. Let f be a SCF and pick an arbitrary individual i and a preference ordering P_i. Suppose alternative a is first-ranked under P_i. Let α denote the probability conditional on P_i that a is the outcome when i announces P_i assuming that other players are truthful as well. Thus α is the sum of $\mu(P_{-i}|P_i)$ over all P_{-i} such that $f(P_i, P_{-i}) = a$. Similarly, let β be the probability that a is the outcome if he announces P'_i, i.e β is the sum of $\mu(P_{-i}|P_i)$ over all P_{-i} such that $f(P'_i, P_{-i}) = a$. If f is OBIC with respect to μ then we must have $\alpha \geq \beta$. Suppose this is false. Then there exists a utility function which gives a utility of one to a and virtually zero to all other outcomes which represents P_i and such that the expected utility from announcing the truth for agent i with preferences P_i is strictly lower than from announcing P_i'. Using a similar argument, it follows that the probability of obtaining the first k ranked alternatives $k = 1, \cdots, m$ according to P_i under truthtelling must be at least as great as under misreporting via P_i'. We make these ideas precise below.

For all $i \in N$, for any $P_i \in \mathbb{P}$ and for any $a \in A$, let $B(a, P_i) = \{b \in A | bP_ia\} \cup \{a\}$. Thus $B(a, P_i)$ is the set of alternatives that are weakly preferred to a under P_i.

Definition 2.8 *The SCF f is OBIC with respect to the belief μ if for all $i \in N$, for all integers $k = 1, \cdots, m$ and for all P_i and P_i',*

$$\begin{aligned} &\mu(\{P_{-i} | f(P_i, P_{-i}) \in B(r_k(P_i), P_i)\} | P_i) \\ \geq\ &\mu(\{P_{-i} | f(P_i', P_{-i}) \in B(r_k(P_i), P_i)\} | P_i) \end{aligned} \tag{2}$$

We omit the proof of the equivalence of the two definitions of OBIC. The proof is easy and we refer the interested reader to Theorem 3.11 in d'Aspremont and Peleg (1988).

3 Uniform Priors

We begin by analyzing the case of uniform priors. Our objective is to provide a weak sufficient condition for OBIC with respect to this prior. Although we shall demonstrate in the next section that this possibility result disappears if the prior is perturbed, it is nevertheless of interest because of the importance of uniform priors in decision theory.

Assumption 3.1 For all i, for all P_i, P_i' and for all P_{-i} and P_{-i}', we have

$$\mu(P_{-i} | P_i) = \mu(P_{-i}' | P_i')$$

We denote these uniform beliefs by $\bar{\mu}$. Restating Definition 2.8 in the present context, we have

PROPOSITION 3.1 The SCF f is OBIC with respect to the belief $\bar{\mu}$ if, for all i, for all integers $k = 1, \cdots, m$, for all P_i and P_i', we have

$$|\{P_{-i}|f(P_i, P_{-i}) \in B(r_k(P_i), P_i)\}| \geq |\{P_{-i}|f(P_i', P_{-i}) \in B(r_k(P_i), P_i\}| \qquad (3)$$

We omit the (trivial) proof of this Proposition. It will be convenient to express equation (3) in a more compact way. For all $P_i \in I\!\!P$ and $x \in A$, let

$$\eta(x, P_i) \equiv |\{P_{-i}|f(P_i, P_{-i}) = x\}|$$

Equation (3) can now be expressed as follows. For all i, for all integers $k = 1, \cdots, m$, for all P_i and P_i', we have

$$\sum_{t=1}^{k} \eta(r_k(P_i), P_i) \geq \sum_{t=1}^{k} \eta(r_k(P_i), P_i') \qquad (4)$$

We now give an example of a non-dictatorial SCF which is OBIC with respect to $\bar{\mu}$.

EXAMPLE 3.1

Let $A = \{a, b, c\}$, $N = \{1, 2\}$. Consider the SCF defined below.

$$
\begin{array}{ccccccc}
 & abc & acb & bac & bca & cab & cba \\
abc & a & a & a & b & c & a \\
acb & a & a & b & a & a & c \\
bac & b & a & b & b & b & c \\
bca & a & b & b & b & c & b \\
cab & a & c & c & b & c & c \\
cba & c & a & b & c & c & c
\end{array}
\tag{5}
$$

In the array above, individual 1's preferences appear along the rows and individual 2's along the columns. The SCF is well-behaved; in particular it is neutral (this will be defined shortly), non-dictatorial and Pareto efficient. To verify that it is OBIC with respect to $\bar{\mu}$, it suffices to observe that for each preference ordering of an individual, the frequency of occurence of its first-ranked alternative is four and of its second and third-ranked alternatives, one each respectively. It is easy to modify the example slightly in order to obtain a SCF which is anonymous (i.e. invariant with respect to the permutation of individuals). Details may be found in Majumdar (2002), Chapter 2.

We introduce some definitions which are required for the main result of this section.

Definition 3.1 *Let $\sigma : A \to A$ be a permutation of A. Let P^σ denote the profile $(P_1^\sigma, \cdots, P_N^\sigma)$ where for all i and for all $a, b \in A$,*

$$aP_ib \Rightarrow \sigma(a)P_i^\sigma\sigma(b)$$

The SCF f satisfies neutrality if, for all profiles P and for all permutation functions σ, we have

$$f(P^{\sigma}) = \sigma[f(P)]$$

Neutrality is a standard requirement for social choice functions and correspondences (see for e.g. Moulin (1983)). All alternatives are treated symmentrically in neutral SCFs i.e. the "names" of the alternatives do not matter.

Let P_i be an ordering and let $a \in A$. We say that P_i' represents an *elementary a-improvement of* P_i if

- for all $x, y \in A \setminus \{a\}$, $xP_iy \Leftrightarrow xP_i'y$
- $[a = r_k(P_i)] \Rightarrow [a = r_{k-1}(P_i')]$, if $k > 1$
- $[a = r_1(P_i)] \Rightarrow [a = r_1(P_i')]$

DEFINITION 3.2 *The SCF f satisfies elementary monotonicity if for all i, P_i,P_i' and P_{-i}*

$$[f(P_i, P_{-i}) = a \textit{ and } P_i' \textit{ represents an a-elementary improvement of } P_i] \Rightarrow$$
$$[f(P_i', P_{-i}) = a]$$

Let P be a profile where the outcome is a. Suppose a moves up one place in some individual's ranking without disturbing the relative positions of any other alternative. Then elementary monotonicity requires a to be the outcome at the new profile. This is relatively weak axiom whose implications we will discuss more fully after stating and proving the main result of this section.

THEOREM **3.1** *A SCF which satisfies neutrality and elementary monotonicity is OBIC with respect to $\bar{\mu}$.*

PROOF: Let f be a SCF which is neutral and satisfies elementary monotonicity. We will show that it is OBIC with respect to $\bar{\mu}$.

Our first step is to show that the neutrality of f implies that, for all i, for all integers $k = 1, \cdots, m$ and for all P_i and P_i', we have $\eta(r_k(P_i), P_i) = \eta(r_k(P_i'), P_i')$. Pick an individual i and orderings P_i and P_i'. Define a permutation function on A as follows: for all integers $k = 1, \cdots, m$,

$$\sigma(r_k(P_i)) = r_k(P_i')$$

Observe that $P_i^\sigma = P_i'$. Fix an integer $k \in \{1, \cdots, m\}$. Let P_{-i} be such that $f(P_i, P_{-i}) = r_k(P_i)$. Since f is neutral,

$$f(P_i', P_{-i}^\sigma) = \sigma[f(P_i, P_{-i})] = \sigma[r_k(P_i)] = r_k(P_i') \tag{6}$$

Equation (6) above establishes that

$$\eta(r_k(P_i), P_i) \leq \eta(r_k(P_i'), P_i').$$

By using the permutation σ^{-1}, the argument above can be replicated to prove the reverse inequality.

The next step in the proof is to show that for all i, for all integers $k = 1, \cdots, m-1$, and for all P_i,

$$\eta(r_k(P_i) \geq \eta(r_{k+1}(P_i))$$

Pick i, $k \in \{1, \cdots, m-1\}$ and P_i. Let P_i' be an elementary $r_{k+1}(P_i)$-improvement of P_i. Since f satisfies elementary monotonicity, we must have

$$\{P_{-i} | f(P_i, P_{-i}) = r_{k+1}(P_i)\} \subseteq \{P_{-i} | f(P_i', P_{-i}) = r_{k+1}(P_i)\} \tag{7}$$

Equation (7) above implies that

$$\eta(r_{k+1}(P_i), P_i') \geq \eta(r_{k+1}(P_i), P_i) \tag{8}$$

But the LHS of equation (8) equals $\eta(r_k(P_i'), P_i')$ which from the first part of the proof, equals $\eta(r_k(P_i), P_i)$. This proves our claim. Observe that this claim implies that

$$\eta(r_k(P_i), P_i) \geq \eta(r_t(P_i), P_i) \text{ whenever } k < t \tag{9}$$

We now complete the proof of the Theorem. Let i be an individual, let $k \in \{1, \cdots, m\}$ be an integer and let P_i and P_i' be orderings. Let $T = \{s | r_s(P_i') = r_t(P_i)\}$. From the first part of the proof we have,

$$\sum_{t=1}^{k} \eta(r_t(P_i), P_i') = \sum_{t \in T} \eta(r_t(P_i), P_i) \tag{10}$$

But from equation (9)

$$\sum_{t \in T} \eta(r_t(P_i), P_i) \leq \sum_{t=1}^{k} \eta(r_t(P_i), P_i) \tag{11}$$

Combining equations (10) and (11), we obtain

$$\textstyle\sum_{t=1}^{k} \eta(r_t(P_i), P_i) \geq \sum_{t=1}^{k} \eta(r_t(P_i), P_i')$$

so that f is OBIC with respect to $\bar{\mu}$. ■

Theorem 3.1 is a positive result. Neutrality and Elementary Monotonicity are relatively weak requirements for SCFs to satisfy. We provide an important class of examples below.

EXAMPLE 3.2 (Scoring Correspondences)

Let $s \equiv (s_1, s_2, \cdots, s_m)$ be a vector in $\mathbb{R}^m$ with the property that $s_1 \geq s_2 \geq \cdots, \geq s_m$ and $s_1 > s_m$. Let P be a profile. The *score* assigned to alternative a in P by individual i is s_k if $r_k(P_i) = a$. The aggregate score of a in P is the sum of its individual scores in P. Let $W_s(P)$ denote the set of alternatives whose scores in P are maximal. The social choice correspondence W defined by this procedure is called a *scoring correspondence* and is discussed in greater detail in Moulin (1983). Important correspondences which belong to this class are the plurality and the Borda correspondences.

We define a SCF f which is a selection from W in the following manner. For all profiles P, $f(P)$ is the alternative in $W_s(P)$ which is maximal according to P_1, i.e. it is the element in the set $W_s(P)$ which is the highest ranked in individual 1's preferences. Observe that f is neutral. We also claim that it satisfies elementary monotonicity. To see this, suppose $f(P) = a$ and let P_i' be an a-improvement of P_i for some individual i. Observe that the score of a in P_i' increases relative to that in P_i while that of the other alternatives either remains constant or falls. Therefore the aggregate score of a in the profile (P_i', P_{-i}) is strictly greater than in P while that of the other alternatives is either the same or less. Therefore $W_s(P_i', P_{-i})) = \{a\} = f(P_i', P_{-i})$ and elementary monotonicity is satisfied. Theorem 3.1 allows to conclude that f is OBIC with respect to $\bar{\mu}$. Indeed any neutral selection from a scoring correspondence will satisfy this property.

Moulin (1983) contains a more extensive discussion of elementary monotonicity (which he calls monotonicity). He shows (Chapter 3, Lemma 1) that in addition to scoring correspondences, Condorcet-type correspondences (those which select majority winners whenever they exist) such as the Copeland and Kramer rules, the Top-cycle and the uncovered set, all satisfy elementary monotonicity. It is easy to show that a neutral selection of these correspondences obtained, for instance, by breaking ties in the manner of the previous example (using the preference ordering of a given individual), generates a SCF which is OBIC with respect to $\bar{\mu}$.

Remark 3.1 Theorem 3.1 only provides a sufficient condition for a SCF to be OBIC with respect to $\bar{\mu}$. In order to see this observe that the SCF in Example 3.1 is neutral but violates elementary monotonicity. For instance, observe that $f(abc, cba) = a$ but $f(abc, cab) = c$.

Remark 3.2 There are SCFs which are not OBIC with respect to $\bar{\mu}$. Consider, for example the SCF which always picks individual 1's second-ranked alternative.

It is clearly neutral. But it violates equation (4).

4 The Generic Case

The main result of this section is to show that the possibility results of the previous section do not hold generally. However we need to make a crucial assumption regarding admissible beliefs.

ASSUMPTION 4.1 All admissible beliefs μ are *independent*, i.e. for all $k = 1, 2, \cdots, N$, there exist probability distributions $\mu_k : \mathbb{P} \to [0, 1]$ such that

$$\mu(P) = \times_{k=1}^{N} \mu_k(P_k)$$

We denote the set of all independent priors by Δ^I. The set Δ^I is the N-th order Cartesian product of unit simplices Δ, where each Δ is of dimension $m! - 1$.

We can now state the main result of this section.

THEOREM **4.1** *Let $m \geq 3$ and assume that beliefs are independent. There exists a subset $\mathcal{C}$ of Δ^I such that*

- *$\mathcal{C}$ is open and dense in Δ^I*
- *$\Delta^I - \mathcal{C}$ has Lebesgue measure zero*
- *if f is unanimous and is OBIC w.r.t μ where $\mu \in \mathcal{C}$, then f is dictatorial.*

The theorem states that there is a subset of the set of independent beliefs which is generic in the latter set which has the property that every unanimous SCF which is OBIC with respect to *any* belief in this set, is dictatorial. We emphasize that the genericity is in the set of independently generated beliefs and not in the space of all probability distribution of types.

The proof of the theorem is contained in the Appendix.

REMARK 4.1 Theorem 4.1 can easily be extended to cover the case where voters do not have common beliefs. Here OBIC has to be defined with respect to an N-tuple of beliefs or a *belief system*. If the belief of each voter is assumed to be independent, then there exists a set of beliefs for each voter with the following property: every unanimous SCF which is OBIC with respect to a belief system where each voter's belief is picked arbitrarily from this set must be dictatorial. In addition, the same genericity properties hold for these sets of beliefs. The arguments required to prove this result are virtually identical to those in the paper. There is however a sense in which this result is more general than the one in the common priors case. We no longer need to assume independence of each voter's beliefs in the case where there are more than two voters - the weaker assumption of *free beliefs* introduced in d'Aspremont and Gérard-Varet (1982) suffices. The common prior assumption in conjunction with free beliefs implies that the common prior satisfies independence. Results relating to the non-common priors case can be found in Majumdar (2002) Chapter 2.

REMARK 3.2 The Gibbard-Satterthwaite Theorem is a corollary of Theorem 4.1. This follows immediately from Remark 2.1.

5 Discussion

In this section we attempt to provide some insight into our results. In order to do so we return to the two person, three alternative SCF described in Example 3. Assume that the common belief is independent; in particular let $\mu_1, \mu_2, \cdots, \mu_6$ denote the row voter's belief that the preferences of the column voter's preferences (types) are $abc, acb, \cdots, cba$ respectively.

Suppose that the row voter's true ordering is abc and she considers misrepresenting her preferences by announcing acb. Observe that by telling the truth, she obtains a with probability $\mu_1 + \mu_2 + \mu_3 + \mu_6$ while by lying, she gets a with probability $\mu_1 + \mu_2 + \mu_4 + \mu_5$. We claim that OBIC requires that

$$\mu_1 + \mu_2 + \mu_3 + \mu_6 \geq \mu_1 + \mu_2 + \mu_4 + \mu_5$$

Suppose that the inequality does not hold, i.e. the quantity on the right hand side strictly exceeds that on the left hand side. Then there exists a cardinalization of *abc* where $u(a) = 1$, $u(b) = \delta > 0$ and $u(c) = 0$ with δ sufficiently small such that the expected utility from truth telling is strictly smaller than that from lying via *acb*.

Now suppose that the row voter's true preference is *acb* and she considers lying by announcing *abc*. Replicating the argument above, we obtain the reverse inequality. (Note that in order to make this claim, we are making use of the independence assumption). Combining the two inequalities, we have

$$\mu_1 + \mu_2 + \mu_3 + \mu_6 = \mu_1 + \mu_2 + \mu_4 + \mu_5$$

This equality does hold in the uniform prior case where both left and right hand side are equal to 4/6. However, it will break down if the prior is perturbed. Therefore the SCF will not be OBIC with respect to these perturbed priors. These observations provide an intuition for our results, both positive and negative. In the uniform priors case, the *position* in which various outcomes occur in the rectangular array which represents the SCF, is (relatively) unimportant because only frequencies of occurrence matter. However, in the generic case these positions are critical - alternatives must "line up" in very specific ways in order to satisfy OBIC. These restrictions precipitate a negative result.

We employ a general version of the arguments above to prove Theorem 4.1. We define the set $\mathcal{C}$ of independent beliefs to be the product of marginal distributions over preference types satisfying the following property: if the probability that a voter's type belongs to a set S equals the probability that his type belongs to the set T, then we must have $S = T$. This is clearly violated (in the most extreme way) when the marginal distribution is uniform. But it is clearly generic in the set of all independent beliefs.

Now consider voter i and preference orderings P_i and $\bar{P}_i$ for this voter with the property that the set of the first k (k is an integer between 1 and $m-1$) ranked alternatives in the two orderings are identical. Denote the set of these first k alternatives by B. For instance, in Example 3.1, if the two orderings for the row voter are *abc* and *acb*, then $k = 1$ and $B = \{a\}$; if the orderings are *abc* and *bac*, then $k = 2$ and $B = \{a, b\}$ etc. Let f be a SCF which is OBIC with respect to some belief lying in the set $\mathcal{C}$. Let S be the set of preferences of voters other than i, P_{-i} such that $f(P_i, P_{-i}) \in B$. Similarly let T be the set of all P_{-i} such that $f(\bar{P}_i, P_{-i}) \in B$. Using the argument outlined earlier, we can conclude (from OBIC) that the probability measures of the sets S and T must be the same. But since beliefs lie in the set $\mathcal{C}$, the sets S and T must be the same. Therefore, if $f(P_i, P_{-i}) \in B$, then $f(\bar{P}_i, P_{-i})) \in B$.

This last condition is a montonicity type of condition on SCFs. Although we are unable to show directly that it implies strategy-proofness, we demonstrate that together with unanimity, it implies dictatorship. The proof proceeds by induction on the number of voters. We show that it holds for two voters and then use a "cloning" of voters argument to establish the induction step.

It is clear that Theorem 4.1 depends heavily on the requirement that the expected utility from truth telling is at least as great as that from lying for various cardinalizations of true preferences. A natural question is whether *all* cardinalizations are required for the result i.e. whether the full force of OBIC is necessary. We can provide a fairly clear answer to this question. Reexamining previous arguments we can verify the following: for any preference ordering and any alternative not ranked last, we require cardinalizations which make the utility gap beteween the weak better-than set (with respect to this alternative) and the strictly worse-than set, as large as possible. For instance, if the utility of the best and worst alternatives are normalized to be one and zero respectively, then for all alternatives x and all real numbers $\delta \in (0, 1)$, there must exist a cardinalization u such that $u(x) - u(y) > \delta$ where y is the alternative ranked immediately below x. If there are exactly three alternatives, then all cardinalizations of each preference ordering are indeed required. However, this is not necessary if there are more than three alternatives.

It is worth pointing out an important feature of the set of beliefs $\mathcal{C}$ that we construct. For all marginal distributions derived from beliefs in this set, all preference orderings gets strictly positive probability. To see this, consider the two person three alternative example once again and suppose that $\mu_1 = 0$. Then $\mu_1 + \mu_2 = \mu_2$ so that the probabilities that the column voter's type belongs to the distinct sets $\{abc, acb\}$ and $\{abc\}$ are the same. This implies that admissible beliefs are such that no domain restriction is introduced. Observe that such a requirement is necessary because it is possible to construct strategy-proof SCFs over restricted domains.

Finally we would like to make an observation that may be of help in interpreting our result. We know that if we require a SCF to be robust in the sense of being incentive compatible with respect to all beliefs, we are, in effect, imposing strategy-proofness. We then immediately obtain dictatorship. In fact, we may not need robustness with respect to all beliefs - even a local version of this requirement may be sufficient (Ledyard (1978)). Our results suggests that a related negative result obtains when robustness is imposed not on beliefs but on utility representations.

6 Conclusion

We have examined the implications of weakening the incentive requirement in the standard voting model from dominant strategies to ordinal Bayesian incentive compatibility. The set of ordinal Bayesian incentive compatible social choice functions clearly depends on the beliefs of each agent. A case of particular interest is the case of uniform priors. We provide a weak sufficient condition for incentive compatibility and show that a large class of well-behaved social choice functions satisfy these conditions. However, we show that these possibility results vanish if we perturb these beliefs. We are thus unable to escape the negative conclusion of the Gibbard-Satterthwaite Theorem for generic priors.

Several questions remain to be answered. Although OBIC is a natural concept in an ordinal setting, it is a reasonably strong requirement. It is vital therefore, to investigate a fully cardinal model where the value of a SCF can depend on the cardinalization of individual preferences. It would also be worthwhile to examine the effects of correlation in the voting model as has been done, quite extensively in models with money and quasi-linear utility functions (see for instance, d'Aspremont, Crémer and Gérard-Varet (2002). We hope to address some of these issues in future work.

7 References

- d'Aspremont, C., and L-A. Gérard-Varet (1982), "Bayesian Incentive Compatible Beliefs", *Journal of Mathematical Economics*, 10:83-103.

- d'Aspremont, Crémer and Gérard-Varet (2002),"Correlation, Independence and Bayesian Incentives", *mimeo.*

- d'Aspremont, C., and B. Peleg (1988), "Ordinal Bayesian Incentive Compatible Representation of Committees", *Social Choice and Welfare*, 5:261-280.

- Gibbard, A. (1973), "Manipulation of Voting Schemes: A General Result", *Econometrica* 41: 587-601.

- Harsanyi, J. (1967), "Games with Incomplete Information Played by 'Bayesian'

Players: I-III", *Management Science* 14: 159-182, 320-334, 486-502.

- Ledyard, J.O. (1978), "Incentive Compatibility and Incomplete Information", *Journal of Economic Theory*, 18: 171-189.
- Majumdar, D. (2002), *Essays in Social Choice Theory*, Ph.d dissertation submitted to the Indian Statistical Institute.
- Maskin, E. (1999), "Nash Equilibrium and Welfare Optimality", *Review of Economic Studies*, 66(1):23-38.
- Moulin, H. (1983), *The Strategy of Social Choice*, Advanced Textbooks in Economics, C.J. Bliss and M.D. Intrilligator (eds), North-Holland.
- Muller, E,. and M. Satterthwaite (1977), "The Equivalence of Strong Positive Association and Strategy Proofness", *Journal of Economic Theory*, 14:412-418.
- Satterthwaite, M. (1975), "Strategy-proofness and Arrow's conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions", *Journal of Economic Theory*, 10:187-217.
- Sen, A. (2001), "Another Direct Proof of the Gibbard-Satterthwaite Theorem", *Economics Letters*, 70:381-385.

8 Appendix

Proof of Theorem 4.1: The proof proceeds in several steps. In Step 1, we define the sets $\mathcal{C}$ and show that they are open and dense subsets of Δ^I and the Lebesgue measure of their complement sets are zero. In Step 2, we show that if f is OBIC with respect to the belief μ where $\mu \in \mathcal{C}$, then f must satisfy a certain property which we call Property M. In Steps 3 and 4, we show by induction on the number of individuals that a SCF which satisfies Property M must be dictatorial. In Step 3, we show that this is true in the case of two individuals. In Step 4, we complete the induction step.

STEP 1

We define the set $\mathcal{C}$ below.

For any $Q \subseteq I\!P^N$, let $\mu(Q) = \sum_{P \in Q} \mu(P)$. The set $\mathcal{C}$ is defined as the set of beliefs μ satisfying the following property: for all $Q, T \subset I\!P^N$

$$[\mu(Q) = \mu(T)] \Rightarrow [Q = T]$$

We first show that $\mathcal{C}$ is open in Δ^I. Pick any $\mu \in \mathcal{C}$ and let

$$\phi(\mu) = \min_{S,T \subset I\!P^N, S \neq T} |\mu(S) - \mu(T)|$$

Observe that $\phi(\mu) > 0$. Since ϕ is a continuous function of μ, there exists $\epsilon > 0$ such that for all $\hat{\mu} \in \Delta^I$ with $d(\hat{\mu}, \mu) < \epsilon$, [2] we have $\phi(\hat{\mu}) > 0$. But this implies that $\hat{\mu} \in \mathcal{C}$. Therefore $\mathcal{C}$ is open in Δ^N.

We now show that $\Delta^I - \mathcal{C}$ has Lebesgue measure zero. We begin with the observation that Δ^I is the Cartesian product of N simplices each of which is of dimension $m! - 1$. On the other hand,

$$\Delta^I - \mathcal{C} = \bigcup_{Q,T \subset I\!P^N} \{\mu \in \Delta^I | \mu(Q) = \mu(T)\}$$

Therefore the set $\Delta^I - \mathcal{C}$ is the union of a finite number of hypersurfaces intersected with Δ^I. It follows immediately that it is a set of lower dimension and hence has zero Lebesgue measure.

Pick $\mu \in \Delta^I - \mathcal{C}$ and consider an open neighbourhood of radius $\epsilon > 0$ with centre μ. Since this neighbourhood has strictly positive measure and since $\Delta^I - \mathcal{C}$

[2] $d(.,.)$ here signifies Euclidean distance

has measure zero, it must be the case that the neighbourhood has a non-empty intersection with the set $\mathcal{C}$. This establishes that $\mathcal{C}$ is dense in Δ^I.

This completes Step 1.

STEP 2

Let f be a SCF which is OBIC with respect to the belief $\mu \in \mathcal{C}$. Our goal in this step of the proof is to show that f must satisfy Property M which we define below.

Let P be a preference profile, let i be an individual and let P'_i be an ordering such that the top k elements in P_i coincide with the top k elements of P'_i. Then Property M requires that if $f(P)$ is one of the top k elements of P_i, then the $f(P'_i, P_{-i})$ must also be one of these top k elements. Formally,

Definition 8.1 *The SCF f satisfies Property M, if for all individuals i, for all integers $k = 1, 2, \cdots, m$, for all P_{-i} and for all P_i, P'_i such that $B(r_k(P_i), P_i) = B(r_k(P'_i), P'_i)$, we have*

$$[f(P_i, P_{-i}) \in B(r_k(P_i), P_i)] \Rightarrow [f(P'_i, P_{-i}) \in B(r_k(P'_i), P'_i)]$$

In order to establish Step 2, we first need to prove an intermediate result.

Let $\mu \in \Delta^I$ and $i \in I$. Then μ_{-i} and μ_i denote respectively, the induced (conditional) probability over preferences of individuals other than i and the marginal distribution over i's preferences. Thus $\mu_{-i}(P_{-i})$ will denote the probability that individuals other than i have preferences P_{-i}. Similarly $\mu_i(P_i)$ will denote the probability that i's preference is P_i.

Lemma 8.1 Let $\mu \in \mathcal{C}$. Then, for all for all $Q, T \subset I\!\!P^{N-1}$

$$[\mu_{-i}(Q) = \mu_{-i}(T)] \Rightarrow [Q = T]$$

PROOF: Suppose not. Then there exists $\mu \in \mathcal{C}$ and $Q, T \subset I\!P^{N-1}$ with Q and T distinct such that $\mu_{-i}(Q) = \mu_{-i}(T)$. Pick an ordering for individual i, P_i and observe that

$$\mu_{-i}(Q)\mu_i(P_i) = \mu_{-i}(T)\mu_i(P_i)$$

which implies that

$$\mu(Q \times \{P_i\}) = \mu(T \times \{P_i\})$$

Since Q and T are distinct $Q \times \{P_i\}$ and $T \times \{P_i\}$ are also distinct. But this contradicts the assumption that $\mu \in \mathcal{C}$. ■

We now complete the proof of Step 2. Let i be an individual and let P_i and P_i' be such that $B(r_k(P_i), P_i) = B(r_k(P_i'), P_i')$. Suppose i's "true" preference is P_i. Since f is OBIC with respect to μ, we have, by using equation (2)

$$\begin{aligned} &\mu(\{P_{-i}|f(P_i, P_{-i}) \in B(r_k(P_i), P_i)\}) \\ \geq\ \ &\mu(\{P_{-i}|f(P_i', P_{-i}) \in B(r_k(P_i), P_i)\}) \end{aligned} \tag{12}$$

Suppose i's "true" preference is P_i'. Applying equation (2), we have

$$\begin{aligned} &\mu(\{P_{-i}|f(P_i', P_{-i}) \in B(r_k(P_i'), P_i')\}) \\ \geq\ \ &\mu_i(\{P_{-i}|f(P_i, P_{-i}) \in B(r_k(P_i'), P_i')\}) \end{aligned} \tag{13}$$

Since $B(r_k(P_i), P_i) = B(r_k(P_i'), P_i')$, equations (12) and (13) imply,

$$\begin{aligned} &\mu_i(\{P_{-i}|f(P_i, P_{-i}) \in B(r_k(P_i), P_i)\}) \\ =\ \ &\mu_i(\{P_{-i}|f(P_i', P_{-i}) \in B(r_k(P_i'), P_i')\}) \end{aligned} \tag{14}$$

Since $\mu \in \mathcal{C}$, it follows from Lemma 9.1 and equation (14) that

$$\{P_{-i}|f(P_i, P_{-i}) \in B(r_k(P_i), P_i)\} = \{P_{-i}|f(P_i', P_{-i}) \in B(r_k(P_i'), P_i')\} \tag{15}$$

Now suppose for some P_i, we have $f(P_i, P_{-i}) \in B(r_k(P_i), P_i)$. Then equation (15) implies that $f(P_i', P_{-i}) \in B(r_k(P_i'), P_i')$. Thus Property M is satisfied and Step 2 is complete.

STEP 3

In this step, we show that in a two-person SCF which satisfies Property M must be dictatorial. Let $N = \{1, 2\}$ and let f satisfy Property M.

CLAIM B: For all profiles (P_1, P_2), either $f(P_1, P_2) = r_1(P_1)$ or $f(P_1, P_2) = r_1(P_2)$ must hold.

Suppose that the Claim is false. Let (P_1, P_2) be a profile where individual 1's first-ranked alternative is a, individual 2's first-ranked alternative is b and suppose $f(P_1, P_2) = c$ where c is distinct from a and b. Consider an ordering $\bar{P}_2$ where a is ranked first and b is ranked second. By unanimity, $f(P_1, \bar{P}_2) = a$. Consider an ordering P_2' where b is ranked first and a second. Observe that the top two elements in the orderings $\bar{P}_2$ and P_2' coincide. Moreover, $f(P_1, \bar{P}_2)$ is one of these top two elements. It follows therefore from Property M that $f(P_1, P_2') \in \{a, b\}$. Now suppose that $f(P_1, P_2') = b$. Since P_2 and P_2' have the same top element, Property M implies that $f(P_1, P_2) = b$ which contradicts our supposition that the outcome at this profile is c. Therefore $f(P_1, P_2') = a$.

Let P_1' be an ordering where a and b are ranked first and second respectively. Since P_1 has the same top element as P_1' (which is a), Property M also implies that $f(P_1', P_2') = a$.

Now consider the profile (P_1', P_2). By considering an ordering $\bar{P}_1$ where b is ranked first and a second, we can duplicate an earlier argument to conclude that $f(P_1', P_2)$ is either a or b. But if it is b, then Property M would imply that

$f(P_1', P_2') = b$ which would contradict our earlier conclusion that the outcome at this profile is a. Therefore $f(P_1', P_2) = a$. But then Property M would imply that $f(P_1, P_2) = a$ whereas we have assumed that the outcome at this profile is c. This proves the Claim.

CLAIM C: If f picks 1's first-ranked alternative at a profile where 1 and 2's first-ranked outcomes are distinct then f picks 1's first-ranked alternative at all profiles.

Let (P_1, P_2) be a profile where the first-ranked alternatives according to P_1 and P_2 are a and b respectively. It follows from Claim B that $f(P_1, P_2)$ is either a or b. Assume without loss of generality that it is a. Holding P_2 fixed, observe that the outcome for all profiles where a is ranked first for 1 must be a, otherwise Property M will be violated. By a similar argument, holding P_1 fixed, the outcome b can never be obtained in all those profiles where 2's top-ranked outcome is b. Now consider an arbitrary profile where a is ranked first for 1 and b for 2. Using Claim B and the arguments above, it follows that the outcome must be a.

Consider an outcome c distinct from a and b. In view of the arguments in the previous paragraph, we can assume without loss of generality that c is second-ranked under P_1. Let P_1' be an ordering where c and a are first and second ranked respectively. Property M implies that $f(P_1', P_2)$ is either a or c. But Claim B requires the outcome at this profile to be either b or c. Therefore $f(P_1', P_2) = c$. Applying the arguments in the previous paragraph, it follows that f always picks 1's first-ranked alternative whenever 2's first-ranked alternative is b.

Let (P_1, P_2) be a profile where a and b are first-ranked in P_1 and P_2 respectively. Pick an alternative x distinct from a and b. Applying earlier arguments, we can assume that x is second-ranked in P_2. Let P_2' be an ordering where x is first and

b is second ranked. It follows from Claim B that $f(P_1, P_2')$ is either x or a. But if it is x Property M would imply that $f(P_1, P_2)$ would either be b or x which we know to be false. Therefore $f(P_1, P_2') = a$. Replicating earlier arguments, it follows that the outcome at any profile is 1's first-ranked altrenative provided that 2's first-ranked alternative is x. Since x is arbitrary, the Claim is proved.

It follows immediately from Claim C that f must be dictatorial. Therefore Step 3 is complete.

STEP 4

We now complete the induction step. Pick an integer N with $N > 2$. We assume the following:

For all K with $K \leq N$, if $f : I\!P^K \to A$ satisfies Property M, then f is dictatorial.

Our goal is to prove:

If $f : I\!P^N \to A$ satisfies Property M then f is dictatorial.

Let $f : I\!P^N \to A$ be a SCF that satisfies Property M. Define a SCF $g : I\!P^{N-1} \to A$ as follows. For all $(P_1, P_3, P_4, \cdots, P_N) \in I\!P^{N-1}$,

$$g(P_1, P_3, P_4, \cdots, P_N) = f(P_1, P_1, P_3, \cdots, P_N)$$

The idea behind this construction is simple and appears frequently in the literature on strategy-proofness, for example in Sen (2001). Individuals 1 and 2 are "cloned" to form a single individual in the SCF g. This coalesced individual in g will be referred to as $\{1, 2\}$.

It is trivial to verify that g satisfies unanimity. We will show that g satisfies Property M. Pick an individual i and suppose P_i and P_i' are such that $B(r_k(P_i), P_i) = B(r_k(P_i'), P_i')$ for some integer k which lies between 1 and m. Further, suppose that for some profile $P_{-i} \in I\!P^{N-2}$, we have $g(P_i, P_{-i}) \in B(r_k(P_i), P_i)$. We will show that $g(P_i', P_{-i}) \in B(r_k(P_i'), P_i')$. Observe that if i is an individual from the set $\{3, \cdots, N\}$, then this follows immediately from our assumption that f satisfies Property M. The only non-obvious case is the one where i is the coalesced individual $\{1, 2\}$. In this case, observe that since f satisfies Property M, $f(P_1, P_1, P_3, \cdots, P_N) \in B(r_k(P_i), P_i)$ implies that $f(P_1', P_1, P_3, \cdots, P_N) \in B(r_k(P_i'), P_i')$ which in turn implies that $f(P_1', P_1', P_3, \cdots, P_N) \in B(r_k(P_i'), P_i')$. Therefore, $g(P_1', P_3, \cdots, P_N) \in B(r_k(P_i'), P_i')$ which is what was required to be proved.

Since g satisfies Property M, our induction assumption implies that g is dictatorial. There are two cases which will be considered separately.

CASE I: The dictator is the cloned individual $\{1, 2\}$. Thus whenever individuals 1 and 2 have the same preferences, the outcome under f is the first-ranked alternative according to this common preference ordering.

Fix an $N-2$ person profile $(P_3, P_4, \cdots, P_N) \in \mathbb{P}^{N-2}$ and define a two-person SCF $h : \mathbb{P}^2 \to A$ as follows: for all $(P_1, P_2) \in \mathbb{P}^2$,

$$h(P_1, P_2) = f(P_1, P_2, P_3, \cdots, P_N)$$

Since $\{1, 2\}$ is a dictator, h satisfies unanimity. Since f satisfies Property M, it follows immediately that h also satisfies Property M. From Step 3, it follows that h is dictatorial. Assume without loss of generality that this dictator is 1. We now show that 1 is a dictator in f. In other words, the identity of the dictator in h does not depend on $(P_3, P_4, \cdots, P_N)$.

Let $j \in \{3, 4, \cdots, N\}$ and suppose that there exists an $N-2$ person profile $(P_1, \cdots, P_N)$ where j can change the identity of the dictator in h (say from 1 to 2) by changing his preferences from P_j to P_j'. We shall show that this is not possible when P_j and P_j' differ only over a pair of alternatives. This is sufficient to prove the general case because the change from P_j to P_j' can be decomposed into a sequence of changes where successive preferences along the sequence differ only over a pair of alternatives. Assume therefore that there exists a pair x, y such that $r_k(P_j) = x$, $r_{k+1}(P_j) = y$ and $r_k(P'_j) = y$, $r_{k+1}(P'_j) = x$. Moreover for any alternative z distinct from x and y, its rank in P_j and P'_j is the same. Consider the profile $P = (P_1, P_2, P_3, \cdots, P_j, \cdots, P_N)$ where P_1 and P_2 have distinct first-ranked alternatives. Then individual j by switching from P_j to P'_j changes the outcome. Observe that P_j and P'_j have the same top s elements where $s = 1, 2, \cdots, k-1, k+1, \cdots, m$. Since f satisfies Property M, it follows that $f(P)$ and $f(P'_j, P_{-j})$ can differ only if $f(P) = f(P_j, P_{-j}) \in \{x, y\}$. But $f(P_j, P_{-j}) \in \{x, y\}$ implies that $f(P'_j, P_{-j}) \in \{x, y\}$. The above statement again follows from the fact that f satisfies Property M. Now pick P_1 and P_2 such that the first-ranked alternatives in these two orderings is x and z respectively where z is distinct from x and y. Since j changes the identity of the dictatator in h from 1 to 2, it follows that $f(P'_j, P_{-j}) = z$ which contradicts our earlier claim that $f(P'_j, P_{-j}) \in \{x, y\}$. Therefore j cannot change the identity of the dictator in h by changing his preferences. Therefore the dictator in h is the dictator in f.

CASE II: The dictator in g is an individual $j \in \{3, \cdots, N\}$. Assume without loss of generality that $j = 3$. Now define a $N-1$ person SCF g' by coalescing individuals 1 and 3 rather than 1 and 2 as in g. Of course, g' satisfies unanimity and Property M. Therefore it is dictatorial (by the induction hypothesis). If the dictator is the coalesced individual $\{1, 3\}$, then Case I applies and we can conclude that f is dictatorial. Suppose therefore that $\{1, 3\}$ is not the dictator. We will show that this is impossible. We consider two subcases.

CASE IIa: The dictator in g' is an individual $j \in \{4, \cdots, N\}$. Assume without loss of generality that $j = 4$. In this subcase, when 1 and 2 have the same preferences, the outcome under f is 3's first-ranked alternative but when 1 and 3 agree, the outcome is 4's first-ranked alternative. Consider an N person profile P

where $P_1 = P_2 = P_3$. Let a be the first-ranked alternative of this ordering. Let the first ranked alternative in P_4 be b which is distinct from a. Since 1 and 2's orderings coincide, $f(P)$ must be individual 3's first-ranked alternative which is a. On the other hand, since 1 and 3's orderings coincide, $f(P)$ must be individual 4's first ranked alternative which is b. We have a contradiction.

CASE IIB: The dictator in g' is individual 2. Let P be an N-person profile where $P_1 = P_3$ and $aP_1bP_1cP_1x$ for all $x \neq a, b, c$. Also let $bP_2aP_2cP_2x$ for all $x \neq a, b, c$ and let P_2 agree with P_1 for all $x \neq a, b, c$ Since 1 and 3 have the same ordering in P, $f(P) = b$. Let P_3' be the ordering obtained by switching b and c in P_3. Since P_3 and P_3' agree on the top and the top three elements, Property M implies that $f(P_3', P_{-3}) \in \{b, c\}$. Suppose that this outcome is c. Then observe that Property M implies that $f(P_1, P_1, P_3', \cdots, P_N) = c$. But since 1 and 2's orderings coincide, the outcome at this profile should be 3's first-ranked alternative a. Therefore $f(P_3', P_{-3}) = b$. Now let $\bar{P}_3$ be the ordering obtained by switching a and c in P_3'. Property M implies that $f(P_1, P_2, \bar{P}_3, \cdots, P_N) = b$. A further application of Property M for individual 2 allows us to conclude that $f(P_1, P_1, \bar{P}_3, \cdots, P_N) \in \{a, b\}$. But 1 and 2 have the same ordering at this profile so that the outcome here must be 3's first-ranked alternative which is c. We have obtained a contradiction.

This concludes Step 4 and the proof of the Theorem. ■