Essays in Social Choice Theory

Dipjyoti Majumdar

Thesis submitted to the Indian statistical Institute in partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy

Essays in Social Choice Theory

Dipjyoti Majumdar

October, 2002

Thesis Supervisor : Professor Arunava Sen

Thesis submitted to the Indian statistical Institute in partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy

To My Parents

"... I have nothing but a book,

Nothing but that to prove your blood and mine."

— W. B. Yeats, *Responsibilities*

Acknowledgements

First and foremost, my supervisor, Professor Arunava Sen. For all that he has given me over the last five years, words are not enough. I will not even attempt to thank him here. I only regret that, many a time, I have not been able to meet his expectations.

My teachers at the Indian Statistical Institute from whom I learnt so much: Professors Bhaskar Dutta, Shurojit Chatterjee, Satya P. Das, Sanghamitra Das, Shubhashis Gangopadhay and Bharat Ramaswami. Special thanks to Professor Bhaskar Dutta without whose invaluable suggestions this thesis would not have come to its present shape. I extend my sincere thanks to Professors Arup Pal and Rahul Roy from the Stat-Math Unit for their invaluable suggestions.

I am grateful to Anirban Kar and Shubhadip Ghosh for helping me in a number of ways. Dwiraj Bose, Paromita Saha and Saraswata Choudhury helped me in a number of computer related problems. I also thank Kunal Dasgupta who has gone through an earlier version of this work and pointed out a number of slips. Thanks to all the hostellers at Delhi ISI, who have made my stay an memorable one.

Finally, I thank my long time friend Runsi for her support and friendship through all these years when I was pursuing my research.

A research fellowship from the Indian Statistical Institute is gratefully acknowledged.

Contents

Chapter 1

Introduction

The purpose of this thesis is to explore some issues in social choice theory and decision theory. Social choice theory provides the theoretical foundations for the field of public choice and welfare economics. It tries to bring together normative aspects like perspective value judgements and positive aspects, like strategic considerations. The second feature which is our focus, is closely related to the problem of providing appropriate incentives to agents, an issue of prime importance in economics.

Consider for example, a set of agents who must elect one among a set of candidates. These candidates may be physical agents or they may be issues such as various economic policies. A voting institution may be thought of as a procedure which selects an outcome or candidate for every profile of voter preferences over candidates. It has long been recognized that a voting institution will typically offer opportunities for some voters to behave strategically. A situation may arise where some voter may find it in his best interest to vote for a candidate other than his most preferred one for doing so changes the final outcome favourably for him. Let us consider a second example. Suppose there is a set of agents whose members are to be matched to members of a second, disjoint set of agents all of whom have preferences over the possible resulting matches. Examples include matching students with universities, men with women, workers with firms etc. Agents report their preferences over possible mates and are then matched to a mate according to some procedure. Here too situations may arise where some agent may find it in his/her best interest to misreport his/her preference. Incentives problems of this nature are pervasive in economic contexts.

The first two chapters of the thesis are concerned with mechanism design issues in two different settings. Chapter 2 considers the classical strategic voting model where a voter's preference ordering over a set of candidates is private information. Chapter 3 considers the familiar two-sided matching model where an agent's preference over his/her possible mates is private information. In both models the central issue is the design of mechanisms or procedures which provide appropriate incentives for agents to reveal their private information truthfully. In the matching model, an additional objective is to ensure that outcomes are always *stable*. There is an extensive literature pertaining to these issues where attention is focused on mechanisms where agents have dominant strategy incentives to tell the truth. This requirement is strong; as a consequence most results are negative. In these two essays we explore the implications of weakening the truth-telling requirement to *ordinal Bayesian incentive compatibility.*

A notion that appears frequently in social choice theory is Maskin-monotonicity. A social choice function satisfies this axiom if it is monotonic with respect to an alternative which improves in a voter's preference ordering. Not only is this axiom normatively appealing but also the key to some important strategic properties. Moreover, it has been shown that if the domain of preferences is unrestricted, Maskin-monotonicity is equivalent to the property of strategy-proofness which in turn implies that these properties are equivalent to dictatorship. In Chapter 4 of this thesis, we formulate a version of monotonicity which is based on improvements of sets of alternatives rather than an alternative. This allows us to identify precisely the monotonicity properties which precipitate dictatorship.

The final essay in the thesis addresses an issue in decision theory. A decision maker is assumed to have beliefs over the possible states of the world. This is a probability distribution over these states, or geometrically, a point in the unit simplex of appropriate dimension. Once new information is made available, these beliefs have to be revised according to some *revision rule*. The main result of this chapter is a characterization of Bayes' rule as a revision rule, employing axioms which are widely used in axiomatic bargaining theory.

We now discuss each of the essays in greater detail.

1.1 Ordinal Bayesian Incentive Compatibility and the Strategic Voting Model

In the typical social choice problem there is a group of agents or voters, a "planner" and a set of outcomes or alternatives. The role of the "planner" (who may not be a "real" agent but simply represent a procedure or a decision making process) is to select an outcome from the feasible set of outcomes. Thus, a collective decision has to be made in a situation where each agent has some information about the "environment", not known to the other agents and to the planner. The latter seeks to achieve certain goals which depends on the information of each agent. This problem is known in the literature as the problem of *mechanism design*. The literature in this area is huge and dates back to the work of Hayek, Lange and Lerner in the 1930s. However it was the work of Hurwicz in the 1950s and 1960s which formalized the insights of Hayek, Lange and Lerner and paved the way for the body of work that followed his pioneering effort.

In the strategic version of this problem, each agent has a preference ordering over the set of feasible outcomes that is *not* known to the other agents or to the planner. The agents report their preferences to the planner. The task of the planner is to pick an outcome given the *reported* preference profile. Thus the selection of the outcome depends on the information that the agents hold. The objective of the planner is represented by a social choice function that associates a feasible outcome with each profile of reported preference orderings. Since the outcome that is selected depends on the reports sent by agents, they realize that they can influence the outcome by changing their report. The mechanism design problem is one of selecting a social choice function that will give the agents incentives to reveal their private information truthfully. The most appealing concept is *strategy-proofness* which requires truth-telling for each agent to be a dominant strategy. If a social choice function is strategy-proof, each agent does at least as well by misreporting as by telling the truth irrespective of his beliefs about what announcements the other agents will make. Unfortunately, while strategy-proofness is a very appealing requirement it is also very stringent and leads to an impossibility result. This is the celebrated Gibbard-Satterthwaite Theorem (Gibbard (1973), Satterthwaite (1975)) which states that under mild assumptions, a social choice function is strategy-proof only if it is *dictatorial.* A dictatorial social choice function is one which always selects the maximal element of a particular agent at all preference profiles.

The negative conclusions of the Gibbard-Satterthwaite Theorem have inspired research in several directions. Perhaps the most fruitful of these has been the exploration of *domain restrictions* which permit the existence of strategy-proof social choice functions satisfying appealing normative properties (in particular non-dictatorship). For instance, in a domain of single-peaked preferences, strategy-proofness is compatible with anonymity and efficiency. (Moulin (1981)). If monetary transfers are introduced in a quasi-linear environment, several interesting possibility results emerge including the rich theory of Groves-Clarke-Vickrey mechanisms (Green and Laffont (1977)). If social choice functions are allowed to pick random outcomes over lotteries and preferences over lotteries are required to satisfy von-Neumann-Morgenstern axioms, the link between strategy-proofness and dictatorship can also be broken (Gibbard(1977)).

We take the relatively less explored approach of weakening the truth-telling requirement. Instead of insisting that truth be a dominant strategy for every agent we require that truth-telling be optimal on average or in expectation. Expected utilities are computed with respect to an agent's prior beliefs about the preferences of other agents and based on the assumption that other agent's will tell the truth. More formally, truth-telling is required to be a Bayes-Nash equilibrium of the revelation game where an agent's type is identified with his preference ordering.

There is a fairly extensive literature on mechanism design using Bayes-Nash equilibrium as a solution concept in the revelation game. However, inspired by the work of d'Aspremont and Gérard-Varet (1979), almost all of it pertains to models where there is money which can be transferred between agents and the planner and where preferences are assumed to be quasi-linear. Since we wish to apply this notion to voting environments, there is another issue which needs to be resolved. Individual preferences are ordinal (rankings over candidates) so there is no "natural" utility function which can be used for expected utility calculations. We assume therefore that truth-telling dominates misrepresentation in terms of expected utility where the latter is evaluated for all utility functions which represent the true preference ordering. This notion was proposed initially by d'Aspremont and Peleg (1988) in their study of committee representation and referred to as *ordinal Bayesian incentive compatibility.*

The ordinal Bayesian incentive compatibility of a social choice function is verified with respect to given (prior) beliefs of each agent. Since strategy-proofness requires this incentive compatibility condition satisfied for all beliefs, ordinal Bayesian incentive compatibility is clearly weaker than strategy-proofness. A priori, we may expect a large class of social choice functions to satisfy ordinal Bayesian incentive compatibility. However, our main result demonstrates that this is not true. Ordinal Bayesian incentive compatibility with respect to *any* belief chosen from a set which is generic in an appropriate sense, implies dictatorship. On the other hand possibility results exist if beliefs are non-generic. In particular, we show that if beliefs are uniform, a wide class of social choice functions satisfy ordinal Bayesian incentive compatibility.

We would like to distinguish our negative result from those which obtain when robustness requirements are imposed on the problem (see, for instance, Ledyard (1978)). Here, typically incentive compatibility of a social choice function is required for a *set* of beliefs. Our result on the other hand holds for any beliefs in a (suitably defined) large set.

1.2 Ordinal Bayesian Incentive Compatibility and Stable Matchings

In the second essay of this thesis we look at two-sided matching model. *Matching problems* refer to the whole gamut of problems which involve matching members of one set of agents to members of a second, disjoint set of agents all of whom have preferences over all possible matches. We concentrate on two-sided *one-to-one* matchings, also known in the literature as the *marriage problem.* One set of agents is then described as the set of *men* and the other set is referred to as the set of *women.* In this context, the main interest is in finding *stable matching procedures* which can be defined as a matching of agents such that no pair of agents would prefer to be matched to each other than to their current partners.

Research on stable matchings has taken basically two separate courses. One line has concentrated on the structure of stable matchings and the computation of efficient algorithms. This literature owes its genesis to the paper of Gale and

Shapley (1962) which showed that the set of stable matchings is non-empty for any matching problem. Their proof was constructive – they provided an algorithm called the deferred acceptance algorithm which always generates stable matchings.

The other strand in the literature on stable matchings is concerned with the strategic issues involved in a matching game. Suppose that there is a centralized authority to whom the agents report their preferences. This authority then selects a stable matching according to some procedure. A question of fundamental interest is the following: does there exist a strategy-proof social choice function (which, in this setting, associates a matching with every profile of preferences) which always selects a stable matching? Roth (1982) answers this question in the negative.

Roth (1989) extends the analysis of Roth (1982) to the case where truth-telling is a Bayes-Nash equilibrium of the revelation game. However he assumes a particular cardinalization of utilities. We study the case where truth-telling requirement is ordinally Bayesian incentive compatible. We consider this more appropriate because stable matchings only consider preferences of agents and not utility specifications.

Our first result is that there does not exist *any* prior belief with the property that there exists a stable matching procedure which is ordinally Bayesian incentive compatible with respect to that prior belief. Our next step is to impose the mild domain restriction that all agents prefer to be matched rather than remain single. These restrictions are not sufficient to guarantee a stable, strategy-proof matching procedure (see Alcalde and Barberà (1994)). However they are sufficient to ensure that both man and woman proposing deferred acceptance algorithms are ordinally Bayesian incentive compatible with respect to the uniform prior. Indeed the result follows immediately from and can be considered to be an equilibrium interpretation of a result in Roth and Rothblum (1999). Our main result in this essay demonstrates that this result is non-generic. Using techniques developed in the previous essay we show the following: if each agent i's prior beliefs are independently distributed, then there exists a set of conditional beliefs C_i which is generic in the set of conditional beliefs (generated under the independence assumption) such that there does not exist *any* stable matching procedure which is ordinally Bayesian incentive compatible with respect to any belief which generates a conditional belief in C_i.

1.3 Monotonicity and Dictatorship

The third essay in the thesis (Chapter 4) takes a closer look at the relationship between the axiom of Maskin monotonicity and dictatorship. Let f be a social choice function and let i be an agent. Consider an alternative x. Let P_i and P_i' be orderings such that x beats all the alternatives in P_i' that it beats in P_i (i.e., $xP_iz \rightarrow xP_i'z$ for all $z \neq x$). Let P_{-i} be a preference profile for all agents other than i such that $f(P_i, P_{-i}) = x$. Maskin monotonicity requires that $f(P_i', P_{-i}) = x$. In other words, f satisfies a monotonicity requirement with respect to "improvements" of an alternative in the preference ordering of an agent. Maskin monotonicity is an appealing normative requirement. In addition, Maskin (1999) has established a fundamental connection between the strategic property of Nash implementability and Maskin monotonicity. It is however, a strong requirement. Muller and Satterthwaite (1977) established the equivalence of Maskin monotonicity and strategy-proofness over the complete domain of preferences. Applying the Gibbard-Satterthwaite Theorem it follows immediately that a social choice function satisfies Maskin monotonicity if and only if it is dictatorial.

In this essay we introduce a new monotonicity axiom. Suppose there are m alternatives. Let t be an integer lying between 1 and m. We say that a social choice function satisfies *Top t-monotonicity* if the following condition is satisfied: for all profiles (P_i, P_{-i}) and (P_i', P_{-i}) such that the set of the top k (where k is an integer $1 \leq k \leq t$) ranked alternatives in P_i (let us refer to this set as B), coincide with the top k ranked alternatives in P_i', then $f(P_i, P_{-i}) \in B$ implies $f(P_i', P_{-i}) \in B$. This is a notion of monotonicity with respect to sets rather than alternatives. The set B (weakly) improves in P_i' relative to P_i. Thus if the outcome in (P_i, P_{-i}) is in the set B, it must also be so in the profile (P_i', P_{-i}).

Our first result establishes that Top $(m-1)$-monotonicity is implied by Maskin monotonicity. The next result shows that Top 2-monotonicity (referred to as Top Pair Monotonicity) in conjunction with the property of unanimity is sufficient to force a social choice function to be dictatorial when there are only two agents. We show by means of an example that if there are three or more agents, then there are non-dictatorial, unanimous social choice functions which satisfy Top Pair monotonicity. However, if Top Pair monotonicity is strengthened to Top 3-monotonicity (or to Top Triple monotonicity), then we obtain dictatorship once again. Both results illustrate the fact that conditions far weaker than Maskin monotonicity precipitate dictatorship.

1.4 A Characterization of Bayes' Rule

The last essay in this thesis (Chapter 5) provides an axiomatic characterization of Bayes' rule. Bayes' rule is a method for updating the beliefs of an agent. This method is used widely in statistics and decision theory. There are several existing characterizations of Bayes' rule. However, most of them are from a no-arbitrage perspective. The arbitrage principle has a long history. In the literature on Bayesian statistics and decision theory, it was introduced as an axiom by de Finetti (1974) for characterizing subjective probability. More recently the "arbitrage principle" has been proposed as a foundation for non-cooperative game theory through its dual relationship with the concept of correlated equilibrium (McCardle and Nau (1990), Nau (1991)). McCardle and Nau (1991) tries to unify decision theory, market theory and game theory by appealing to no-arbitrage principle. However in all these settings money plays a crucial role. In environments where money is available as a medium of exchange and measurement, no-arbitrage is synonymous with subjective utility maximization in personal decisions. In this essay, we axiomatize Bayes' Rule without introducing money in the model.

Bayes' Rule is viewed in this essay as a *revision rule*. The task for a revision rule is to assign posterior probabilities given a prior belief and given some new information. We characterize Bayes' Rule by imposing axioms on the revision rule. The paper which is closest in spirit to our analysis is Rubinstein and Zhou (1999). They consider a general decision situation where an agent chooses a point in a convex set S given some reference point e. In such a general setting they characterize the choice rule that picks the point in S that is closest to e. However, since their environment is general, their axioms are strong and inappropriate for the more specialized problem of belief revision. Observe for instance, that the set of beliefs is not any general convex set but the unit simplex of appropriate dimension.

The critical axiom is *Path Independence*. This axiom requires that the posterior belief should not be affected by the order in which new information appears. For instance, suppose that the prior belief has placed positive probabilities over some set T. Some new information arrives that rules out say, the states $t_1, t_2 \in T$. The revision rule then assigns probabilities over the elements in $T \setminus \{t_1, t_2\}$. Now suppose a second new piece of information comes in that rules out $t_3, t_4 \in T \setminus \{t_1, t_2\}$. The revision rule would now assign the posterior over the set $T \setminus \{t_1, t_2, t_3, t_4\}$. Suppose instead of the information coming in steps the initial information had ruled out the states t_1, t_2, t_3 and t_4. Path independence says that the posterior distribution over the set $T \setminus \{t_1, t_2, t_3, t_4\}$ should remain the same in both cases.

The other axioms are quite standard and have been motivated by the axioms used in the axiomatic theory of surplus sharing. One is a symmetry (or anonymity) axiom which requires that the names of the states of the world are not material for the revision rule. The continuity axiom requires the revision rule to be continuous with respect to the prior. The monotonicity axiom requires that the revised probability on a state should not be less than the prior on that state. This axiom is intuitive. A revision rule redistributes the probability weights on the states that have been ruled out over the states that remain. So the probability weight on any existing state should not go down after the revision. Finally a "no mistake hypothesis" is imposed which requires that if an agent believes initially that the occurrence of a particular state is impossible, then she continues to believe this even after the arrival of new information. (actually this axiom is required only in the very special case where a revision eliminates all but only two states of the world). Detailed discussions of the axioms can be found in the chapter.

The problem that we consider is similar to the structure of the so called "bargaining problem with claims" (Chun and Thomson(1992)). The problem in that context is a triple (S, e, c) with the interpretation that S is the set of feasible utility vectors, e is the disagreement point and c is the "claims" point. Chun and Thomson characterize the proportional solution whose functional form is the same as Bayes' Rule. Chun (1988) also characterizes the proportional solution in the context of the *bankruptcy problem* using a stronger version of the Pareto optimality criterion — the *No Advantageous Reallocation.* Both the models emphasize the utility interpretation of choices and as a consequence Pareto optimality or its stronger version, the no advantageous reallocation is imposed as an axiom. However, our model does not have a utility interpretation so that these assumptions are meaningless in our context. Therefore our result, though similar in spirit is very different from theirs, in interpretation and substantive details.

Chapter 2

Ordinally Bayesian Incentive Compatible Voting Schemes

2.1 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 essay 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, modelled 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 result. Suppose all voters have beliefs generated from a common prior which is independently distributed. Then, for each voter i, there exists a set of conditional beliefs C_i such that, any social choice function is ordinally Bayesian incentive compatible with respect to *any* belief whose conditionals lie in the set C_i, $i = 1, \cdots, N$, only if it is dictatorial. Moreover, the set C_i is "generic" in the set of all conditional beliefs generated under the independence assumption, i.e., it is open and dense in this set and its complement set has Lebesgue measure zero. 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 (provided independence holds), the much 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 essay is organized as follows. In Section 2.2 we set out the basic notation and definitions. In Sections 2.3 and 2.4, we consider respectively the generic case and the case of uniform priors. Section 2.5 concludes. The proof of the main result is contained in the Appendix.

2.2 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 2.2.1 *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 this essay is that we assume SCFs to be *ordinal*. In other words, the only information 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 essay, we assume that SCFs under consideration satisfy the axiom of *unanimity*. 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 2.2.2 *A SCF f is unanimous if $f(P) = a_j$ whenever $a_j = r_1(P_i)$ for all individuals $i \in N$.*

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

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 2.2.3 *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 2.2.4 *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 **2.2.1** *Gibbard (1973), Satterthwaite (1975)*

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

in this essay, 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 2.2.5 *A belief for an individual i is a probability distribution on the set $\mathbb{P}^N$, i.e., it is a map $\mu_i : \mathbb{P}^N \rightarrow [0,1]$ 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 $m!^N - 1$ which we denote by $\tilde{\Delta}$. 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 2.2.6 *The utility function $u : A \to \Re$ represents $P_i \in I\!P$, 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 essay.

DEFINITION 2.2.7 *A SCF f is Ordinally Bayesian Incentive Compatible (OBIC) with respect to the belief μ if for all $i \in N$, for all $P_i, P_i' \in I\!P$, for all $u \in \mathcal{U}(P_i)$, we have*

$$\sum_{P_{-i} \in I\!P^{N-1}} u\left(f(P_i, P_{-i})\right) \mu(P_{-i}|P_i) \geq \sum_{P_{-i} \in I\!P^{N-1}} u\left(f(P_i', P_{-i})\right) \mu(P_{-i}|P_i) \qquad (2.1)$$

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 $I\!P$ 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 SCFs under consideration are ordinal by assumption, there is no "natural" utility function for expected utility 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 system. We record without proof the precise relationship between the two concepts below.

REMARK 2.2.1 A SCF is strategy-proof if and only if 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 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 truth-telling 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 2.2.8 *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',*

$$\begin{aligned} &\mu(\{P_{-i}|f(P_i, P_{-i}) \in B(r_k(P_i), P_i)\}|P_i) \\ \geq \ &\mu(\{P_{-i}|f(P_i', P_{-i}) \in B(r_k(P_i), P_i)\}|P_i) \end{aligned} \tag{2.2}$$

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).

2.3 The Generic Case

The main result of this section is to show that generically, OBIC implies dictatorship. However, we need to make a crucial assumption regarding admissible beliefs.

DEFINITION 2.3.1 *Individual i's beliefs are independent if for all $k = 1, \cdots, N$ there exist probability distributions $\mu_k : I\!P \to [0, 1]$ such that for all P*

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

An individual's belief is independent if his belief is a product measure of the marginals over the types of all the individuals. 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 $m!-1$. If μ is the common prior for all voters, the conditional beliefs for voter i will be denoted by μ_{-i}, i.e., $\mu_{-i}(P_{-i}) = \times_{k \neq i} \mu_k(P_k)$. We denote the set of conditional beliefs by Δ^{CI} which is the $N-1$-th order Cartesian product of the unit simplices Δ. Clearly $\mu_{-i} \in \Delta^{CI}$.

We can now state the main result of this section.

THEOREM 2.3.1 *Let $m \geq 3$ and assume that all individuals have independent beliefs. Then for all $i \in N$, there exists a subset C_i of Δ^{CI} such that*

- *C_i is open and dense in Δ^{CI}*
- *$\Delta^{CI} - C_i$ has Lebesgue measure zero*
- *if f is unanimous and is OBIC w.r.t the belief μ where $\mu_{-i} \in C_i$ for all $i \in N$, then f is dictatorial.*

The proof of the Theorem is contained in the Appendix. Here we only describe briefly the construction of the sets C_i in order to clarify the nature of generic beliefs.

For any $Q \subseteq \mathbb{P}^{N-1}$, let $\mu(Q) = \sum_{P_{-i} \in Q} \mu_{-i}(P_{-i})$. The set C_i is defined as the set of conditional beliefs μ_{-i} satisfying the following property: for all $Q, T \subseteq \mathbb{P}^{N-1}$

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

For any belief μ and agent i the conditional belief μ_{-i} belongs to C_i if it assigns equal probabilities to two "events" Q and T only if $Q = T$. Obviously the events Q and T are defined over the preference orderings of individuals other than i. The first step in the proof is to show the sets C_i are open and dense in Δ^{CI} and that its complement set has Lebesgue measure zero. Observe that C_i is generic in the space of *conditional probabilities generated by an independent prior.* It is *not* generic in the space of all probability distributions. The next step in the argument is to show that any SCF which is OBIC with respect to a belief μ where each individual i's conditional beliefs (generated from μ) belong to the set C_i must satisfy a certain property which we call Property M. The final step in the argument consists in showing that an SCF which satisfies Property M must be dictatorial. This is accomplished by induction on the number of individuals starting with the case $N = 2$.

Remark 2.3.1 The Gibbard-Satterthwaite Theorem is a corollary of Theorem 2.3.1. This follows immediately from Remark 2.2.1.

The result stated in Theorem 2.3.1 continues to hold if we assume "non-common" priors instead of "common" priors. Let μ_i denote the prior for individual i. We shall let $\tilde{\Delta}$ denote the set of all beliefs μ_i. As before, the set of all beliefs $\tilde{\Delta}$ is the unit simplex of dimension $m!^N - 1$. For all μ_i, for all P_{-i} and P_i, we shall let $\mu_i(P_{-i}|P_i)$ denote the conditional probability of P_{-i} given P_i. We shall also refer to an N-tuple of beliefs $(\mu_1, \cdots, \mu_N)$ as a *belief system.*

Definition 2.3.2 *A SCF f is Ordinally Bayesian Incentive Compatible (OBIC) with respect to the belief system $(\mu_1, \cdots, \mu_N)$ 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*

$$\sum_{P_{-i} \in \mathbb{P}^{N-1}} u\left(f(P_i, P_{-i})\right) \mu_i(P_{-i}|P_i) \geq \sum_{P_{-i} \in \mathbb{P}^{N-1}} u\left(f(P_i', P_{-i})\right) \mu_i(P_{-i}|P_i) \qquad (2.3)$$

In the case of "non-common" priors, a condition weaker than independence is sufficient for our result. This is the condition of "free beliefs" introduced in d'Aspremont - Gèrard-Varet (1982).

Definition 2.3.3 *Individuals i's beliefs are free if for all P_{-i} and for all P_i and P_i' we have*

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

If a player's beliefs are free, then his beliefs about the types of the other players are independent of the realization of his own type. Observe that if voters have common prior, then free beliefs implies independence. We shall denote the set of all free beliefs for individual i by $\mathcal{F}_i$.

We can now state the result.

THEOREM **2.3.2** *Let $m \geq 3$. For all $i \in N$, there exists a subset $\mathcal{M}_i$ of $\tilde{\Delta}$ such that*

- *$\mathcal{M}_i$ is open and dense in $\tilde{\Delta}$*
- *$\tilde{\Delta} - \mathcal{M}_i$ has Lebesgue measure zero*
- *if f is unanimous and is OBIC w.r.t the belief system $(\mu_1, \cdots, \mu_N)$ where $\mu_i \in \mathcal{M}_i \cap \mathcal{F}_i$ for all $i \in N$, then f is dictatorial.*

The prof of this result is contained in the Appendix.

REMARK 2.3.2: We wish to point out a subtle difference in the "spirit" of Theorems 2.3.1 and 2.3.2. In the former, we demonstrate the existence of sets $\mathcal{C}_i$ which are generic in the space of conditional distributions generated under the independence hypothesis. Any SCF which is OBIC with respect to a conditional distribution in $\mathcal{C}_i$ for all i, is dictatorial. In Theorem 2.3.2 we construct the sets $\mathcal{M}_i$ which are generic in the space of *all* beliefs. Any SCF which is OBIC with respect to a belief system where each belief belongs to $\mathcal{M}_i$ *and* satisfies free beliefs, is dictatorial. Of course we can formulate a version of Theorem 2.3.1 along the lines of Theorem 2.3.2 and vice versa. For instance, we have:

THEOREM **2.3.3** *Let $m \geq 3$. There exists a subset $\mathcal{M}$ of $\tilde{\Delta}$ such that*

- *$\mathcal{M}$ is open and dense in $\tilde{\Delta}$*
- *$\tilde{\Delta} - \mathcal{M}$ has Lebesgue measure zero*
- *if f is unanimous and is OBIC w.r.t the belief μ where $\mu \in \mathcal{M}$ and μ is independent, then f is dictatorial.*

We make a few remarks about the proof of Theorem 2.3.3 in the Appendix.

Theorems 2.3.1– 2.3.3 make it emphatically clear that in "almost all cases", weakening the incentive-compatibility requirement from strategy-proofness to OBIC does not expand the set of incentive compatible SCFs. However, the definition of "almost all cases" above leaves out at least one very important case. This is the case where an individual's beliefs over the types of others, conditional on his own type is uniform. Observe that in this case, the probabilities of events Q and T are equal whenever Q and T have the same cardinality. We analyse this case extensively in the next section and show that the results are dramatically different from the generic case.

2.4 A Non-Generic Case: Uniform Priors

In this section, we make the following further assumption on beliefs. (We continue to assume independence)

ASSUMPTION 2.4.1 For all profiles P and P' we have

$$\mu(P) = \mu(P')$$

Thus, all individual have a common prior which is uniform. We denote these beliefs by $\bar{\mu}$. As remarked in the earlier section, $\bar{\mu}$ is non-generic. Restating Definition 2.2.8 in the present context, we have

PROPOSITION 2.4.1 *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*

$$|\{P_{-i}|f(P_i, P_{-i}) \in B(r_k(P_i), P_i)\}| \geq |\{P_{-i}|f(P_i', P_{-i}) \in B(r_k(P_i), P_i\}| \qquad (2.4)$$

We omit the (trivial) proof of this Proposition. It will be convenient to express equation (2.4) in a more compact way. For all $P_i \in \mathbb{P}$ and $x \in A$, let

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

Equation (2.4) 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

$$\sum_{t=1}^{k} \eta(r_k(P_i), P_i) \geq \sum_{t=1}^{k} \eta(r_k(P_i), P_i') \tag{2.5}$$

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

EXAMPLE 2.4.1

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

$$\begin{array}{ccccccc} & abc & acb & bac & bca & cab & cba \\ abc & a & a & b & a & c & a \\ acb & a & a & a & b & a & c \\ bac & b & a & b & b & b & c \\ bca & a & b & b & b & c & b \\ cab & c & a & b & c & c & c \\ cba & a & c & c & b & c & c \end{array} \tag{2.6}$$

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 anonymous and efficient (for a definition of these terms, see Moulin (1983)). 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 occurrence of its first-ranked alternative is four and of its second and third-ranked alternatives, one each respectively.

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

DEFINITION 2.4.1 *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 symmetrically 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 2.4.2 *The SCF f satisfies elementary monotonicity if for all i, P_i, P_i' and P_{-i}*

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

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 a relatively weak axiom whose implications we will discuss more fully after stating and proving the main result of this section.

THEOREM 2.4.1 *A SCF which satisfies neutrality and elementary monotonicity is OBIC with respect to the beliefs $\bar{\mu}$.*

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,

$$f(P_i', P_{-i}^\sigma) = \sigma[f(P_i, P_{-i})] = \sigma[r_k(P_i)] = r_k(P_i') \tag{2.7}$$

Equation (2.7) 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

$$|\{P_{-i} | f(P_i, P_{-i}) = r_{k+1}(P_i)\}| \subseteq |\{P_{-i} | f(P_i', P_{-i}) = r_{k+1}(P_i)\}| \tag{2.8}$$

Equation (2.8) above implies that

$$\eta(r_{k+1}(P_i), P_i') \geq \eta(r_{k+1}(P_i), P_i) \tag{2.9}$$

But the LHS of equation (2.9) 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

$$\eta(r_k(P_i), P_i) \geq \eta(r_t(P_i), P_i) \text{ whenever } k < t \tag{2.10}$$

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,

$$\sum_{t=1}^{k} \eta(r_t(P_i), P_i') = \sum_{t \in T} \eta(r_t(P_i), P_i) \tag{2.11}$$

But from equation (2.10)

$$\sum_{t \in T} \eta(r_t(P_i), P_i) \leq \sum_{t=1}^{k} \eta(r_t(P_i), P_i) \tag{2.12}$$

Combining equations (2.11) and (2.12), 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 2.4.1 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 2.4.2 (Scoring Correspondences)

Let $s \equiv (s_1, s_2, \cdots, s_m)$ be a vector in $I\!R^m$ 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 2.4.1 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 $\bar{\mu}$.

The next example shows that Theorem 2.4.1 only provides sufficient conditions for a SCF to be OBIC with respect to the uniform prior.

EXAMPLE 2.4.3

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

$$
\begin{array}{ccccccc}
 & abc & acb & bac & bca & cab & cba \\
abc & a & a & a & b & c & a \\
acb & a & a & b & a & a & c \\
bac & b & a & b & b & b & c \\
bca & a & b & b & b & c & b \\
cab & a & c & c & b & c & c \\
cba & c & a & b & c & c & c
\end{array}
\tag{2.13}
$$

We claim that f is OBIC with respect to $\bar{\mu}$. To see this, observe that (as in Example 2.4.1), for both individuals and for all six preference orderings, the frequency of occurrence of the first, second and third alternatives is four. one and one respectively. It is also easy to verify that f is neutral unlike the SCF in Example 2.4.1. However, it does not satisfy elementary monotonicity. For instance. observe that $f(abc, cba) = a$ but $f(abc, cab) = c$.

Our final observation in this section is that 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 (2.5).

2.5 Conclusion

We have examined the implications of weakening the incentive requirement in voting theory from dominant strategies to ordinal Bayesian incentive compatibility.

Truth-telling is no longer assumed to be optimal for every conceivable strategy-tuple of the other players. It is only required to maximize expected utility given an agent's prior beliefs about the types of other players and the assumption that these players are following truth-telling strategies. The set of ordinal Bayesian incentive compatible social choice functions clearly depends on the beliefs of each agent. However, we show that for generic beliefs appropriately defined, the only social choice functions which are incentive compatible in this sense are dictatorial. We are thus unable to escape the negative conclusion of the Gibbard-Satterthwaite Theorem. However, a dramatically different picture can emerge for non-generic cases. A case of particular interest is the case of uniform priors or "complete ignorance". We provide a weak sufficient condition for incentive compatibility and show that a large class of well-behaved social choice functions satisfy these conditions.

2.6 Appendix

PROOF OF THEOREM 2.3.1: The proof proceeds in several steps. In Step 1, we define the sets C_i and show that they are open and dense subsets of Δ^{CI} 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 $\mu_{-i} \in C_i$ for all i, 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

Pick an arbitrary individual i. We define the set C_i below.

For any $Q \subseteq I\!P^{N-1}$, let $\mu_{-i}(Q) = \sum_{P_{-i} \in Q} \mu_{-i}(P_{-i})$. The set $\mathcal{C}_i$ is defined as the set of conditional beliefs μ_{-i} satisfying the following property: For all $Q, T \subset I\!P^{N-1}$

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

We first show that $\mathcal{C}_i$ is open in Δ^{CI}. Consider any μ such that for all $i \in N$, $\mu_{-i} \in \mathcal{C}_i$. Let

$$\phi(\mu) = \min_{S,T \subset I\!P^{N-1}, S \neq T} |\mu_{-i}(S) - \mu_{-i}(T)|$$

Observe that $\phi(\mu) > 0$. Since ϕ is a continuous function of μ, there exists $\epsilon > 0$ such that for all product measures $\hat{\mu} \in \Delta^I$ with $d(\hat{\mu}, \mu) < \epsilon$, [2] we have $\phi(\hat{\mu}) > 0$. But this implies that $\hat{\mu}_{-i} \in \mathcal{C}_i$. Therefore $\mathcal{C}_i$ is open in Δ^{N-1}.

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

$$\Delta^{CI} - \mathcal{C}_i = \bigcup_{Q,T \subset I\!P^{N-1}} \{\mu \in \Delta^{CI} | \mu_{-i}(Q) = \mu_{-i}(T)\}$$

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

Pick a product measure μ such that for some i, $\mu_{-i} \in \Delta^{CI} - \mathcal{C}_i$ and consider an open neighbourhood of radius $\epsilon > 0$ with centre μ_{-i}. Since this neighbourhood has strictly positive measure and since $\Delta^{CI} - \mathcal{C}_i$ has measure zero, it must be the case that the neighbourhood has a non-empty intersection with the set $\mathcal{C}_i$. This establishes that $\mathcal{C}_i$ is dense in Δ^{CI}.

This completes Step 1.

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

STEP 2

Let f be a SCF which is OBIC with respect to the belief μ where $\mu_{-i} \in C_i$ for all i. Our goal in this step of the proof is to show that f must satisfy Property M. 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. We give the formal definition below.

DEFINITION 2.6.1 *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*

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

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 and μ is independent, we have, by using equation (2.2)

$$\begin{aligned} &\mu_{-i}(\{P_{-i} | f(P_i, P_{-i}) \in B(r_k(P_i), P_i)\}) \\ &\quad \geq \mu_{-i}(\{P_{-i} | f(P_i', P_{-i}) \in B(r_k(P_i), P_i)\}) \end{aligned} \tag{2.14}$$

Suppose i's "true" preference is P_i'. Applying equation (2.2), we have

$$\begin{aligned} &\mu_{-i}(\{P_{-i} | f(P_i', P_{-i}) \in B(r_k(P_i'), P_i')\}) \\ &\quad \geq \mu_{-i}(\{P_{-i} | f(P_i, P_{-i}) \in B(r_k(P_i'), P_i')\}) \end{aligned} \tag{2.15}$$

Since $B(r_k(P_i), P_i) = B(r_k(P_i'), P_i')$, equations (2.14) and (2.15) imply,

$$\begin{aligned} &\mu_{-i}(\{P_{-i} | f(P_i, P_{-i}) \in B(r_k(P_i), P_i)\}) \\ &\quad = \mu_{-i}(\{P_{-i} | f(P_i', P_{-i}) \in B(r_k(P_i'), P_i')\}) \end{aligned} \tag{2.16}$$

Since $\mu_{-i} \in C_i$, it follows from and equation (2.16) that

$$\{P_{-i} | f(P_i, P_{-i}) \in B(r_k(P_i), P_i)\} = \{P_{-i} | f(P_i', P_{-i}) \in B(r_k(P_i'), P_i')\} \tag{2.17}$$

Now suppose for some P_i, we have $f(P_i, P_{-i}) \in B(r_k(P_i), P_i)$. Then equation (2.17) 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 two-person SCF which satisfies Property M must be dictatorial. Let $N = \{1, 2\}$ and let f satisfy Property M.

CLAIM A: 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 B: 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 A 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 A 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 A 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 alternative provided that 2's first-ranked alternative is x. Since x is arbitrary, the Claim is proved.

It follows immediately from Claim B 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 $f : \mathbb{P}^K \to A$ satisfies Property M, then f is dictatorial.

Our goal is to prove:

If $f : \mathbb{P}^N \to A$ satisfies Property M then f is dictatorial.

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

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

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 "coalesced" 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 $P_{-i} \in \mathbb{P}^{N-2}$, 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 coalesced 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 dictator 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 a 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 $N-1$ 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. ∎

PROOF OF THEOREM 2.3.2: For all $i \in N$, define $\mathcal{M}_i \subset \tilde{\Delta}$ as the set of measures μ_i for which the following holds: for all $Q, T \subseteq I\!P^N$

$$[\mu_i(Q) = \mu_i(T)] \Rightarrow [Q = T] \tag{2.18}$$

Replicating the arguments in the proof of Theorem 2.3.1, it follows that $\mathcal{M}_i$ is open and dense in $\tilde{\Delta}$ and that $\tilde{\Delta} - \mathcal{M}_i$ has Lebesgue measure zero.

Let $\mu_i \in \mathcal{M}_i \cap \mathcal{F}_i$. We now show that the induced conditional belief μ_{-i} (defined unambiguously because $\mu_i \in \mathcal{F}_i$) satisfies a similar property. In particular,we show that, for all $P, Q \subseteq I\!P^{N-1}$,

$$[\mu_{-i}(Q) = \mu_{-i}(T)] \Rightarrow [Q = T] \tag{2.19}$$

In order to prove this suppose that equation (2.18) is true but equation (2.19) is not. Suppose therefore that there exists i, P_i and sets $Q, T \subseteq I\!P^{N-1}$ such that

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

Let μ_i^m be the marginal distribution corresponding to μ_i. Therefore,

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

$$\Rightarrow \mu_i(Q \times \{P_i\}) = \mu_i(T \times \{P_i\})$$

Since Q and T are distinct, so are $Q \times \{P_i\}$ and $T \times \{P_i\}$. Since the latter sets are subsets of $I\!P^N$, we obtain a contradiction to equation (2.18) which we assumed to be true.

Suppose f is OBIC with respect to $(\mu_1, \cdots, \mu_N)$ with $\mu_i \in \mathcal{M}_i \cap \mathcal{F}_i$ for all $i = 1, \cdots, N$. Since $\mu_i \in \mathcal{F}_i$, inequalities (2.14) and (2.15) hold, so that (2.16) holds as well. Applying (2.19), we obtain (2.17) again. Therefore Property M holds and Steps 3 and 4 can be replicated. ■

In order to prove Theorem 2.3.3, define $\mathcal{M}$ to be the set of (common) prior beliefs μ such that, for all $P, Q \subseteq I\!P^{N-1}$,

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

It is clear that the arguments in the proof of Theorem 2.3.2 go through in this case with $\mathcal{M}$ substituted for $\mathcal{M}_i$.

Chapter 3

Ordinal Bayesian Incentive Compatibility and Stable Matchings

3.1 Introduction

In this chapter we explore issues in incentives related to matching problems and the design of matching procedures. *Matching problems* refer to problems which involve matching members of one set of agents to members of a second, disjoint set of agents all of whom have preferences over the possible resulting matches. We focus attention on two-sided, one-to-one matching where each agent is matched to at most one mate. A fundamental notion in this context is a *stable matching* which can be defined as a matching such that there does not exist a pair of agents who would prefer to be matched to each other than to their current partners. Such a matching is in the core of the corresponding cooperative game which would result if individual agents were able to freely negotiate their own matches. Gale and Shapley (1962) show that the set of stable matchings is non-empty.

In the strategic version of the model the preferences of the agents are private information. Therefore any stable matching is computed on the basis of the reported preferences. The agents know that by reporting different preferences they can alter the stable matching that is selected and hence change their mate. A natural question which arises is whether *matching procedures* can be designed which give the agents inventive to truthfully reveal their preferences, and which produce stable matchings. The truth-telling concept mostly used in the literature is *strategy-proofness.* Under strategy-proofness it is a dominant strategy for all the agents to truthfully reveal their preferences. The question is does there exist a stable marriage procedure that is strategy-proof. Roth (1982) demonstrates that there does not exist any matching procedure which is strategy-proof and which also generates stable matching at every profile of preferences. This result is similar in spirit to a number of impossibility results present in the social choice literature, in the context of designing non-dictatorial social choice procedures which operate in fairly unrestricted domains (Gibbard (1973), Satterthwaite(1975)).

In this essay we weaken the truth-telling requirement from strategy-proofness to *ordinal Bayesian incentive compatibility (OBIC).* This notion was introduced in d'Aspremont and Peleg (1988) in the context of a different problem, that of representation of committees and analysed in standard voting environments in Chapter 2. Truth-telling is required to maximize the expected utility of each individual where expected utility is computed with reference to the individual's *prior beliefs* about the (possible) preferences of other individuals and based on the assumption that other individuals follow the truth-telling strategy. However, this truth-telling notion has one important difference with the standard notion of Bayesian incentive compatibility used widely in incentive theory (for example auction theory). Under OBIC truth-telling is required to maximize expected utility for *every* representation of an individual's true preference ordering. Roth (1989) applies the notion of Bayesian incentive compatibility to the stable matching problem. He generalizes the Roth (1982) result to the case where truth-telling is a Bayes-Nash equilibrium of the revelation game. However, he assumes particular cardinalization of utilities and makes specific assumptions about priors. Since stable matchings only considers preferences and since individual preferences are ordinal, a more appropriate equilibrium notion would be ordinal Bayesian incentive compatibility.

As we have noted in Chapter 2 ordinal Bayesian incentive-compatibility is a significant weakening of the strategy-proofness requirement. We might therefore expect a possibility result to emerge if ordinal Bayesian incentive compatibility was used as the truth-telling requirement. However our first result states that there does not exist any prior such that there exists a stable matching procedure that is ordinally Bayesian incentive compatible with respect to it.

Our next step is to look for possibility results by putting restrictions on the set of allowable preferences of the agents. Alcalde and Barberá (1994) look at possibility results by restricting the set of allowable preferences but maintaining strategy-proofness as the notion of truth-telling. We restrict attention to the class of preferences where each agent prefers to be matched than to remain single and show that when each individual's belief about the preferences of others is uniformly and independently distributed then there *exist* stable matching procedures that are ordinally Bayesian incentive compatible. In a recent paper, Roth and Rothblum (1999) consider stable matching in an incomplete information environment where agents have what they call "symmetric beliefs". If beliefs are uniform then they are symmetric. Roth and Rothblum discuss stochastic dominance of one strategy over others in such an environment. They show that if the stable matching procedure is the *man proposing deferred acceptance algorithm* then for any woman with symmetric beliefs any strategy that changes her true preference ordering of men is stochastically dominated by a strategy that states the same number of acceptable men in their correct order. Ehlers (2001) gives an alternative condition to the symmetry condition on beliefs that leads to the same result. However, none of the papers analyse equilibrium behaviour of agents. Our possibility result with uniform priors follows immediately from and can be seen to be an equilibrium interpretation of the Roth and Rothblum (1999) and the Ehlers (2001) results.

Our main result in this chapter is to show that this possibility result is non-generic. Following the analysis of the previous chapter we assume common independently distributed prior for all individuals and show that for each individual i there exists a set of conditional beliefs $\mathcal{C}_i$ which is open and dense in the set of all conditional beliefs and whose complement set is of Lebesgue measure zero, such that *no* stable matching procedure exists that is ordinally Bayesian incentive compatible with respect to a prior belief μ such that the conditionals generated by μ lie in $\mathcal{C}_i$.

The essay is organised as follows. In section 3.2 we set out the basic notation and definitions. Section 3.3 deals with the case of unrestricted preferences. In section 3.4 we consider restricted preferences. In subsection 3.4.1 we deal with uniform priors while subsection 3.4.2 considers generic priors. Section 3.5 concludes. Appendix A contains the deferred acceptance algorithm while Appendix B briefly discusses symmetric beliefs.

3.2 Preliminaries

We assume that there are two disjoint sets of individuals which we refer to as the set of men and women. These sets are denoted by M and W respectively. Elements in M are denoted by m, m' etc and elements is W are denoted by w, w' etc. Let $I \equiv M \cup W$ denote the entire set of agents. Each man $m \in M$ has a preference ordering P_m over the set $W \cup \{m\}$. Let $\mathcal{P}_m$ be the set of all possible preference orderings for man m. Each woman w has a preference ordering P_w over the set $M \cup \{w\}$. Let $\mathcal{P}_w$ denote the set of all possible preference orderings for woman w. We denote by $P = ((P_m)_{m \in M}, (P_w)_{w \in W})$, a preference profile for all the agents. Let

$\mathcal{P} = \times_{i \in I} \mathcal{P}_i$ denote the set of all such preference profiles. We assume that these orderings are strict. We denote by P_{-i} the collection of preferences for all agents other than i. The set of all such P_{-i}'s is denoted by $\mathcal{P}_{-i} = \times_{j \neq i} \mathcal{P}_j$.

We will usually describe an agent's preferences by writing only the ordered set of people that the agent *weakly* prefers to remaining single. Thus the preference P_m described below,

$$P_m := w_1 P_m w_2 P_m m P_m, \cdots, P_m w_k$$

will be abbreviated to,

$$P_m := w_1 P_m w_2 P_m m$$

For reasons that will be obvious shortly, it will suffice only to consider these abbreviated preferences.

DEFINITION 3.2.1 *A matching is a function $\nu : I \to I$ satisfying the following properties:*

- $\nu(m) \in W \cup \{m\}$
- $\nu(w) \in M \cup \{w\}$
- $\nu(\nu(i)) = i \qquad \forall i \in I$

We now define a *stable matching.* Let $A(P_i) = \{j \in I | j P_i i\}$ denote the set of acceptable mates for agent i. Obviously for a man m with preference ordering P_m, $A(P_m) \subseteq W$ and similarly for a woman w with preference P_w, $A(P_w) \subseteq M$.

DEFINITION 3.2.2 *A matching ν is stable if the following two conditions are satisfied*

- *for all $i \in I$, $\nu(i) \in A(P_i) \cup \{i\}$*

- *there does not exist* $(m,w) \in M \times W$ *such that* $wP_m\nu(m)$ *and* $mP_w\nu(w)$

Let $\mathcal{S}(P)$ denote the set of stable matches under P. Gale and Shapley (1962) shows that $\mathcal{S}(P)$ is always non-empty for all $P \in \mathcal{P}$.

Let $\mathcal{M}$ denote the set of all possible matchings. A *matching procedure* is a mapping that associates a matching with every preference profile P.

DEFINITION 3.2.3 *A matching procedure is a function* $f : \mathcal{P} \to \mathcal{M}$

If f is a matching procedure and P is a profile, then $f_i(P)$ denotes the match for i selected by f under P.

A *stable* matching procedure f selects an element from the set $\mathcal{S}(P)$ for every $P \in \mathcal{P}$. The rest of the essay is concerned only with stable matching procedures.

We now look at strategic issues in the model. In the strategic version of this problem each agent's preference over his/her possible mates is private information. A question of fundamental interest is the following: does there exist a stable, strategy-proof matching procedure? The answer is negative.

DEFINITION 3.2.4 *A matching procedure* f *is strategy-proof if there does not exist* $i \in I$, $P_i, P_i' \in \mathcal{P}_i$, *and* $P_{-i} \in \times_{j \neq i}\mathcal{P}_j$ *such that*

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

THEOREM 3.2.1 *Roth (1982)*

A stable and strategy-proof matching procedure does not exist.

In this essay, we explore the consequences of weakening the incentive requirement for stable matching procedures 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 3.2.5 *A belief for an individual i is a probability distribution on the set $\mathcal{P}$, i.e. it is a map $\mu_i : \mathcal{P} \to [0,1]$ such that $\sum_{P \in \mathcal{P}} \mu_i(P) = 1$.*

We assume that all individuals have a common prior belief μ. 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.

Consider a man m. The utility function $u_m : W \cup \{m\} \to \Re$ represents $P_m \in \mathcal{P}_m$, if and only if for all $i, j \in W \cup \{m\}$,

$$iP_m j \Leftrightarrow u_m(i) > u(j)$$

The utility function u_w for a woman w is similarly defined.

For any agent $i \in I$ we will denote the set of utility functions representing P_i by $\mathcal{U}_i(P_i)$.

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

DEFINITION 3.2.6 *A matching procedure f is ordinally Bayesian Incentive Compatible (OBIC) with respect to the belief μ if for all $i \in I$, for all $P_i, P_i' \in \mathcal{P}_i$, for all $u_i \in \mathcal{U}(P_i)$, we have*

$$\sum_{P_{-i} \in \mathcal{P}_{-i}} u_i\left(f_i(P_i, P_{-i})\right) \mu(P_{-i}|P_i) \geq \sum_{P_{-i} \in \mathcal{P}_{-i}} u_i\left(f_i(P_i', P_{-i})\right) \mu(P_{-i}|P_i) \qquad (3.1)$$

As in the earlier chapter it is possible to give an alternative definition of OBIC in terms of stochastic dominance. For any agent $i \in I$, let I_i be the set of possible mates for i. Thus if $i \equiv m \in M$, then $I_i = W \cup \{m\}$ and if $i \equiv w$ then $I_i = M \cup \{w\}$. For all $P_i \in \mathcal{P}_i$ and $k = 1, \cdots, |I_i|$, let $r_k(P_i)$ denote the k th ranked mate in P_i, i.e., $r_k(P_i) = j$ implies that $|\{l \neq j | lP_i j\}| = k - 1$. For all $i \in I$, for any $P_i \in \mathcal{P}_i$ and for any $j \in I_i$, let $B(j, P_i) = \{l \in I_i | lP_i j\} \cup \{j\}$. Thus $B(j, P_i)$ is the set of mates that are weakly preferred to j under P_i.

The stable matching procedure f is OBIC with respect to the belief μ if for all $i \in I$, for all integers $k = 1, \cdots, |I_i|$ and for all P_i and P_i',

$$\mu(\{P_{-i}|f(P_i, P_{-i}) \in B(r_k(P_i), P_i)\}|P_i) \geq \mu(\{P_{-i}|f(P_i', P_{-i}) \in B(r_k(P_i), P_i)\}|P_i) \tag{3.2}$$

3.3 The Case of Unrestricted Preferences

The main result of this section is to show that there does not exist any stable marriage procedure that is OBIC with respect to any prior belief μ[1]. In an earlier paper, Roth (1989) extends the analysis of Roth (1982) by weakening the truth-telling requirement to Bayesian incentive compatibility. However, he assumes particular cardinalization of utilities. The paper shows that there exists specific utility values and probability distributions for which no stable matching procedure is Bayesian incentive compatible. The paper therefore, does not rule out the possibility that there may exist utility profiles and probability distributions for which there exist Bayesian incentive compatible stable procedures. However, since stable matchings are based only on ordinal preferences, it is possible to argue that OBIC is a more appropriate equilibrium notion. We have the following strong negative result.

THEOREM **3.3.1** *Let $|M|$, $|W| \geq 2$ and assume that there are no restrictions on the preferences of individuals. Then for any prior belief μ, there does not exist a stable matching procedure f such that f is OBIC with respect to μ.*

Let f be a stable matching procedure. We first establish a lemma which says the following: consider an agent $i \in I$ and two preference orderings P_i and P_i' such that $r_1(P_i) = r_1(P_i') = j$. However under preference ordering P_i' agent i prefers to

[1]The result holds even if we do away with the assumption of common priors

remain single than to be matched to any agent other than j. Lemma 3.1 shows that if for some combination of others preferences P_{-i}, f picks j to be i's mate when i reports P_i, then f should pick j as i's mate when i reports P_i'. Formally we show the following:

Lemma 3.3.1 *Consider an agent $i \in I$ and two preferences P_i and P_i' such that $r_1(P_i) = r_1(P_i') = j$ and $r_2(P_i') = i$. Then for any $P_{-i} \in \mathcal{P}_{-i}$,*

$$[f_i(P_i, P_{-i}) = j] \Rightarrow [f_i(P_i', P_{-i}) = j]$$

PROOF: It follows from the definition of stable matching that $f_i(P_i', P_{-i}) \in \{j, i\}$. Suppose that $f_i(P_i', P_{-i}) = i$. Observe that for agent j, $i \in A(P_j) \cup \{j\}$.. Also since the preferences for all the agents other than i have not changed, we claim that any k such that kP_ji will not be matched to j under the preference profile (P_i', P_{-i}). Suppose that the claim is not true and suppose that there exists a k with kP_ji such that, $k = f_j(P_i', P_{-i})$. Since f is a stable matching procedure it follows that $f_k(P)P_kj$, otherwise (k, j) would have blocked the matching selected by f under the profile P. Let $l = f_k(P)$. Replicating the arguments above one can show that $f_l(P_i', P_{-i})P_lf_l(P) = k$. Otherwise k and l would block the matching $f(P_i', P_{-i})$. Let $f_l(P_i', P_{-i}) = k' \neq k$. Observe that $k' \neq i$ for in the matching $f(P_i', P_{-i})$, i is remaining single. Again by analogous arguments it follows that $f_{k'}(P)P_{k'}f_{k'}(P_i', P_{-i}) = l$. Thus there exists a sequence of pairs $\{(k_n, l_n) | n = 1, 2, 3, \cdots\}$ where any two pairs are distinct (i.e., for any n_1 and n_2, $k_{n_1} \neq k_{n_2}$ and $l_{n_1} \neq l_{n_2}$) such that $k_1 = i$, $l_1 = j$ and

$$l_n = f_{k_n}(P)P_{k_n}f_{k_n}(P_i', P_{-i}) = l_{n+1} \text{ and}$$
$$k_{n+1} = f_{l_n}(P_i', P_{-i})P_{l_n}f_{l_n}(P) = k_n$$

Since I is finite there exists a n^* such that,

$$l_{n^*} = f_{k_{n^*}}(P)P_{k_{n^*}}f_{k_{n^*}}(P_i', P_{-i}) \text{ and,}$$

there does not exist a $\hat{k} \in I \setminus \{k_n\}_{n=1}^{n^*}$ such that $\hat{k}P_{l_{n^*}}f_{l_{n^*}}(P)$. Then (k_{n^*}, l_{n^*}) will block the matching $f(P_i', P_{-i})$. Therefore it follows that any k such that $kP_j i$ will not be matched to j under (P_i', P_{-i}). So if $f_i(P_i', P_{-i}) = i$ it implies that for agents i and j,

$$jP_i' f_i(P_i', P_{-i}) = i \text{ and}$$
$$iP_j f_j(P_i', P_{-i})$$

Then f is not a stable matching procedure. We thus have a contradiction. Therefore $f_i(P_i', P_{-i}) = j$. ■

PROOF OF THEOREM 3.3.1 Let f be OBIC with respect to μ. Pick $i \in I$ and preferences P_i and P_i'. From (3.2) we get,

$$\mu(\{P_{-i}|f_i(P_i, P_{-i}) = r_1(P_i)\}|P_i) \geq \mu(\{P_{-i}|f_i(P_i', P_{-i}) = r_1(P_i)\}|P_i) \tag{3.3}$$

Consider a preference profile P such that, $P_{m_1} := w_1 P_{m_1} w_2 P_{m_1} m_1$; $P_{m_2} := w_2 P_{m_2} w_1 P_{m_2} m_2$; $P_{w_1} := m_2 P_{w_1} m_1 P_{w_1} w_1$; $P_{w_2} := m_1 P_{w_2} m_2 P_{w_2} w_2$; also let $P_j := j$ for all $j \in I \setminus \{m_1, m_2, w_1, w_2\}$. It is easy to check that $\mathcal{S}(P)$ consists of two matchings ν_1 and ν_2 where $\nu_1(m_1) = w_1$, $\nu_1(m_2) = w_2$, $\nu_1(j) = j$ for all $j \in I \setminus \{m_1, m_2, w_1, w_2\}$, $\nu_2(m_1) = w_2$, $\nu_2(m_2) = w_1$ and $\nu_2(j) = j$ for all $j \in I \setminus \{m_1, m_2, w_1, w_2\}$. Suppose $f(P) = \nu_1$. Now consider $P_{w_2}' := m_1 P_{w_2}' w_2$. Then we claim that the only stable matching in the profile (P_{w_2}', P_{-w_2}) is ν_2. Suppose $f(P_{w_2}', P_{-w_2}) = \nu$. Note that $\nu(w_2)$ is either m_1 or w_2. Suppose $\nu(w_2) = w_2$. Then either $\nu(m_1) = m_1$ or $\nu(m_2) = m_2$. If $\nu(m_1) = m_1$, then (m_1, w_2) blocks ν. Therefore $\nu(m_1) = w_1$ and $\nu(m_2) = m_2$. Then (m_2, w_1) blocks ν. Therefore $\nu(w_2) = m_1$ and $\nu(w_1) = m_2$. But then $\nu = \nu_2$. Since $f_{w_2}(P) = m_2$, $f_{w_2}(P_{w_2}', P_{-w_2}) = m_1$ and $r_1(P_{w_2}) = m_1$, it must be the case in order for (3.3) to hold that there exists $\tilde{P}_{-w_2}$

such that $f_{w_2}(P_{w_2}, \tilde{P}_{-w_2}) = m_1$ and $f_{w_2}(P'_{w_2}, \tilde{P}_{-w_2}) \neq m_1$. But from Lemma 3.3.1 this will never be the case. Thus $f(P) \neq \nu_1$. Therefore $f(P) = \nu_2$. Now consider $P'_{m_1} := w_1 P'_{m_1} m_1$. The only stable matching under the profile (P'_{m_1}, P_{-m_1}) is ν_1. By replicating the earlier argument it follows that if $f(P'_{m_1}, P_{-m_1}) = \nu_1$ then $f(P)$ can never be ν_2. But this is a contradiction. This completes the proof of the theorem. ∎

The result in this section assumes unrestricted preferences, i.e., each man m is allowed to have any ordering over the set $W \cup \{m\}$ and similarly each woman w is allowed to have any ordering over the set $M \cup \{w\}$. Alcalde and Barberà (1994) put strong restrictions on preferences to obtain strategy-proof stable matchings. In the next section we put weaker restrictions on preferences to see whether possibility results with OBIC can be obtained.

3.4 Restricted Preferences

In this section we examine the stable matching problem for a special class of preferences. We restrict our attention to the class of preferences where remaining single is the worst alternative for every agent. That is, each agent prefers to be matched to some other agent than to remain single.

Formally, the domain $\mathcal{D}$ consists of *all* preferences (P_m, P_w) satisfying the following conditions:

- for all $w_i \in W$, $w_i P_m m$
- for all $m_i \in M$, $m_i P_w w$

In this environment a stable matching procedure is a function $f : \mathcal{D} \to \mathcal{M}$ with the restriction that $f(P) \in \mathcal{S}(P)$ for all $P \in \mathcal{D}$. We denote by $\mathcal{D}_{-i}$ the set of all P_{-i}'s, where P_{-i} is the collection of preferences of all agents other than i.

The man proposing and the woman proposing *deferred acceptance* algorithms are ways to obtain a stable matching given the preference reports of men and women. Both algorithms are discussed in Appendix A.

Let $f^{DA(m)}$ denote the stable matching procedure that uses the man proposing deferred acceptance algorithm and let $f^{DA(w)}$ denote the woman proposing deferred acceptance algorithm. Roth (1982) demonstrates that with the man proposing deferred acceptance algorithm it is a dominant strategy for men to truthfully reveal their preferences i.e., it is strategy-proof for men. Since men and women are symmetric in this model, the woman proposing deferred acceptance algorithm is strategy-proof for women.

THEOREM **3.4.1** *Roth (1982)*

The stable matching procedure $f^{DA(m)}$, is strategy-proof for men. Similarly, $f^{DA(w)}$ is strategy-proof for women.

3.4.1 Uniformly and Independently Distributed Priors

In this section, we assume that the beliefs are independently and uniformly distributed.

DEFINITION 3.4.1 *Individual i's beliefs are independent if for all $k = 1, \cdots, |I|$ there exist probability distributions $\mu_k : \mathcal{P}_k \rightarrow [0,1]$ such that, for all P_{-i} and P_i,*

$$\mu(P_{-i}|P_i) = \times_{k \neq i} \mu_k(P_k)$$

An individual's belief is independent if his conditional belief about the types of the other individuals is a product measure of the marginals over the types of the other individuals. We also assume that the beliefs are uniform.

DEFINITION 3.4.2 *For all profiles $P, P' \in \mathcal{P}$, we have*

$$\mu(P) = \mu(P')$$

We denote these independent, uniform priors by $\bar{\mu}$. Restating Definition 3.6 in the present context, we have

PROPOSITION 3.4.1 *The matching procedure f is OBIC with respect to the belief $\bar{\mu}$ if, for all i, for all integers $k = 1, \cdots, |I_i|$, for all P_i and P_i', we have*

$$|\{P_{-i}|f_i(P_i, P_{-i}) \in B(r_k(P_i), P_i)\}| \geq |\{P_{-i}|f_i(P_i', P_{-i}) \in B(r_k(P_i), P_i)\}| \quad (3.4)$$

We omit the trivial proof of this Proposition.

Roth and Rothblum (1999) define a particular type of belief for agents which they call "symmetric" beliefs. Symmetric beliefs are discussed in Appendix B. We note that independent, uniform beliefs are symmetric. They show that if the stable matching procedure is $f^{DA(m)}$ then for a woman with symmetric beliefs, a strategy that changes her true preference ordering of men is stochastically dominated by a strategy that states the same number of acceptable men in their correct order, i.e., in the order of the true preference ordering. The same is true for men when the matching procedure is $f^{DA(w)}$. The following theorem can be treated as an equilibrium interpretation of the Roth and Rothblum results.

THEOREM 3.4.2 *The stable marriage procedures $f^{DA(m)} : \mathcal{D} \to \mathcal{M}$ and $f^{DA(w)} : \mathcal{D} \to \mathcal{M}$ are OBIC with respect to the uniform prior.*

PROOF: We give the proof for $f^{DA(m)}$. The proof for $f^{DA(w)}$ is analogous. From Theorem 3.4.1 we know that $f^{DA(m)}$ is strategy-proof for men. So we only need to check whether $f^{DA(m)}$ is OBIC with respect to the uniform prior for women. Observe that if any $w \in W$ has uniformly and independently distributed prior belief then her conditional belief is $\{M\}$-symmetric. So Proposition 3.7.1 applies and hence any strategy that changes her true preference ordering of men is stochastically dominated by a strategy that states the same number of acceptable men in their correct order. However, when preference profiles are in $\mathcal{D}$, for any $w \in W$ with preference order P_w, the only strategy that states w's set of acceptable men in their correct order is P_w. Since OBIC is equivalent to the stochastic domination of truth-telling this proves the theorem. ∎

3.4.2 Generic Priors

The main result in this section is to show that the possibility result of the previous section vanish if the beliefs are perturbed appropriately. We continue to assume first that the beliefs are independent.

For each agent i, we let $\Delta(i)$ denote the set of all beliefs over the possible types of i. If i is a man, $\Delta(i)$ is a unit simplex of dimension $(|W|+1)-1$. If i is a woman, $\Delta(i)$ is a unit simplex of dimension $(|M|+1)-1$. The set of all independent priors $\Delta^I = \times_{i \in I}\Delta(i)$. For an agent i and belief $\mu \in \Delta^I$, we shall let μ_{-i} i's conditional belief over the types of agents other than i. For instance $\mu_{-i}(P_{-i})$ will denote the probability under μ that the preferences of agents other than i, is P_{-i}. The set of all such conditional beliefs will be denoted by Δ^{CI}. Clearly, $\Delta^{CI} = \times_{k \neq i}\Delta(k)$.

We now state the main result of this section.

THEOREM **3.4.3** *Let $|M| = |W| \geq 3$ and assume that all individuals have independent beliefs. Then for all $i \in I$, there exists a subset C_i of $\Delta^{CI}(i)$ such that*

- *C_i is open and dense in $\Delta^{CI}(i)$*
- *$\Delta^{CI}(i) - C_i$ has Lebesgue measure zero*
- *there does not exist a stable marriage procedure $f : \mathcal{D} \to \mathcal{M}$ that is OBIC w.r.t the belief μ where $\mu_{-i} \in C_i$ for all $i \in I$.*

PROOF: The proof proceeds in three steps. In Step 1 we define the sets C_i and show that they are open and dense subsets of $\Delta^{CI}(i)$ and the Lebesgue measure of their complement sets are zero. In Step 2 we show that if a matching procedure f is OBIC with respect to μ with $\mu_{-i} \in C_i$ for all i, then f must satisfy a certain property which we call *Top Monotonicity(TM).* In Step 3 we complete the proof by showing that stable matching procedure violates TM.

STEP1:

Pick an individual i. We define the set C_i below.

For any $Q \subseteq \mathcal{D}_{-i}$, let $\mu_{-i}(Q) = \sum_{P_{-i} \in Q} \mu_{-i}(P_{-i})$. The set C_i is defined as the set of conditional beliefs μ_{-i} satisfying the following property:For all $Q, T \subseteq \mathcal{D}_{-i}$

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

By arguments similar to the one in the previous chapter we can show that C_i is open and dense in $\Delta^{CI}(i)$ and that $\Delta^{CI}(i) - C_i$ is of Lebesgue measure zero. We omit the details here as the proof is provided in the previous chapter.

STEP 2:

Let f be a matching procedure that is OBIC with respect to the belief μ where $\mu_{-i} \in C_i$ for all i. In this step we show that f must satisfy property TM which we describe below.

DEFINITION 3.4.3 *The marriage procedure f satisfies TM, if for all individuals i, for all P_{-i} and for all P_i, P_i' such that $r_1(P_i) = r_1(P_i')$, we have*

$$f_i(P_i, P_{-i}) = r_1(P_i) \Rightarrow f_i(P_i', P_{-i}) = r_1(P_i')$$

Let i be an individual and let P_i and P_i' be such that $r_1(P_i) = r_1(P_i')$. Suppose i's "true" preference is P_i. Since f is OBIC with respect to μ, we have, by using equation (3.2)

$$\mu(\{P_{-i}|f_i(P_i,P_{-i})=r_1(P_i)\}) \geq \mu(\{P_{-i}|f_i(P_i',P_{-i})=r_1(P_i')\}) \tag{3.5}$$

Suppose i's true preference is P_i'. Applying equation (3.2) we have

$$\mu(\{P_{-i}|f_i(P_i',P_{-i})=r_1(P_i')\}) \geq \mu(\{P_{-i}|f_i(P_i,P_{-i})=r_1(P_i)\}) \tag{3.6}$$

Combining (3.5) and (3.6) and using the fact that $r_1(P_i)=r_1(P_i')$ we get,

$$\mu(\{P_{-i}|f_i(P_i,P_{-i})=r_1(P_i)\}) = \mu(\{P_{-i}|f_i(P_i',P_{-i})=r_1(P_i')\}) \tag{3.7}$$

Since $\mu(P_{-i}) \in \mathcal{C}_i$ it follows from (3.7) that,

$$\{P_{-i}|f_i(P_i,P_{-i})=r_1(P_i)\} = \{P_{-i}|f_i(P_i',P_{-i})=r_1(P_i')\} \tag{3.8}$$

Thus, if for some P_i $f_i(P_i,P_{-i})=r_1(P_i)$, then (3.8) implies that $f_i(P_i',P_{-i}) = r_1(P_i')$. Therefore f satisfies TM.

STEP 3: In this step we complete the proof of the theorem by showing that a stable matching procedure does not satisfy TM.

Let $|M|=|W|\geq 3$ and let $f:\mathcal{D}\to\mathcal{M}$ be a stable matching procedure, i.e., for all $P\in\mathcal{D}$, $f(P)\in\mathcal{S}(P)$. Consider a preference profile P defined as follows:

$$\begin{aligned}
P_{m_1} &:= w_2 P_{m_1} w_1 P_{m_1} w_3 P_{m_1}, \cdots, P_{m_1} m_1\\
P_{m_2} &:= w_1 P_{m_2} w_2 P_{m_2} w_3 P_{m_2}, \cdots, P_{m_2} m_2\\
P_{m_3} &:= w_1 P_{m_3} w_2 P_{m_3} w_3 P_{m_3}, \cdots, P_{m_3} m_3\\
P_{w_1} &:= m_1 P_{w_1} m_3 P_{w_1} m_2 P_{w_1}, \cdots, P_{w_1} w_1\\
P_{w_2} &:= m_3 P_{w_2} m_1 P_{w_2} m_2 P_{w_2}, \cdots, P_{w_2} w_2\\
P_{w_3} &:= m_1 P_{w_3} m_2 P_{w_3} m_3 P_{w_3}, \cdots, P_{w_3} w_3
\end{aligned}$$

For all $k\neq 1,2,3$, $P_{m_k} := w_k P_{m_k}, \cdots, P_{m_k} m_k$ and $P_{w_k} := m_k P_{w_k}, \cdots, P_{w_k} w_k$. We claim that $\mathcal{S}(P)=\{\nu_1,\nu_2\}$ where,

$$\nu_1 = [(m_1, w_2), (m_2, w_3), (m_3, w_1), (m_k, w_k), \ k \neq 1, 2, 3]$$
$$\nu_2 = [(m_1, w_1), (m_2, w_3), (m_3, w_2), (m_k, w_k), \ k \neq 1, 2, 3]$$

Observe that , in any stable matching m_2 must be matched with w_3; otherwise either (m_1, w_2) or (m_3, w_1) will block. Given that, there are only two other possible combinations: one where m_1 is matched with w_1 and the other where m_1 is matched to w_2. Both give rise to stable outcomes since there is no pair that will block the matching. Let $f(P) = \nu_1$. Then $f_{w_1}(P) = m_3$. Now consider the preference ordering $\hat{P}_{w_1}$ given by

$$\hat{P}_{w_1} := m_1 \hat{P}_{w_1} m_2 \hat{P}_{w_1} m_3 \hat{P}_{w_1}, \cdots, w_1$$

We claim that $\mathcal{S}(\hat{P}_{w_1}, P_{-w_1}) = \nu_2$. Observe that in any stable matching in the profile $(\hat{P}_{w_1}, P_{-w_1})$, m_3 must be matched to w_2; otherwise, either (m_1, w_1) or (m_3, w_2) will block. Also, m_2 has to be matched to w_3; otherwise, m_1 and w_1 would block the matching. Hence the only stable matching is ν_2. Then $f_{w_1}(\hat{P}_{w_1}, P_{-w_1}) = m_1$. But if $f_{w_1}(\hat{P}_{w_1}, P_{-w_1}) = m_1$ it follows from TM that, $f_{w_1}(P)$ should also be m_1. Hence $f(P) \neq \nu_1$. Therefore, $f(P) = \nu_2$. Now consider a preference ordering for m_1, $\hat{P}_{m_1}$ given by,

$$\hat{P}_{m_1} := w_2 \hat{P}_{m_1} w_3 \hat{P}_{m_1} w_1 \hat{P}_{m_1}, \cdots, \hat{P}_{m_1} m_1$$

Replicating the earlier arguments we conclude that $\mathcal{S}(\hat{P}_{m_1}, P_{-m_1}) = \nu_1$. Then $f_{m_1}(\hat{P}_{m_1}, P_{-m_1}) = w_2$. But then TM implies that $f_{m_1}(P)$ should also be w_2 i.e., $f(P) = \nu_1$. But this is a contradiction for we have shown above that $f(P) \neq \nu_1$. This completes the proof of the theorem. ■

REMARK 3.4.1: The result in Theorem 3.4.3 is valid even when $|M| \neq |W|$. Let $M = \{m_1, \cdots, m_n\}$ and $W = \{w_1, \cdots, w_m\}$. Without loss of generality assume that $m < n$. Consider the preference profile P defined in the following way: for all $k \leq m$, P_{i_k} is defined in the same way as above; for $k > m$, $P_{m_k} :=$

$w_3 P_{m_k}, \cdots, P_{m_k} m_k$. Observe that under the preference profile P any selection from the set of stable marriages divides the set of agents into three groups: men m_1, m_2 and m_3 and women w_1, w_2 and w_3 form matchings among themselves; w_k is matched to m_k for all $3 < k \leq m$ and the remaining set of men are forced to remain single. Now replicating the arguments above we obtain the result in Theorem 3.4.3.

REMARK 3.4.2: When there are only two agents on each side of the market and preferences are restricted to the set $\mathcal{D}$, Alcalde and Barberà (1994) show that the stable matching selections obtained by the man-proposing and woman-proposing deferred acceptance algorithms are both strategy-proof.

3.5 Conclusion

We have examined the implications of weakening the incentive requirement in the theory of two-sided one-to-one matching from dominant strategies to ordinal Bayesian incentive compatibility. Truth-telling is no longer assumed to be optimal for every conceivable strategy-tuple of the other players. It is only required to maximize expected utility given an agents' prior beliefs about the types of other players and the assumption that these players are following truth-telling strategies. The set of ordinal Bayesian incentive compatible stable matching procedures clearly depends on the beliefs of each agent. However, we show that when preferences are unrestricted, there is no stable matching procedure that is ordinally Bayesian incentive compatible with respect to *any* prior. When we put restrictions on the set of allowable preferences, by requiring that every agent prefers to be matched than to remain single, one obtains possibility results with independently and uniformly distributed priors. However the possibility result is non-generic. If we perturb beliefs we get back the impossibility result.

3.6 Appendix A: Deferred Acceptance Algorithm

Man Proposing Deferred Acceptance Algorithm

STEP 1: Each man makes an offer to the first woman on his preference list of acceptable women. Each woman rejects the offer of any firm that is unacceptable to her, and each woman who receives more than one acceptable offer rejects all but her most preferred of these which she "holds".

STEP K: Any man whose offer was rejected at the previous step makes an offer to his next choice (i.e., to his most preferred woman among those who have not yet rejected it), so long as there remains an acceptable woman to whom he has not yet made an offer. If a man has already made an offer to all the women he finds acceptable and has been rejected by all of them, then he makes no further offers. Each woman receiving offers rejects any from unacceptable men, and also rejects all but her most preferred among the set consisting of the new offers together with an offer she may have held from the previous step.

STOP: The algorithm stops after any step in which no man's offer has been rejected. At this point, every man is either being matched to some woman or his offer has been rejected by every woman in his list of acceptable women. The output of the algorithm is the matching at which each woman is matched to the man whose offer she is holding at the time the algorithm stops. Women who do not receive any acceptable offer or men who were rejected by all women acceptable to them remain unmatched.

3.7 Appendix B: Symmetric beliefs

In this section, we briefly discuss symmetric beliefs. For the ensuing analysis some definitions are in order. For a given preference profile, denote by P_S the preference orders of the agents in the subset $S \subseteq I$. Denote by $P_S^{m \leftrightarrow m'}$ the preference orders of the agents in S obtained from P by switching m and m', i.e., each woman in S exchanges the places of m and m' in her preference list and if m is in S his preference is $P_{m'}$ and if m' is in S its preference is P_m. Note that if woman w's true preferences are given by P_w, then $P_w^{m\leftrightarrow m'}$ is the preference in which she reverses the order of m and m' (but otherwise states her true preferences). Similarly, P_{-w} and $P_{-w}^{m\leftrightarrow m'}$ are assessments by agent w of the preferences of all other agents that are identical except that the roles of m and m' are everywhere reversed.

We model agent w's uncertainty about the about the differences in the preferences of men m and m', and about the other women's preferences for the two men as follows:

DEFINITION 3.7.1 *Given distinct men m and m' we say woman w's conditional belief $\mu(.|P_w)$ is $\{m, m'\}$ -symmetric if $\mu(P_{-w}|P_w) = \mu(P_{-w}^{m\leftrightarrow m'}|P_w)$.*

Observe that w may know a great deal about m and m' (for example w may know that both men prefer w' to some w''. What w does not know about m and m', if her conditional beliefs are $\{m, m'\}$-symmetric are any *differences* in their preferences, or in other women's preferences between them.

DEFINITION 3.7.2 *For a woman $w \in W$ and a set of men $U \subseteq M$, we say that w's conditional belief $\mu(.|P_w)$ is $\{U\}$-symmetric if it is $\{m, m'\}$-symmetric for every pair (m, m') of distinct members of U.*

If $U = M$ then woman w's belief is $\{M\}$-symmetric. We can similarly define $\{W\}$-symmetric beliefs for a man $m \in M$.

PROPOSITION 3.7.1 *(Corollary 1 in Roth and Rothblum (1999))*

For a woman with $\{M\}$-symmetric conditional belief, any strategy that changes the true preference ordering of men is stochastically dominated by a strategy that states the same number of acceptable men in their correct order.

Observe that the uniform prior $\bar{\mu}$ is $\{M\}$-symmetric for the women and $\{W\}$-symmetric for men.

Chapter 4

Top-Pair and Top-Triple Monotonicity

4.1 Introduction

An important part of social choice theory is the study of elections and committee decisions. The central object of interest in these settings is a *social choice function (SCF)*. A SCF is a mapping which associates a feasible outcome with every profile of voter preferences. It can be thought of as representing the goals of a "principal" or a "planner". A typical question that is addressed is the following: what is the class of SCFs which satisfies certain "desirable" properties or axioms in a particular environment?

An axiom that appears frequently in the literature is *Maskin Monotonicity (MM)* or *Strong Positive Association (SPA)*. This axiom expresses a very natural idea. If an outcome is selected at a given profile of preferences and if it "improves" in some agent's preference ordering, the preferences of all the others remaining the same, then this alternative must continue to be selected. More formally, consider an alternative x, an agent i and preference profiles (P_i, P_{-i}) and (P'_i, P_{-i}) such that if P_i ranks alternative x above alternative y, then P_i' also ranks x above that y. A SCF f satisfies SPA if $f(P_i, P_{-i})) = x$ implies $f(P_i', P_{-i}) = x$. For a more extensive discussion of this axiom and several variants, the reader is referred to Moulin(1983).

There is a deeper reason why MM/SPA is of great interest. It turns out to be of critical importance in the theory of strategic voting. For instance, Maskin (1999) demonstrates that if a SCF is implementable in Nash equilibrium then it satisfies SPA [1]. Moreover, SPA is "almost" sufficient for Nash-implementability. Similarly there is an extremely close relationship between SPA and the property of *strategy-proofness.* In a situation where an agent's preference ordering is private information, a strategy-proof SCF provides each agent with dominant strategy incentives to reveal their private information truthfully. Muller and Satterthwaite (1977) show that if the domain of preferences consists of all linear orders over the set of alternatives, then strategy-proofness and SPA are equivalent. Tanaka (2001) analyses this relationship in the case where indifference is permitted in individual preferences and Nehring (2000) considers the case of social choice correspondences.

What is the class of SCFs that satisfy SPA over a universal domain of preferences? An answer can be found immediately by combining the Muller-Satterthwaite result with the classical Gibbard-Satterthwaite Theorem. If there are at least three alternatives, a SCF defined over a domain consisting of all possible linear orders satisfies SPA and unanimity only if it is dictatorial [2]. In fact some proofs (e.g. Reny (2001)) of the Gibbard-Satterthwaite Theorem proceed by showing directly that SPA implies dictatorship and then exploiting the equivalence of SPA and strategy-proofness.

In this essay we show that a *much weaker requirement than SPA implies dicta-*

[1]Maskin considered the more general social choice correspondences (SCCs) which are set-valued SCFs.

[2]If indifference is permitted, an impossibility result obtains.

torship. The implications of this result are two-fold. First, it becomes clear that "very little" of SPA is required to prove dictatorship. Secondly, it demonstrates quite dramatically, the robustness of dictatorship results. Below we describe two monotonicity criteria which we call *Top-Pair Monotonicity (TPM)* and *Top-Triple Monotonicity (TTM)*. Both TPM and TTM embody the same "invariance under improvement" idea as SPA. However, the improvement is relative to a *set* rather than an alternative. Let f be a SCF. Let i be an agent and let (P_i, P_{-i}) and (P_i', P_{-i}) be two preference profiles where P_i and P_i' are such that the first k-ranked elements in these two orderings is the same. We can then think of the set of the first k-ranked alternatives in P_i (let us refer to this set as B) "improving" between (P_i, P_{-i}) and (P_i', P_{-i}). In the spirit of monotonicity we might require that if $f(P_i, P_{-i}) \in B$, then $f(P_i', P_{-i}) \in B$. The TPM condition is precisely this requirement but only for the cases of $k = 1, 2$. Similarly TTM is the same requirement but for the cases $k = 1, 2, 3$. Thus TPM is equivalent to the following two requirements:(i) if $f(P_i, P_{-i}) = x$ where x is the top-ranked alternative in P_i and x remains the top-ranked alternative in P_i', then $f(P_i', P_{-i}) = x$,(ii) if $f(P_i, P_{-i}) = x$ where $\{x, y\}$ is the set of first and second - ranked alternatives in P_i and $\{x, y\}$ remains the set of first and second - ranked alternatives in P_i', then $f(P_i', P_{-i}) \in \{x, y\}$. On the other hand, in addition to (i) and (ii) TTM requires the following third condition (iii) if $f(P_i, P_{-i}) = x$ where $\{x, y, z\}$ is the set of the three top-ranked alternatives in P_i and $\{x, y, z\}$ remains the set of first, second and third- ranked alternatives in P_i', then $f(P_i', P_{-i}) \in \{x, y, z\}$. It is clear that TTM implies TPM. We will show that SPA implies TTM

We give two characterization results. One result is for the two-person case and states that if there are at least three alternatives, a two-person SCF satisfies TPM and unanimity only if it dictatorial. A striking fact is that if there are more than two individuals TPM *does not* imply dictatorship. We give an example to show that there exist three-person SCFs that satisfy TPM and unanimity and are yet non-dictatorial. This leads to our second characterization result for an arbitrary number of individuals where we show that any SCF that satisfies TTM and unanimity must be dictatorial. We think that the result is striking because both TPM and TTM appear to be significantly weaker than SPA. It places restrictions on the SCF for improvements which occur only "near the top" of the preference ordering of an agent. It is clear that SPA, on the other hand, considers a much wider class of improvements.

The essay is organised as follows. In section 4.2 we set out the basic notation and concepts and illustrate our monotonicity criterion through some examples. Section 4.3 contains the case of two individuals. In section 4.4 we consider the general case of N individuals. Section 4.5 concludes.

4.2 Preliminaries

The set $N = \{1, ..., N\}$ 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[3] 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 4.2.1 *A Social Choice Function or (SCF) f is a mapping $f : \mathbb{P}^N \to A$.*

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

Throughout the chapter, we assume that SCFs under consideration satisfy the axiom of *unanimity.* 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 4.2.2 *A SCF f is unanimous if $f(P) = a_j$ whenever $a_j = r_1(P_i)$ for all individuals $i \in N$.*

We now introduce some definitions which are well-known in the literature.

Let P_i be an ordering and let $x \in A$. We say that P_i' represents a *x-improvement of* P_i if for all $y \in A$, $xP_iy \Rightarrow xP_i'y$.

DEFINITION 4.2.3 *The SCF f satisfies Strong Positive Association (SPA) if for all i, P_i, P_i' and P_{-i}*

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

A SCF satisfies SPA if it is the case that whenever an alternative is selected at a profile, it continues to be selected if the alternative "improves" for some agent. The next definition is that of a strategy-proof SCF. Such an SCF has the property that truth-telling is a dominant strategy for all agents.

DEFINITION 4.2.4 *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*

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

Muller and Satterthwaite (1977) characterized the relation between SPA and strategy-proofness.

THEOREM **4.2.1** *Muller and Satterthwaite(1977)*

A SCF satisfies strategy-proofness if and only if it satisfies SPA.

A special class of SCFs is described below.

DEFINITION 4.2.5 *A SCF f is dictatorial if there exists an individual i such that, for all profiles P we have $f(P) = r_1(P_i)$.*

A fundamental result which characterizes the class of strategy-proof SCFs is the Gibbard-Satterthwaite Theorem which we state below.

THEOREM 4.2.2 *Gibbard (1973), Satterthwaite (1975)*

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

The theorems above immediately lead to a characterization of the class of SCFs that satisfy SPA and unanimity over the full domain of preferences. This is the class of dictatorial SCFs.

Let t be an integer, $1 \leq t \leq m$. We now introduce the definition of *Top-t Monotonicity* where $1 \leq t \leq m$.

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

DEFINITION 4.2.6 *The SCF f satisfies Top-t Monotonicity if, for all $i \in N$, for all integers $k = 1, 2, \cdots, t$, for all P_{-i} and for all P_i and P_i' such that $B(r_k(P_i), P_i) = B(r_k(P_i'), P_i')$, we have*

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

As outlined in the Introduction, Top-t Monotonicity embodies an "invariance under set improvement" idea. Pick an integer k lying between 1 and t. Let P_i and P_i' denote two preference orderings whose set of top k, denoted by B coincides. If, at the profile (P_i, P_{-i}), f picks an element in B, then it must pick an element in

B at profile (P_i', P_{-i}) if it is to satisfy Top-t Monotonicity. The idea is that the set B has (weakly) improved in P_i vis-a-vis P_i'.

We now introduce some special cases. If $t = 2$, the resulting monotonicity property is TPM. For $t = 3$, it is TTM and if $t = 1$, we obtain Top Monotonicity (TM).

REMARK 4.2.1 Observe that Top-t Monotonicity implies Top-t' Monotonicity, where $t > t'$ and $1 \leq t, t' \leq m$.

PROPOSITION 4.2.1 *If a SCF satisfies SPA then it satisfies Top-$(m-1)$ Monotonicity.*

PROOF: Let f be a SCF satisfying SPA. We will show that f satisfies Top-$(m-1)$ Monotonicity. Let P be a preference profile. Consider any $k \in \{1, \cdots, m-1\}$. Let i be an individual and let P_i' be an ordering such that $B(r_k(P_i), P_i) = B(r_k(P_i'), P_i')$. Let $f(P) = x$ where $x \in B(r_k(P_i), P_i)$. We claim $f(P_i', P_{-i}) \in B(r_k(P_i'), P_i')$. Suppose that the claim is false. Let $f(P_i', P_{-i}) = z$ where $z \in A - B(r_k(P_i'), P_i')$. Observe that x is ranked above z in P_i'. Let $\hat{P}_i$ be another ranking such that x is the top-ranked alternative and z is the second-ranked alternative. Since $\hat{P}_i$ is a x-improvement of P_i and f satisfies SPA, $f(P_i, P_{-i}) = x$ implies $f(\hat{P}_i, P_{-i}) = x$. Again since $\hat{P}_i$ is a z-improvement of P_i' and f satisfies SPA, $f(P_i', P_{-i}) = z$ implies $f(\hat{P}_i, P_{-i}) = z$. We therefore have a contradiction. ∎

REMARK 4.2.2 Since Top-$(m-1)$ Monotonicity implies TTM which in turn implies TPM, it follows that SPA implies TTM and TPM.

4.3 The $N = 2$ Case

The main result of this section is to show that for a two-person SCF, TPM along with unanimity implies dictatorship.

THEOREM **4.3.1** *Assume $m \geq 3$. If a SCF $f : \mathbb{P}^2 \to A$ is unanimous and satisfies TPM then it is dictatorial.*

PROOF: Assume $m \geq 3$ and let $f : \mathbb{P}^2 \to A$ be a SCF satisfying unanimity and TPM. The proof follows from the two lemmas.

Lemma 4.3.1 *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.*

PROOF: Suppose not. 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 $\bar{P}_2$ such that $a\bar{P}_2 b\bar{P}_2 x$ for all $x \neq a, b$. 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 TPM that $f(P_1, P_2') \in \{a, b\}$. Suppose that $f(P_1, P_2') = b$. Observe that P_2 and P_2' have the same top element b. Applying TPM it follows that $f(P_1, P_2) = b$ which contradicts our assumption 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 and P_1' have the same top element (which is a), TPM 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 TPM 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 TPM would imply that $f(P_1, P_2) = a$ whereas we have assumed that the outcome at this profile is c. This proves Lemma 4.3.1.

Lemma 4.3.2 *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.*

PROOF: 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 Lemma 4.3.1 that $f(P_1, P_2)$ is either a or b. Holding P_2 fixed observe that the outcomes for all the profiles where a is th top-ranked outcome for individual 1 must be a, otherwise TPM will be violated. By a similar argument, holding P_1 fixed, the outcome for all profiles where b is ranked first for individual 2 can never be b. Now consider an arbitrary profile where a is ranked first for 1 and b for 2. Using Lemma 4.3.1 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. Then TPM implies that $f(P_1', P_2)$ is either a or c. But Lemma 4.3.1 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 Lemma 4.3.1 that $f(P_1, P_2')$ is either x or a. But if it is x, TPM 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 alternative provided that 2's first-ranked alternative is x. Since x is arbitrary, Lemma 4.3.2 is proved. ∎

It follows immediately from Lemma 4.3.2 that f must be dictatorial. Therefore Theorem 4.3.1 is proved. ∎

REMARK 4.3.1 The arguments above borrows from those in Sen(2001)

We now give examples to show that both parts of TPM i.e., TM and $k = 2$ monotonicity are essential for dictatorship in the two-person case.

EXAMPLE 4.3.1: We now demonstrate the existence of a non-dictatorial SCF satisfying TM but not TPM. For an arbitrary number of individuals the plurality correspondence is defined in the following way: for any profile P the correspondence selects all those alternatives (called plurality winners) which are top-ranked by the largest number of individuals. Consider a SCF f obtained from the Plurality correspondence by breaking ties at each profile according to an ordering "$>$" of the elements of A.

We claim that f satisfies TM. To see this consider a profile P where f picks an alternative a which is individual i's top-ranked alternative under P_i. Let P_i' be any other ordering where a is top-ranked. Observe that the set of Plurality winners at (P_i', P_{-i}) is exactly the same as at P. Since ties are broken with respect to the same ordering $>$, it follows immediately that $f(P_i', P_{-i}) = a$ so that TM holds.

We now claim that f does not satisfy TPM. Consider the special case of $N = \{1,2\}$ and $A = \{a,b,c\}$ with $a > b > c$. In the profiles (P_1, P_2) and (P_1, P_2') described below we have $f(P_1, P_2) = a$ and $f(P_1, P_2') = b$. But TPM would require $f(P_1, P_2') \in \{a, c\}$

$$f \begin{array}{c} \begin{array}{cc} P_1 & P_2 \end{array} \\ \begin{pmatrix} a & b \\ c & a \\ b & c \end{pmatrix} \end{array} = a \tag{4.1}$$

But,

$$f \begin{array}{c} \begin{array}{cc} P_1' & P_2 \end{array} \\ \begin{pmatrix} c & b \\ a & a \\ b & c \end{pmatrix} \end{array} = b \tag{4.2}$$

EXAMPLE 4.3.2: We now give an example to show that there exist non-dictatorial SCFs that satisfy the monotonicity criterion for $k = 2$ but not for $k = 1$. Consider a society with two individuals, $N = \{1, 2\}$ and suppose that there are four alternatives i.e., $A = \{a, b, c\}$. Define the SCF in the following way: in any profile P, if the set of the first and the second-ranked alternatives for individuals 1 and 2 are the same, (the orderings need not be the same) then the outcome is 2's top-ranked alternative. Otherwise, it is 1's top-ranked alternative.

This SCF is clearly non-dictatorial. We claim that this SCF satisfies $k = 2$ monotonicity. For any profile where the top two alternatives are the same for individuals 1 and 2, the outcome is 2's top. So if we interchange the positions of the alternatives in either 1 or 2's ranking, the outcome continues to remain 2's top-ranked alternative which obviously is in the set of the top two alternatives. Consider now a profile where the top two alternatives are different for 1 and 2.

The outcome is 1's top. If the positions of the two top-ranked alternatives in 1's preference ordering are interchanged, the outcome continues to be 1's top and thus top-two monotonicity is satisfied. Consider now the case where the two top-ranked alternatives in 2's preference ordering are interchanged. If 1's top-ranked alternative is distinct from the two top-ranked alternatives for 2, $k = 2$ monotonicity is *vacuously* satisfied. Otherwise if 1's top is the same as either 2's top or 2's second-ranked alternative, observe that $k = 2$ monotonicity is satisfied because the outcome remains unchanged.

But now, consider the profile (P_1, P_2) and (P_1', P_2) described below. We have $f(P_1, P_2,) = a$ and $f(P_1', P_2) = c$.

$$f \begin{array}{c} \begin{array}{cc} P_1 & P_2 \end{array} \\ \begin{pmatrix} a & b \\ c & a \\ b & c \end{pmatrix} \end{array} = a \tag{4.3}$$

Notice that,

$$f \begin{array}{c} \begin{array}{cc} P_1' & P_2 \end{array} \\ \begin{pmatrix} a & b \\ b & a \\ c & c \end{pmatrix} \end{array} = b \tag{4.4}$$

But TM requires $f(P_1', P_2) = a$.

4.4 The $N \geq 3$ Case

We begin this section with an example to show that if there are more than two individuals, TPM no longer implies dictatorship.

EXAMPLE 4.4.3: Let $I = \{1, 2, 3\}$, and let $A = \{a, b, c, d\}$. A SCF, f is defined in the following way. For any profile P,

(i) if the top-ranked alternatives for individuals 1 and 2 are the same then the outcome is individual 3's top-ranked outcome

(ii) if the top-ranked alternatives for individuals 1 and 2 are not the same then there are two sub-cases to consider:

(iia) Suppose that the top-ranked alternative for individual 3 is one of individual 1 or 2's top-ranked alternative. Then, f picks the alternative which is either individual 1's top or individual 2's top but *not* individual 3's top.

(iib) Suppose that individuals 1,2 and 3 have distinct top-ranked alternatives. Then f picks the (unique) alternative which is *not* top-ranked by any of the individuals 1,2 and 3.

Cases (i) and (ii) exhaust all cases, so that f is well-defined.

It is clear that f is non-dictatorial. We will show that it satisfies TPM. In order to do so, we will consider, in turn profiles satisfying cases (i), (iia) and (iib) and examine the changes in outcomes which occur when individual preferences are changed.

First consider the case of a profile satisfying (i). Assume without loss of generality that 1's and 2's top is a while 3's top is b. The outcome is then b. If 3's preferences are changed the outcome will remain 3's top. It is trivial to verify that TPM will hold for such preference changes. Consider now a change in 2's preferences (identical arguments hold for changes in 1's preferences). In order for TPM not to be satisfied vacuously, it is necessary to assume that b is ranked second in 2's initial preference ordering. Consider a new preference ordering for 2 where b is first and a second-ranked. For the new profile, case (iia) applies and the outcome is a. Clearly TPM holds.

Now consider the case of a profile satisfying (iia). Assume without loss of generality that 1,2 and 3's top-ranked alternatives are a, b and a respectively. Once again we consider changes in the preference orderings of 1,2 and 3 and show that TPM is satisfied in each case. For changes in 1's preferences, the only non-vacuous case to consider is the one where b is ranked second in the initial ordering. Suppose 1's preferences are changed so that b is ranked first and a second. Case (i) applies for the new profile and the outcome here is a. It follows that TPM holds. A very similar argument can be made for changes in 3's preferences. The only non-vacuous case to consider in order to check TPM is the one where a is ranked second in 3's preference ordering. In the changed preferences for 3, b is ranked first and a second. Case (i) applies to the new profile and the outcome is a. Clearly TPM holds. Finally consider a change in 2's preferences. Assume without loss of generality that in the initial preference ordering for 2, alternative c is second-ranked while in the new ordering, c is first-ranked and b second ranked. Case (iia) applies to the new profile. The outcome is c and TPM holds.

Finally consider the case of a profile satisfying (iib). Assume without loss of generality that 1,2 and 3's top-ranked alternatives are a, b and c respectively. Consider a preference change for 1. The only non-trivial case to consider for the purpose of verifying TPM is the one where d is ranked second. Now change 1's preference so that d is ranked first and a second. The outcome in the new profile is a (case (iib) applies) and TPM holds. An identical argument holds for changes in 2's and 3's preferences. Therefore f satisfies TPM. We now show that f does not satisfy TTM. Consider the profiles (P_1, P_2, P_3) and (P_1, P_2, P_3') below. We have

$f(P_1, P_2, P_3) = b$ and $f(P_1, P_2, P_3') = d$.

$$f \begin{array}{c} \begin{array}{ccc} P_1 & P_2 & P_3 \end{array} \\ \begin{pmatrix} a & b & a \\ b & a & c \\ c & c & b \\ d & d & d \end{pmatrix} \end{array} = b \tag{4.5}$$

$$f \begin{array}{c} \begin{array}{ccc} P_1 & P_2 & P_3' \end{array} \\ \begin{pmatrix} a & b & c \\ b & a & a \\ c & c & c \\ d & d & d \end{pmatrix} \end{array} = d \tag{4.6}$$

However since the top three alternatives in P_3 and P_3' coincide and since $f(P_1, P_2, P_3)$ is one of these alternatives, TTM requires $f(P_1, P_2, P_3')$ to be one of these alternatives too, i.e., $f(P_1, P_2, P_3') \in \{a, b, c\}$. Clearly f violates TTM. In the next section we show that it is TTM that guarantees dictatorship in the general N-person case.

Our main result in this section is to show that if TPM is strengthened to TTM, dictatorship is obtained once again.

THEOREM **4.4.1** *Assume $m \geq 3$. If a SCF $f : \mathbb{P}^N \to A$, $N \geq 3$, satisfies unanimity and TTM, then it is dictatorial.*

PROOF: We will prove the result by induction on N, the number of individuals. An important observation that we make at the outset, is that Theorem 4.4.1 is valid for the case $N = 2$. This follows from Theorem 4.3.1 and the fact that TTM implies TPM. In order to prove Theorem 4.4.1, it therefore suffices to prove Statement A below.

STATEMENT A: Pick an integer $N > 2$. Suppose for all K, $2 \leq K < N$, $f : I\!P^K \to A$ satisfies unanimity and TTM implies f is dictatorial. Then $f : I\!P^N \to A$ satisfies unanimity and TTM implies f is dictatorial.

We now prove Statement A. Assume $m \geq 3$. Let $f : I\!P^N \to A$ be a SCF that satisfies unanimity and TTM. Define a SCF $g : I\!P^{N-1} \to A$ as follows. For all $(P_1, P_3, P_4, \cdots, P_N) \in I\!P^{N-1}$,

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

In this construction individuals 1 and 2 are "coalesced" 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 TTM. 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 $k = 1, 2, 3$. 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 TTM. The only non-obvious case is the one where i is the coalesced individual $\{1, 2\}$. In this case, observe that since f satisfies TTM, $f(P_1, P_1, P_3, \cdots, P_N) \in B(r_k(P_1), P_1)$ implies that $f(P_1', P_1, P_3, \cdots, P_N) \in B(r_k(P_1'), P_1')$ which in turn implies that $f(P_1', P_1', P_3, \cdots, P_N) \in B(r_k(P_1'), P_1')$. Therefore, $g(P_1', P_3, \cdots, P_N) \in B(r_k(P_1'), P_1')$, which is what was required to be proved.

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

CASE I: The dictator is the coalesced 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 TTM, it follows immediately that h also satisfies TTM. From Theorem 4.3.1 and Remark 4.2.1, 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 a $N-2$ person profile $(P_3, \cdots, P_N)$ 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'. Let us first consider the case where $r_1(P_j) = r_1(P_j') = x$ (say). Consider a P_1 and P_2 such that $r_1(P_1) = x \neq y = r_1(P_2)$. According to our hypothesis $f(P_j, P_{-j}) = x$ and $f(P_j', P_{-j}) = y$. But since P_j and P_j' have the same top elements, it follows from TTM that, $f(P_j', P_{-j}) = x$. So when P_j and P_j' have the same top elements j cannot change the identity of the dictator. The only other case to consider is where the top elements of P_j and P_j' are distinct. Let $r_1(P_j) = x$ and $r_1(P_j') = y$ $(x \neq y)$. Pick alternative a distinct from x and y and assume that xP_1aP_1w for all $w \neq a, x$ and aP_2yP_2z for all $z \neq a, y$. Consider the preference ordering for individual j, $\bar{P}_j$ where $x\bar{P}_jy\bar{P}_ju$ for all $u \neq x, y$. By TTM, $f(\bar{P}_j, P_{-j}) = x$. Let $\hat{P}_j$ be an ordering where y is ranked first and x second. It follows from TTM that, $f(\hat{P}_j, P_{-j}) \in \{x, y\}$. But y is not the top ranked outcome for either 1 or 2. Therefore $f(\hat{P}_j, P_{-j}) = x$. Let $\hat{P}_2$ be an ordering where y and a are ranked first and second respectively. Since either 1 or 2 is the dictator in f, it follows that $f(P_1, \hat{P}_2, \cdots, \hat{P}_j, \cdots, P_N) \in \{x, y\}$. But if it is y, TTM implies that $f(\hat{P}_j, P_{-j}) \in \{a, y\}$. But by our earlier claim $f(\hat{P}_j, P_{-j}) = x$. Therefore, $f(P_1, \hat{P}_2, \cdots, \hat{P}_j, \cdots, P_N) = x$. By assumption 2 is the dictator when j's preference ordering is P_j'. Therefore $f(P_1, \hat{P}_2, \cdots, P_j', \cdots, P_N) = y$. Therefore

TTM implies $f(P_1, \hat{P}_2, \cdots, \hat{P}_j, \cdots, P_N) = y$, which contradicts our earlier claim that $f(P_1, \hat{P}_2, \cdots, \hat{P}_j, \cdots, P_N) = x$. 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 $N-1$ SCF g' by coalescing individuals 1 and 3 rather than 1 and 2 as in g. Of course, g' satisfies unanimity and TTM. 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 $x, y \neq a, b, c$. Since 1 and 3 have the same ordering in P, $f(P) = b$. Let P_3' be the ordering obtained by interchanging the positions of b and c in P_3. Since $f(P_1, P_1, P_3', \cdots, P_N) = a$, it follows from TTM that $f(P_1, P_2, P_3', \cdots, P_N) \in \{a, b\}$. Suppose $f(P_1, P_2, P_3', \cdots, P_N) = a$. Then from TTM it follows that $f(P_1, P_2, P_3, \cdots, P_N) = a$, which contradicts our earlier claim that $f(P) = b$. Thus $f(P_1, P_2, P_3', \cdots, P_N) = b$. Let $\bar{P}_3$ be an ordering obtained by interchanging the positions of a and c in P_3'. Since the top three elements of P_3' and $\bar{P}_3$ coincide it follows from TTM that $f(P_1, P_2, \bar{P}_3, \cdots, P_N) \in \{a, b, c\}$. But $f(P_1, P_2, \bar{P}_3, \cdots, P_N) \notin \{a, c\}$ otherwise TPM will be violated. Therefore $f(P_1, P_2, \bar{P}_3, \cdots, P_N) = b$. A further application of TTM 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 completes case II. Cases I and II complete the proof of Theorem 4.4.1. ∎

REMARK 4.4.1: We note that with only three alternatives, TPM suffices to guarantee dictatorship even when there are more than two individuals. The reason is that, with three alternatives TPM and TTM are equivalent, both being identical to Top-$(m-1)$ Monotonicity.

4.5 Conclusion

In this chapter we introduced new monotonicity criteria, TPM and TTM and showed that they are sufficient to force dictatorship in conjunction with unanimity. This generalizes the Muller-Satterthwaite(1977) result on the equivalence between Strong Positive Association and strategy-proofness. In this context a paper of related interest is the one by Aswal, Chatterjee and Sen(2001). They construct restricted domains of preferences where strategy-proofness implies dictatorship. The critical aspect of the preference orderings in these domains is the way alternatives are ranked at the "top". The approach in our essay may be thought of as "dual" to theirs. We consider unrestricted domains but weaken the requirement on SCFs from strategy-proofness (equivalent to SPA) to axioms (TPM and TTM) which place restrictions on the values of SCFs when changes are made at the top of individual preferences.

Chapter 5

An Axiomatic Characterization of Bayes' Rule

5.1 Introduction

Bayes' Rule is pervasive in theoretical economics, its widest use being for the purpose of updating beliefs. From the perspective of probability theory, Bayes' Rule can be derived as a consequence of the basic axioms of probability and the definition of conditional probability. This essay offers an alternative characterization of Bayes' Rule based on axioms inspired by those in the axiomatic theory of surplus sharing.

The central notion in the essay is that of a *revision rule*. Consider a situation where an agent has an initial or prior belief about the true state of the world. This belief is expressed in the form of a probability distribution over the set of "possible" states of the world, or geometrically by a point in the unit simplex of appropriate dimension. Now new information emerges which conclusively rules out the occurrence of certain states. A revision rule formulates an updated or posterior belief, which is a probability distribution over the states which remain "possible".

It is clear that Bayes' Rule is a revision rule. In particular, it redistributes the aggregate probability weight of the states which are eliminated, amongst the states which remain, in proportion to the probability that is assigned to each of these remaining states by prior belief.

Revision Rules are also sometimes described in the literature as "Evidence based Rules". Evidence based rules appear in a wide variety of related contexts. In models of learning Stahl ((1996),(1998),(1999)) introduces a family of such evidence based rules in the context of learning dynamics. Belief revision rules are widely applied in other contexts as well, one prominent area being artificial intelligence or more specifically computer simulations of autonomous agents (Bhargava and Branley (1995)). Computer simulations form an important aspect of what is known as *decision support technology* and is widely used in formulating combat or military strategies. In such computer simulations there are several schemes for representing meaningful information and various techniques for reasoning with information (Pearl (1988), Sanchez and Zadeh (1988)). In such a computer simulated world, an agent has a previous belief (prior) and a set of information at any instant. The agent combines the set of information with the previous belief using some belief revision rule to obtain a current belief. Even generalizations of probability measures such as Dempster-Shafer type belief functions (Dempster (1967), Shafer (1976)) use belief revision rules for combining ex-ante uncertainty with current information. There are many ways to formulate belief revision rules, candidates being Bayesian methods and weighted combination of beliefs.

There is an extensive literature on characterizing Bayes' Rule. Most of these characterizations rely on a *no-arbitrage* perspective. The arbitrage principle has a long history. In the literature on Bayesian statistics and decision theory it was introduced as an axiom by de Finetti (1974) to characterize subjective probability. More recently it has been proposed as a foundation for non-cooperative game theory through its dual relationship with the concept of correlated equilibrium (McCardle and Nau (1990), Nau (1991)). It has been used in McCardle and Nau (1991) to unify decision theory and equilibrium theory. However in all these settings money plays a crucial role. In environments where money is available as a medium of exchange and measurement, no-arbitrage is synonymous with subjective utility maximization in personal decisions. The fundamental point of difference between these models and ours is that we do not introduce money in our model.

The main result of the essay is a characterization of Bayes' rule in terms of axioms imposed on revision rules. The most potent of these axioms is *Path Independence*, an axiom which has been employed in a variety of contexts such as the theory of rational choice (Plott (1973)) and axiomatic bargaining (Kalai (1977)). The axiom requires that the posterior belief be unaffected by the order in which the new information appears. In section 5.3 this axiom is illustrated by means of an example. The other axioms in the characterization are relatively innocuous. One is a symmetry (or anonymity) axiom which requires that the names of the states of the world are not material for the revision rule. The continuity axiom requires the revision rule to be continuous with respect to the prior. The monotonicity axiom requires that the revised probability on a state should not be less than the prior on that state. Finally a "no mistake hypothesis" is imposed which requires that if an agent believes initially that the occurrence of a particular state is impossible, then she continues to believe this even after the arrival of new information. (actually this axiom is required only in the very special case where a revision eliminates all but only two states of the world.)

A paper, which is related in spirit to the present one, is Rubinstein and Zhou (1999). They consider a general decision situation where an agent chooses an element from a set S given a reference point e. The set S is a suitable subset of an ambient space X. For the case of updating beliefs X can be the set that includes all possible theories (point beliefs) about the world. Assuming S to be a convex subset of an Euclidean space they axiomatize the choice rule that selects a point in S that is closest to e. Their paper uses a strong symmetry axiom that forces choice decisions along the line joining the minimum distance point and e. This essay considers choices on unit simplices and characterizes a different rule.

Other than the special structure of the feasible set in this model there is another feature which distinguishes it from some other related models. This is the fact that there is no utility interpretation of the model so that axioms such as Pareto-optimality have no place here. The analysis therefore differs substantively from that in the "bargaining with claims" problem analysed in Chun and Thomson (1992). There a problem is a triple (S, e, c) with the interpretation that S is the set of feasible utility vectors, $e \in S$ is the disagreement point and $c \notin S$ is the vector of claims that cannot be fulfilled. In such a setting, Chun and Thomson characterize the proportional solution, which is similar in functional form to the Bayes' Rule. That model however emphasizes the utility interpretation of choices and as a consequence Pareto optimality is imposed as an axiom.

A similar remark is also appropriate with respect to the analysis of bankruptcy problems (see O'Neill (1982), Aumann and Maschler(1985)). The issue here is to divide the liquidation value of a bankrupt firm among its creditors. In this context, Chun (1988) characterizes the proportional solution which is again equivalent to Bayes' Rule. However he uses a strong axiom the *No-Advantageous Reallocation*(NAR) (for a discussion of NAR, see Moulin (1987)), which is a stronger version of the Pareto optimality criterion.

The chapter is organized as follows: In sections 5.2 and5. 3 we give the model and the axioms. Section 5.4 gives the main result, while section 5.5 checks the tightness of the axiomatic characterization. Section 5.6 concludes.

5.2 Model

Let T $= \{1, ..., t\}$ denote the finite set of states of the world. Let $\mathcal{P}(\mathrm{T})$ denote the class of all nonempty subsets of T. Generic elements of $\mathcal{P}(\mathrm{T})$ are denoted by P, Q, R etc. For any $P \in \mathcal{P}(\mathrm{T})$ define Δ^P= conv-hull $\{e^i\}_{i \in P}$ where e^i is a vector in $\Re^P$ for which the i-th coordinate is 1 and the rest are zeros. Thus Δ^P is the $|P| - 1$ dimensional simplex.

Before proceeding further some preliminary definitions are needed.

DEFINITION 5.2.1 *(Revision Rule:) Consider any $Q \in \mathcal{P}(T)$ and $x \in \Delta^Q$. Consider any $P \subset Q$ such that there is at least one $j \in P$ for which $x_j > 0$. A revision rule $F(P, Q, x)$ is a function that assigns a unique point $F(P, Q, x) \in \Delta^P$ with the restriction $F(Q, Q, x) = x$.*

Now $F(P, Q, x)$ is a $|P|$ dimensional vector. The i-th element is $F_i(P, Q, x)$. Thus

$$F(P, Q, x) = (F_i(P, Q, x))_{i \in P} \tag{5.1}$$

DEFINITION 5.2.2 *(Bayes' Rule:) Consider $Q \in \mathcal{P}(T)$, $x \in \Delta^Q$. Consider any $P \subset Q$ such that there exists at least one $i \in P$ for which $x_i > 0$. Then Bayes' Rule $BR(P, Q, x)$ is the revision rule having the following expression: $\forall i \in P$,*

$$BR_i(P, Q, x) = x_i + \left(\sum_{j \in Q \setminus P} x_j \right) \frac{x_i}{\sum_{k \in P} x_k} \tag{5.2}$$

5.3 Axioms

We would like to characterize Bayes' Rule. To that end we consider the following axioms

[i] **Path Independence (PI):** Consider $P, Q, R \in \mathcal{P}(\mathrm{T})$, $P \subset Q \subset R$ and $x \in \Delta^R$. A revision rule satisfies PI if and only if

$$F(P, Q, F(Q, R, x)) = F(P, R, x) \tag{5.3}$$

The expression in (3) can alternatively be written in the following way: consider $P \in \mathcal{P}(\mathrm{T})$ and take $Q_1,\ Q_2 \in \mathcal{P}(\mathrm{T})$ such that $Q_1 \supset P$ and $Q_2 \supset P$. PI then says,

$$F(P,T,x) = F(P,Q_1,F(Q_1,T,x)) = F(P,Q_2,F(Q_2,T,x)) \qquad (5.4)$$

Path Independence is a consistency requirement. Path Independence implies that the order in which information comes in does not matter. The axiom is illustrated by the following example. Suppose that a person running a high fever consults a doctor. Initial symptoms suggest to the doctor that the true disease is one in the set $\{D_1,\ D_2,\ D_3,\ D_4,\ D_5\}$. His beliefs are represented by a probability distribution over this set. The doctor orders blood test B_1 which can correctly identify D_4 and D_5 and blood test B_2 which can correctly identify D_3. Both the tests are negative. The doctor's revision rule transforms his prior beliefs into a probability distribution over $\{D_1,\ D_2\}$. Suppose the results on B_1 arrive before that on B_2. The posterior on $\{D_1,\ D_2\}$ can be thought of as passing through an intermediate belief on $\{D_1,\ D_2,\ D_3\}$. If on the other hand the report on B_2 precedes that on B_1, the prior is first revised to $\{D_1,\ D_2,\ D_4,\ D_5\}$ and eventually to $\{D_1,\ D_2\}$. If a revision rule satisfies path independence the same posterior (on $\{D_1,\ D_2\}$) obtains in both the cases.

[ii] Symmetry (SYM): Consider any $P, Q \in \mathcal{P}(\mathrm{T})$, $P \subset Q$ and $x \in \Delta^Q$. Consider any permutation function $\sigma : Q \longrightarrow Q$ such that

$$\begin{array}{ll} [\mathrm{i}] & \sigma(i) \in P \text{ if } i \in P \\ [\mathrm{ii}] & \sigma(i) = i\ \forall i \notin P \end{array}$$

A revision rule satisfies SYM if and only if

$$\forall i \in P, \qquad F_{\sigma(i)}(P,Q,\sigma(x)) = F_i(P,Q,x) \qquad (5.5)$$

$$\text{where,} \quad \sigma(x) = \left(x_{\sigma(k)}\right)_{k \in Q}$$

This is an anonymity requirement. It forces the revision rule to ignore the names of the states of the world. In the disease example, the doctor should not be putting more weight on a disease just because it carries a particular name, say tuberculosis.

[iii] **Continuity (CONT):** The revision rule $F(P, Q, x)$ is continuous in x.

The requirement here is that small changes in the priors should not lead to large changes in the revised beliefs.

[iv] **Monotonicity (MON):** Consider any $P, Q \in \mathcal{P}(\mathrm{T})$, $P \subset Q$ and $x \in \Delta^Q$. A revision rule *satisfies monotonicity* if for all $i \in P$,

$$F_i(P, Q, x) \geq x_i \tag{5.6}$$

This monotonicity requirement says that, if a state is not ruled out by some new information coming in, then the revised probability on that state is not going to be less than the prior probability.

[v] **No Mistake Hypothesis** (NM): For all $P \in \mathcal{P}(\mathcal{T})$ with $|P| = 2$, if $x_i = 0$ for some $i \in P$, then

$$F_j(P, T, x) = 1\,, \quad j \in P\,, j \neq i \tag{5.7}$$

Let us consider the disease example again. Suppose that the prior belief of a doctor about disease D_1 is zero. This axiom says that if she believes that it is impossible for disease D_1 to occur and if the test conducted does not rule out D_1, then, after the revision process, she is never going to put positive probability weight on D_1. The agent is therefore not allowed to make mistakes of a particular kind.

5.4 The Main Result

Let F be a revision rule and let $x \in \Delta^Q$, $P \subset Q \subset T$. Without loss of generality, we can write

$$F_i(P, Q, x) = x_i + \phi_i^{P,Q}(x), \qquad \forall i \in P \tag{5.8}$$

where $\phi_i^{P,Q} : \Delta^Q \longrightarrow \Re$ is a real valued function with the restriction $-x_i \leq \phi_i^{P,Q}(x) \leq 1 - x_i$, for any $x \in \Delta^Q$. Since for any $Q \in \mathcal{P}(\mathrm{T})$, $x \in \Delta^Q$ necessarily means $x \in \Delta^T$, we ignore the second superscript in $\phi_i^{P,Q}$.

THEOREM **5.4.1** *Suppose $|T| = 3$. A revision rule satisfies SYM, CONT, MON and NM if and only if it is Bayes' Rule.*

Without loss of generality we can take $T = \{1, 2, 3\}$. Before going into the proof of the theorem let us consider the following lemma.

Lemma 5.4.1 *Let $P = \{1, 2\}$, $T = \{1, 2, 3\}$. A revision rule satisfies SYM, CONT and MON if and only if there exists a continuous function $g : \Re^2{}_+ \longrightarrow \Re$ such that $\forall i \in P$, and for all $x \in \Delta^T$,*

$$F_i(P, T, x) = x_i + \frac{x_i}{x_1 + x_2}\{x_3 - 2g(x_1 + x_2,\, x_3)\} + g(x_1 + x_2,\, x_3)$$

PROOF: As mentioned above, for each i in P, the revision rule can be written as

$$F_i(P, T, x) = x_i + \phi_i^P(x_1, x_2, x_3) \tag{5.9}$$

where x_1 is the first element of the vector, x_2 is the second element and so on. Using MON we can say that $\phi_i^P(x_1, x_2, x_3) \geq 0$. Now consider a $\sigma : \{1, 2, 3\} \longrightarrow \{1, 2, 3\}$ in the following manner: $\sigma(1) = 2$, $\sigma(2) = 1$, $\sigma(3) = 3$. Then from SYM it follows,

$$F_2(P, T, x) = F_1(P, T, \sigma(x)) = x_2 + \phi_1^P(x_2, x_1, x_3)$$

Therefore, $\phi_2^P(x_1, x_2, x_3) = \phi_1^P(x_2, x_1, x_3)$. Since $F_1(P,T,x) + F_2(P,T,x) = 1$, it follows that

$$\phi_1^P(x_1, x_2, x_3) + \phi_1^P(x_2, x_1, x_3) = x_3 \tag{5.10}$$

Since $(x_1 + x_2, 0, x_3) \in \Delta^T$, it also follows that

$$\phi_1^P(x_1 + x_2, 0, x_3) + \phi_1^P(0, x_1 + x_2, x_3) = x_3 \tag{5.11}$$

Combining (5.10) and (5.11), we have

$$\phi_1^P(x_1, x_2, x_3) + \phi_1^P(x_2, x_1, x_3) = \phi_1^P(x_1 + x_2, 0, x_3) + \phi_1^P(0, x_1 + x_2, x_3) \tag{5.12}$$

Define the function $f : \Re^3 \longrightarrow \Re$ as follows:

Let $z = (z_1, z_2, z_3) \in \Re^3$,

$$f(z_1, z_2, z_3) = \phi_1^P(z_1, z_2 - z_1, z_3) - \phi_1^P(0, z_2, z_3).$$

Then, $f(x_1, x_1 + x_2, x_3) + f(x_2, x_1 + x_2, x_3)$

$$\begin{aligned} &= \phi_1^P(x_1, x_2, x_3) + \phi_1^P(x_2, x_1, x_3) - 2\phi_1^P(0, x_1 + x_2, x_3) \\ &= \phi_1^P(x_1 + x_2, 0, x_3) - \phi_1^P(0, x_1 + x_2, x_3) \\ &= f(x_1 + x_2, x_1 + x_2, x_3) \end{aligned} \tag{5.13}$$

Thus f is additive with respect to the first argument. Since f is continuous (follows from CONT), applying the theorem on Cauchy Equation to (5.13), [1] it follows that there exists a function $h : \Re^2 \longrightarrow \Re$ such that,

$$f(x_i, x_1 + x_2, x_3) = x_i h(x_1 + x_2, x_3) \tag{5.14}$$

Since $\phi_1^P(x_1, x_2, x_3) - \phi_1^P(0, x_1 + x_2, x_3) = f(x_1, x_1 + x_2, x_3)$, we have,

$$x_1 h(x_1 + x_2, x_3) + \phi_1^P(0, x_1 + x_2, x_3) = \phi_1^P(x_1, x_2, x_3) \tag{5.15}$$

[1] For a treatment of Cauchy Equations, see Eichhorn (1978)

Similarly,

$$x_2 h(x_1 + x_2, x_3) + \phi_1^P(0, x_1 + x_2, x_3) = \phi_1^P(x_2, x_1, x_3) \tag{5.16}$$

Adding (5.14) and (5.15) and using (5.10), we obtain

$$(x_1 + x_2)h(x_1 + x_2, x_3) + 2\phi_1^P(0, x_1 + x_2, x_3) = x_3$$

$$\Rightarrow h(x_1 + x_2, x_3) = \frac{1}{x_1 + x_2}\{x_3 - 2\phi_1^P(0, x_1 + x_2, x_3)\} \tag{5.17}$$

Substituting (5.17) in (5.14) and (5.15) we obtain,

$$\begin{aligned} F_i(P,T,x) &= x_i + \frac{x_i}{x_1 + x_2}\left\{x_3 - 2\phi_1^P(0, x_1 + x_2, x_3)\right\} \\ &\qquad + \phi_1^P(0, x_1 + x_2, x_3) \\ &\qquad i = 1,\ 2. \end{aligned} \tag{5.18}$$

Writing the function $\phi_1^P(0, x_1 + x_2, x_3)$ as the function $g : \Re^2{}_+ \longrightarrow \Re$, we obtain the desired conclusion. ■

Proof of Theorem 5.4.1: Without loss of generality let $P = \{1,2\}$ and $T = \{1,2,3\}$ and $x \in \Delta^T$. Let $x_1 = 0$ be given. Then by NM we have, $F_1(P,T,x) = 0$. Now from the definition of g given in (17) this implies that $g(x_1 + x_2, x_3) = 0$. Observe that $g(\ ,\ )$ is the same for all $i \in P$. Thus we have $F_i(P,T,x) = \frac{x_i}{x_1+x_2}$ for all $i \in P$. Suppose now that $x \in \Delta^T$ and $x_i > 0\ \forall i \in T$. Consider another vector $y \in \Delta^T$ defined as follows:

$$y_1 = 0,\ y_2 = x_1 + x_2,\ y_3 = x_3$$

Observe that g is the same for both x and y. But $g(y_1 + y_2, y_3) = 0$. This implies $F_i(P,T,x) = \frac{x_i}{x_1+x_2}$ for all $i \in P$ as desired. ■

Now we consider the more general case.

THEOREM 5.4.2 *Suppose $|T| \geq 4$. Then a revision rule satisfies SYM, PI, CONT, MON and NM if and only if it is Bayes' Rule.*

The proof of the theorem follows from the given lemma:

Lemma 5.4.2 *Consider T such that $|T| = t \geq 4$ and $x \in \Delta^T$. If $F(P,T,x) = BR(P,T,x)$ for all $P \in \mathcal{P}(T)$ such that $|P| = m$ $(2 \leq m < t)$, then $F(Q,T,x) = BR(Q,T,x)$ for all Q such that $|Q| > m$.*

PROOF: The following cases are considered.

Case I: Consider $x \in \Delta^T$ such that $x_k > 0$ for all $k \in T$. Consider $P, Q \in \mathcal{P}$(T) such that $Q = P \cup \{j'\}$, $j' \in T \setminus P$.

Fix an $i \in P$. From PI we get,

$$\begin{aligned} F_i(P,T,x) &= F_i(P, Q, F(Q,T,x)) \\ \Longrightarrow \phi_i^Q(x) + \phi_i^P\left((x_j + \phi_j^Q(x))_{j \in Q}\right) & \\ &= \phi_i^P(x) \end{aligned} \tag{5.19}$$

$$\begin{aligned} \Longrightarrow \phi_i^Q(x) + (x_{j'} + \phi_{j'}^Q(x)) \frac{x_i + \phi_i^Q(x)}{\sum\limits_{k \in P} x_k + \phi_k^Q(x)} & \\ = \left(\sum_{m \notin P} x_m\right) \frac{x_i}{\sum\limits_{k \in P} x_k} \end{aligned} \tag{5.20}$$

The last equality follows from the fact that $F(P,T,x) = BR(P,T,x)$. Let $x_{j'} + \phi_{j'}^Q(x) = A$. Then we get,

$$\frac{\phi_i^Q(x) + Ax_i}{1 - A} = (1 - \sum_{k \in P} x_k) \frac{x_i}{\sum\limits_{k \in P} x_k} \tag{5.21}$$

The last equality holds for any $j \in P$. So for any $j \in P$ we get,

$$\phi_j^Q(x) = \frac{x_j}{x_i} \phi_i^Q(x) \tag{5.22}$$

Now consider a P' in which a $j \in P \setminus i$ is replaced by state of the world j'. Thus $P' = (P \setminus j) \cup \{j'\}$. And one gets,

$$\phi_{j'}^{Q}(x) = \frac{x_{j'}}{x_i}\phi_i^{Q}(x) \tag{5.23}$$

Now, $\sum_{l \notin Q} x_l = \sum_{j \in Q} \phi_j^{Q}(x)$. This implies that for any $i \in Q$,

$$\phi_i^{Q}(x) = \left(\sum_{l \notin Q} x_l\right) \frac{x_i}{\sum_{j \in Q} x_j} \tag{5.24}$$

Case II: Suppose that $x_k = 0$ for some $k \in T$. Consider a P with $|P| = m$ such that $k \in P$. Consider $Q \supset P$ such that $|Q| = |P| + 1$. Proceeding as above one can show that

$$\phi_k^{Q}(x) + A\frac{\phi_k^{Q}(x)}{1-A} = 0$$
$$\Longrightarrow \phi_k^{Q}(x) = 0$$

For any other $i \in Q$ such that $x_i > 0$ application of Case A gives

$$\phi_i^{Q}(x) = \left(\sum_{l \notin Q} x_l\right) \frac{x_i}{\sum_{j \in Q} x_j} \tag{5.25}$$

Thus we have seen that given $F(P,T,x) = BR(P,T,x)$, for any P with $|P| = m$, $F(Q,T,x) = BR(Q,T,x)$ whenever $Q = P \cup \{j\}$ for any $j \in T \setminus P$. Suppose that $F(Q,T,x) = BR(Q,T,x)$ for any Q such that $m < |Q| \leq n < t$. Consider $Q' = Q \cup \{j'\}$ where $j' \in T \setminus Q$. Applying the procedure used above we can show that $F(Q',T,x) = BR(Q',T,x)$. Therefore we have the desired result. ■

PROOF OF THEOREM 5.4.2: Consider $x \in \Delta^T$ and $P \in \mathcal{P}(\mathrm{T})$. Now take $P' \subset Q \subset P$, such that $|P'| = 2$, $|Q| = 3$. Let $F(Q,T,x) = y$. Now $y \in \Delta^Q$. From Theorem 5.4.1 we get $F(P',Q,y) = BR(P',Q,y)$. Now from PI we get, $F(P',Q,y) = F(P',Q,F(Q,T,x)) = F(P',T,x)$. So, $F(P',T,x) = BR(P',T,x)$. Now from Lemma 2 we know that if $F(P',T,x) = BR(P',T,x)$,

then for any $Q \supset P'$,
$F(Q,T,x) = BR(Q,T,x)$. Since $P \supset P'$ we have $F(P,T,x) = BR(P,T,x)$. ∎

REMARK 5.4.1: There is a possible extension to the model considered above. Observe that the revision process analyzed in this essay always takes place from one set to its subsets. A possible way to extend this model would be to consider revisions that takes place from one set to another which is not necessarily a subset of the former. For the revision process to be meaningful the two sets should have non-empty intersection. Consider for example $P, Q \in \mathcal{P}(\mathrm{T})$, $P \cap Q \neq \emptyset$ and $x \in \Delta^T$. The choice rule for any such P, Q would be defined as $F(P,T,x) \in \Delta^P$ with the additional restrictions:

$$F(P,Q,F(Q,T,x)) \in \Delta^{P \cap Q} \tag{5.26}$$

In this extension let us consider an alternative version of the path independence axiom, which is due to Rubinstein-Zhou(1999).

[vi] **PI*** Consider P, $Q \in \mathcal{P}(\mathrm{T})$ $x \in \Delta^T$, $P \cap Q \neq \emptyset$. Then,

$$F(P,Q,F(Q,T,x)) = F(P \cap Q,T,x) \tag{5.27}$$

Let $T = \{1,2,3\}$. Let $P = \{1\,,\,2\}$, $Q = \{2\,,\,3\}$. From PI* we get
$F_2(P,Q,F(Q,T,x)) = F_2(\{2\},T,x) = 1$. This implies
$F_1(\{1\,,\,2\},\{2\,,\,3\},F(\{2\,,\,3\},T,x)) = 0$. Let $F(\{2\,,\,3\},T,x) = y$. Now $y_1 = 0$. i.e., $F_1(\{1\,,\,2\},\{2\,,\,3\},(0,y_2,y_3)) = 0$. Applying this to the expression in (5.17) we get $g(\,,\,) = 0$. Hence $F(P,T,x) = BR(P,T,x)$. For T with $|T| \geq 4$ the result follows from lemma 5.4.2.

Thus we get an alternative characterization:

THEOREM **5.4.3** *A choice rule satisfies SYM, MON, PI* and CONT if and only if it is Bayes' Rule.* ∎

Below we show that the five axioms are independent. For each axiom we give an example of a function that satisfies the remaining three but fails to satisfies it.

5.5 Independence of the axioms

1. Example of a revision rule that satisfies PI, MON, SYM and CONT but not NM.

Let $T = \{1,2,3,4\}$. For any $R \in \mathcal{P}(\mathrm{T})$, define $F(R, T, x)$ as follows:

$$F_i(R, T, x) = 1/r \ \forall i \in R$$
$$\text{where } r = |R|.$$

Consider $P = \{1,2\}$, $x \in \Delta^T$, $x = (0, \alpha, \beta, \gamma,)$ where $\alpha, \beta, \gamma \in (0, 1]$.This revision rule satisfies PI, MON, CONT and SYM but not NM as $F_1(\{1, 2\}, T, x) = 1/2 \neq 0$.

2. Example of a revision rule that satisfies PI, MON, SYM and NM but not CONT.

Again let $T = \{1,2,3,4\}$. For any $R \in \mathcal{P}(\mathrm{T})$, define $F(R, x)$ as follows:

$$F_i(R,T,x) = 1/m \quad \text{if } x_i > 0$$
$$= 0 \text{ otherwise.}$$
$$\text{where } m = |\{j \in M | x_j > 0\}|.$$

Let $P = \{1,2,3\}$, $x \in \Delta^T$, $x = (0, \alpha, \beta, \gamma,)$ where $\alpha, \beta, \gamma \in (0, 1]$. Consider $x_\epsilon = (3\epsilon, \alpha - \epsilon, \beta - \epsilon, \gamma - \epsilon)$; $F_1(\{1,2,3\}, T, x_\epsilon) = 1/3$
but $F_1(\{1,2,3\}, T, x) = 0$.

3. Example of a revision rule that satisfies NM, MON, SYM and CONT but not PI.

Let $T = \{1,2,3,4\}$. Define,

$$F_i(P,T,x) = \frac{x_i}{\sum\limits_{k\in P} x_k} \text{ if } |P| = 2.$$
$$= 1/p \text{ (where } p = |P|) \text{ otherwise.}$$

Consider $P = \{1,2\}$, $Q = \{1,2,3\}$. Consider $x \in \Delta^T$ such that $x = (0.1, 0.2, 0.3, 0.4)$. This revision rule satisfies NM, MON, CONT, SYM but not PI.

4. Example of a revision rule that satisfies NM, PI, CONT, SYM but not MON. Let $T = \{1,2,3,4\}$. Define the revision rule as follows:
If $R \in \mathcal{P}(\mathrm{T})$ and $|R| = 2$,

$$F_i(R,T,x) = \frac{x_i^2}{\sum\limits_{k\in R} x_k^2}$$

Otherwise,

$$F_i(R,T,x) = \frac{x_i}{\sum\limits_{k\in R} x_k}$$

Consider $x \in \Delta^T$ such that $x = (0.05, 0.85, 0.025, 0.075)$. Take $R = \{1,2\}$. Then $F_1(R,x) = 0.0034 < .05$.

5. Example of a revision rule that satisfies NM, PI, MON and CONT but not SYM.

Let $T = \{1,2,3,4\}$. For any $R \in \mathcal{P}(\mathrm{T})$ define F as follows:
If $|R| = 2$

$$F_i(R,T,x) = \frac{2x_i}{2x_i + x_j} \text{ if } i = \max\{k | k \in R\}$$
$$= \frac{x_i}{x_i + 2x_j} \text{ ootherwise.}$$

Otherwise,

$$F_i(R,T,x) = \frac{x_i}{\sum\limits_{k\in R} x_k}$$

Consider $P = \{1,2\}$ and $\sigma\{1,2,3,4\}$ as follows: $\sigma(1) = 2$; $\sigma(2) = 1$; $\sigma(3) = 3$; $\sigma(4) = 4$. Then $F_{\sigma(2)}(\{1,2\},T,\sigma(x)) = x_2/(x_2 + 2x_1)$ but $F_2(\{1,2\},T,x) = 2x_2/(x_1 + 2x_2)$. This revision rule satisfies PI, NM, MON, CONT but not SYM.

5.6 Conclusion

In this essay we axiomatically characterized Bayes' Rule by imposing axioms on the revision rule. One important feature of the model considered here is that the set of states of the world is finite. There are many economic problems where the set of states of the world is not finite. In some cases, the set is not even countable. An interesting extension of the present analysis would be to find axioms that characterize Bayes' rule in such situations.

References

1. Alcalde, J.,and S. Barberà (1994), "Top Dominance and the Possibility of Strategy-proof Stable Solutions to Matching Problems", *Economic Theory*, 4: 417-435.

2. d'Aspremont, C., and L.A. Gérard-Varet (1979), "Incentives and Incomplete Information", *Journal of Public Economics* 11: 25-45.

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

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

5. Aswal, N., Chatterjee, S. and A. Sen (2001), "Dictatorial Domains", *mimeograph.*

6. Aumann, R., and M. Maschler (1985), "Game Theoretic Analysis of a Bankruptcy Problem from the Talmud", *Journal of Economic Theory*, 36: 195-213.

7. Bhargava, H.K., and W.C. Branley (1995), "Simulating Belief Systems of Autonomous Agents", *Decision Support Systems*, 14: 329-348.

8. Chun, Y. (1988), "The Proportional Solution for the Rights Problem", *Mathematical Social Sciences*, 15: 231-246.

9. Chun, Y., and W. Thomson (1992), "Bargaining Problem with Claims", *Mathematical Social Sciences*, 24: 19-33.

10. Dempster, A.P. (1967), "Upper and Lower Probabilities induced by a Multivalued Mapping", *Annals of Mathematical Statistics*, 38: 325-339.

11. Ehlers, L. (2001), "In Search of Advice for Participants in the Matching Markets which use the Deferred-Acceptance Algorithm" *mimeograph.*

12. Eichhorn, W. (1978), *Functional Equations in Economics*, Addison-Wesley.

13. de Finetti, B. (1974), *Theory of Probability*, Wiley, New York.

14. Gale, D., and L. Shapley (1962), "College Admissions and the Stability of Marriage", *American Mathematical Monthly*, 69: 9-15.

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

16. Gibbard, A. (1977), "Manipulation of Schemes that mix Voting with Chance", *Econometrica*, 45: 665-681.

17. Green, J. and J.J. Laffont (1977), "Characterization of Satisfactory Mechanisms for the Revelation of Preferences for Public goods", *Econometrica*, 45: 427-438.

18. Harsanyi, J. (1967), "Games with Incomplete Information Played by 'Bayesian' Players: I-III", *Management Science*, 14: 159-182, 320-334, 486-502.

19. Kalai, E. (1977), "Proportional Solutions to Bargaining Situations: Interpersonal Utility Comparisons, *Econometrica*, 45: 1623-1630.

20. Ledyard, J.O. (1978), "Incentive Compatibility and Incomplete Information", *Journal of Economic Theory*, 18: 171-189.

21. Maskin, E. (1999), "Nash Equilibrium and Welfare Optimality", *Review of Economic Studies*, 66(1):23-38.

22. McCardle, K. and R. Nau (1990), "Coherent Behavior in Non-cooperative Games", *Journal of Economic Theory*, 50(2): 424-444.

23. McCardle, K. and R. Nau (1991), "Arbitrage, Rationality and Equilibrium", *Theory and Decision*, 31: 199-240.

24. Moulin, H. (1981), "Implementing Just and Efficient Decision Making", *Journal of Public Economics*, 16: 193-213.

25. Moulin, H. (1983), *The Strategy of Social Choice*, Advanced Textbooks in Economics, C.J. Bliss and M.D. Intrilligator (eds), North-Holland.

26. Moulin, H. (1987), "Equal or proportional Division of a Surplus and other Methods", *International Journal of Game Theory*, 16: 161-186.

27. Muller, E,. and M. Satterthwaite (1977), "The Equivalence of Strong Positive Association and Strategy Proofness", *Journal of Economic Theory*, 14:412-418.

28. Nau, R. (1991), "Joint Coherence in Games of Incomplete Information", *Management Science*

29. Nehring, K. (2000), "Monotonicity implies Generalized Strategy-proofness for Correspondences", *Social Choice and Welfare*, 17(2): 367-375.

30. O'Neill, B. (1982), "A Problem of Rights Arbitration in the Talmud", *Mathematical social Sciences*, 2: 345-371.

31. Pearl, J. (1988), *Probabilistic reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann Publishers Inc., San Mateo, CA.

32. Plott, C.R. (1973), "Path Independence, Rationality and Social Choice", *Econometrica*, 41: 1075-1091.

33. Reny, P. (2001), "Arrow's Theorem and the Gibbard-Satterthwaite Theorem: A Unified Approach", *Economics Letters* 99-105.

34. Roth, A. (1982), "The Economics of Matching: Stability and Incentives", *Mathematics of Operations Research*, 7: 617-628.

35. Roth, A. (1989), "Two-sided Matching with Incomplete Information about Others' Preferences", *Games and Economic Behavior*, 1: 191-209.

36. Roth, A., and U. Rothblum (1999), "Truncation Strategies in Matching Markets – in Search of Advice for Participants", *Econometrica*, 67(1): 21-43.

37. Rubinstein, A., and L. Zhou (1999), "Choice Problems with a reference Point",*Mathematical Social Sciences*, 37: 205-209.

38. 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.

39. Sanchez, E., and L.A. Zadeh (1988), *Approximate Reasoning in Intelligent Systems, Decision and Control*, Pergamon Press, New York, NY.

40. Sen, A. (2001), "Another Direct Proof of the Gibbard-Satterthwaite Theorem", *Economics Letters*, 70:381-385.

41. Shafer, G. (1976), *A Mathematical Theory of Evidence*, Princeton University Press.

42. Stahl, D. (1996), "Boundedly Rational Rule Learning in a Guessing Game", *Games and Economic Behavior*, 6: 303-330.

43. Stahl, D. (1998), "Rule Learning in Symmetric Normal Form Games: Theory and Evidence", *mimeograph*

44. Stahl, D. (1999), "Evidence Based Rules and Learning in Symmetric Normal Form Games", *International Journal of Game Theory*, 28: 111-130.

45. Tanaka, Y. (2001), "Generalized Monotonicity and Strategy-proofness: A Note", *mimeograph*