

# Nash Implementation with Partially Honest Individuals\*

Bhaskar Dutta and Arunava Sen<sup>†</sup>

July 9, 2011

## Abstract

We investigate the problem of Nash implementation in the presence of “partially honest” individuals. A partially honest player is one who has a strict preference for revealing the true state over lying when truthtelling does not lead to a worse outcome than that which obtains when lying. We show that when there are at least three individuals, all social choice correspondences satisfying No Veto Power can be implemented. If all individuals are partially honest and if the domain is separable, then all social choice functions can be implemented in strictly dominant strategies by a mechanism which does not use “integer/modulo games”. We also provide necessary and sufficient conditions for implementation in the two-person case, and describe some implications of these characterization conditions.

---

\*We are most grateful to M.Kandori, N.Kartik, H.Matsushima and T.Sjostrom for discussions and helpful comments. We are particularly indebted to two anonymous referees of the journal for remarkably insightful and detailed comments.

<sup>†</sup>Dutta is in the Department of Economics, University of Warwick, Coventry CV4 7AL, England. Sen is in the Planning Unit, Indian Statistical Institute, 7 SJS Sansanwal Marg, New Delhi 110016, India.

# 1 Introduction

The theory of mechanism design investigates the goals that a planner or principal can achieve when these goals depend on private information held by various agents. The planner designs a mechanism and elicits the private information from the agents. The cornerstone of the theory is the assumption that agents act purely to further their own *self-interest*. This is of course in common with much of the literature in economics which too assumes that individual agents' behaviors are solely motivated by material self-interest. However, there is a fair amount of both empirical and experimental evidence suggesting that considerations of fairness, and reciprocity do influence individual behavior. The recent literature in behavioral economics builds on this evidence to construct theoretical models of individual behavior.<sup>1</sup>

In this paper, we too depart from the traditional assumption that all agents are solely motivated by the pursuit of self-interest. In particular we assume that there are some agents who have a “small” intrinsic preference for honesty. In the context of mechanism design, this implies that such agents have preferences not just on the outcomes but also directly on the *messages* that they are required to send to the “mechanism designer”. Specifically, we assume the following: these agents strictly prefer to report the “true” state rather than a “false” state when reporting the former leads to an outcome (given some message profile of the other agents) which is at least as preferred as the outcome which obtains when reporting the false state (given the same message profile of the other agents). Suppose for instance, that an agent  $i$  believes that the other agents will send the message profile  $m_{-i}$ . Suppose that the true state is  $R$  and the message  $m_i$  reports  $R$  while the message  $m'_i$  reports a false state. Now let the message profiles  $(m_i, m_{-i})$  and  $(m'_i, m_{-i})$  lead to the same outcome in the mechanism, say  $a$ . Then this agent will *strictly* prefer to report  $m_i$  rather than  $m'_i$ . Of course, in the conventional theory, the agent would be indifferent between the two.

It is important to emphasize that the agent whose preferences have been described above has only a *limited* or *partial* preference for honesty.<sup>2</sup> She has a strict preference for telling the truth only when truthtelling leads to an outcome which is not worse than the outcome which occurs when she lies.

---

<sup>1</sup>For a sample of some papers, see for instance Kahneman et al. (1986), Roth (1995), Fehr and Schmidt (1999), Fehr et al. (2002), Rabin (1993).

<sup>2</sup>Of course, if agents have a very strong or outright preference for telling the truth, then the entire theory of mechanism design may be rendered trivial and redundant.

We consider such behaviour quite plausible at least for some agents.

We investigate the theory of Nash implementation pioneered by Maskin (1999)<sup>3</sup> in the presence of partially honest individuals. Our conclusion is that even a small departure from the standard model in this respect can lead to dramatically different results. In the case where there are at least three or more individuals, the presence of even a single partially honest individual implies that all social choice correspondences satisfying the weak requirement of No Veto Power can be Nash implemented. The stringent requirement of Monotonicity is no longer a necessary condition. It is vital to emphasize here that the informational requirements for the planner are minimal; although he is assumed to know of the existence of at least one partially honest agent, he does not know of her identity (or their identities).

We show that a startlingly permissive result can be obtained if further behavioural and domain assumptions are made. Specifically, if all individuals are partially honest and the domain is *separable* (a condition satisfied by classical exchange economies), then all social choice functions are implementable in strictly dominant strategies. The implementing mechanism makes no use of “integer” or “modulo” games and announcing the true state of the world is a strictly dominant strategy.<sup>4</sup>

Although our “many person” results are striking enough, an important feature of our paper is the treatment of the *two* agent case. The two-agent implementation problem is important in view of its potential applications to bilateral contracts and bargaining. We consider separately the case where there is exactly one partially honest individual and the case where both individuals are partially honest. We derive necessary and sufficient conditions for implementation in both cases under the assumption that individuals have strict preferences over outcomes. We then go on to derive some implications of these characterization conditions. In contrast to the many-person case, it turns out that non-trivial restrictions remain on the class of social choice correspondences which can be implemented *even when both individuals are known to be partially honest*. This reflects the fact that the two-person Nash implementation problem is “harder” than the many-person case. However, we show that despite these non-trivial restrictions, the pos-

---

<sup>3</sup>See Jackson(2001) and Maskin and Sjostrom (2002) for comprehensive surveys of the literature.

<sup>4</sup>However, this result should not be confused with results in the literature on strategy-proofness since our result pertains to a model where individuals know the entire state of the world.

sibilities for implementation increase. In particular, the negative result of Hurwicz and Schmeidler (1978) and Maskin (1999) no longer applies - there are sub-correspondences of the Pareto correspondence which can be implemented.

We also demonstrate that our permissive result in the three or more individuals case is robust to a particular change in the informational assumption. Specifically it is assumed that there exists a particular individual who is partially honest with a strictly positive but arbitrary probability. We show that any social choice correspondence satisfying No Veto Power can be implemented in Bayes-Nash equilibrium.

Some recent papers similar in spirit to ours, though different in substance are Matsushima (2008a), Matsushima (2008b), Kartik and Tercieux (2010) and Corchón and Herrero (2004). We discuss their work in greater detail in Section 4.1.

In the next section we describe the model and notation. Section 3 introduces the notion of partially honest individuals. Sections 4 and 5 present results pertaining to the many-person and the two-person implementation problems respectively. Section 6 analyzes the implementation problem in separable domains while Section 7 investigates a model with incomplete information model on honesty. Section 8 concludes.

## 2 The Background

Consider an environment with a *finite* set  $N = \{1, 2, \dots, n\}$  of *agents* or *individuals* and a set  $A$  of feasible *outcomes*. Each individual  $i$  has a preference ordering  $R_i$  over  $A$  where for all  $x, y \in A$ ,  $xR_iy$  signifies “ $x$  is at least as preferred as  $y$  under  $R_i$ ”. A preference profile  $R \equiv (R_1, \dots, R_n)$  specifies a preference ordering for each  $i \in N$ . For any profile  $R$  and agent  $i$ ,  $P_i$  and  $I_i$  will denote the asymmetric and symmetric components of  $R_i$  respectively. Letting  $\mathcal{R}$  be the set of all orderings over  $A$ ,  $\mathcal{R}^n$  will denote the set of all preference profiles. A *domain* is a set  $\mathcal{D} \subset \mathcal{R}^n$ . An *admissible* preference profile will be represented by  $R, R' \in \mathcal{D}$ . We will also refer to an admissible preference profile as a *state of the world*.

We assume that each agent observes the state of the world, so that there is *complete information*. Of course, the planner does not observe the state of the world. This gives rise to the implementation problem since her objective or goal does depend upon the state of the world.

**Definition 1** A social choice correspondence (*scc*) is a mapping  $f$  that specifies a nonempty set  $f(R) \subseteq A$  for each  $R \in \mathcal{D}$ . A *scc* which is always singlevalued will be called a social choice function (*scf*).

The social choice correspondence represents the goals of the planner. For any  $R \in \mathcal{D}$ ,  $f(R)$  is the set of “socially desirable” outcomes which the planner wants to achieve. Since the planner does not observe the state of the world, she has to use a *mechanism* which will induce individuals to reveal their private information.

**Definition 2** A mechanism  $g$  consists of a pair  $(M, \pi)$ , where  $M$  is the product of individual strategy sets  $M_i$  and  $\pi$  is the outcome function mapping each vector of individual messages into an outcome in  $A$ .

A mechanism  $g$  together with any state of the world induces a game with player set  $N$ , strategy sets  $M_i$  for each player  $i$ , and payoffs given by the composition of the outcome function  $\pi$  and preference ordering  $R_i$ . Let  $N(g, R)$  denote the set of *Nash equilibrium outcomes* in the game corresponding to  $(g, R)$ .

**Definition 3** A *scc*  $f$  is implementable in Nash equilibrium if there is some game form  $g$  such that for all  $R \in \mathcal{D}$ ,  $f(R) = N(g, R)$ .

We introduce some notation which we will need later on.

For any set  $B \subseteq A$  and preference  $R_i$ ,  $M(R_i, B) = \{a \in B \mid a R_i b \forall b \in B\}$ , is the set of *maximal elements* in  $B$  according to  $R_i$ . The *lower contour set* of  $a \in A$  for individual  $i$  and  $R_i \in \mathcal{R}$ , is  $L(R_i, a) = \{b \in A \mid a R_i b\}$ .

### 3 Partially Honest Individuals

With a few exceptions, the literature on implementation assumes that individuals are completely strategic - they only care about the outcome(s) obtained from the mechanism. However, it is not unrealistic to assume that at least some individuals may have an *intrinsic preference for honesty*. Of course, there are various options about how to model such a preference for honesty. In this paper, we adopt a very weak notion of such preference for honesty. In particular, we assume the following. Suppose the mechanism used by the planner requires each agent to announce the state of the world.

Then, an individual is said to have a preference for honesty if she prefers to announce the true state of the world whenever a lie does not change the outcome given the messages announced by the others. Notice that this is a very weak preference for honesty since an “honest” individual may prefer to lie whenever the lie allows the individual to obtain a more preferred outcome. An alternative way of describing an honest individual’s preference for honesty is that the preference ordering is *lexicographic* in the sense that the preference for honesty becomes operational only if the individual is indifferent on the outcome dimension.

We focus on mechanisms in which one component of each individual’s message set involves the announcement of the state of the world. We know from Maskin (1999) that there is no loss of generality in restricting ourselves to mechanisms of this kind. Therefore, consider a mechanism  $g$  in which for each  $i \in N$ ,  $M_i = \mathcal{R}^n \times S_i$ , where  $S_i$  denotes the other components of the message space. For each  $i$  and  $R \in \mathcal{D}$ , let  $T_i(R) = \{R\} \times S_i$ . For any  $R \in \mathcal{D}$  and  $i \in N$ , we interpret  $m_i \in T_i(R)$  as a *truthful* message as individual  $i$  is reporting the true state of the world.

Given such a mechanism, we need to “extend” an individual’s ordering over  $A$  to an ordering over the message space  $M$  since the individual’s preference between being honest and dishonest depends upon what messages others are sending as well as the outcome(s) obtained from them. Let  $\succeq_i^R$  denote individual  $i$ ’s ordering over  $M$  in state  $R$ . The asymmetric component of  $\succeq_i^R$  will be denoted by  $\succ_i^R$ .

**Definition 4** *Let  $g = (M, \pi)$  be a mechanism where  $M_i = \mathcal{D} \times S_i$ . An individual  $i$  is partially honest whenever for all states  $R \in \mathcal{D}$  and for all  $(m_i, m_{-i}), (m'_i, m_{-i}) \in M$ ,*

(i) *If  $\pi(m_i, m_{-i}) R_i \pi(m'_i, m_{-i})$  and  $m_i \in T_i(R)$ ,  $m'_i \notin T_i(R)$ , then  $(m_i, m_{-i}) \succ_i^R (m'_i, m_{-i})$ .*

(ii) *In all other cases,  $(m_i, m_{-i}) \succeq_i^R (m'_i, m_{-i})$  iff  $\pi(m_i, m_{-i}) R_i \pi(m'_i, m_{-i})$ .*

The first part of the definition captures the individual’s (limited) preference for honesty - she strictly prefers the message vector  $(m_i, m_{-i})$  to  $(m'_i, m_{-i})$  when she reports truthfully in  $(m_i, m_{-i})$  but not in  $(m'_i, m_{-i})$  provided the outcome corresponding to  $(m_i, m_{-i})$  is at least as good as that corresponding to  $(m'_i, m_{-i})$ .

Since individuals who are not partially honest care only about the outcomes associated with any set of messages, their preference over  $M$  is straightforward to define. That is, for any state  $R$ ,  $(m_i, m_{-i}) \succeq_i^R (m'_i, m_{-i})$  iff only  $\pi(m_i, m_{-i})R_i\pi(m'_i, m_{-i})$ .

Any mechanism together with the preference profile  $\succeq^R$  now defines a modified normal form game, and the objective of the planner is to ensure that the set of Nash equilibrium outcomes corresponds with  $f(R)$  in every state  $R$ .<sup>5</sup> We omit formal definitions.

## 4 Many Person Implementation

The seminal paper of Maskin (1999) derived a necessary and “almost sufficient” condition for Nash implementation. Maskin showed that if a social choice correspondence is to be Nash implementable, then it must satisfy a monotonicity condition which requires that if an outcome  $a$  is deemed to be socially desirable in state of the world  $R$ , but not in  $R'$ , then *some* individual must reverse her preference ranking between  $a$  and some other outcome  $b$ . This condition seems mild and innocuous. However, it has powerful implications. For instance, only constant single-valued social choice correspondence can satisfy this condition on the complete domain of preferences (Saijo (1987)). Maskin also showed that when there are three or more individuals, this monotonicity condition and a very weak condition of No Veto Power are *sufficient* for Nash implementation. No Veto Power requires that  $(n - 1)$  individuals can together ensure that if they unanimously prefer an alternative  $a$  to all others, then  $a$  must be socially desirable. Notice that this condition will be vacuously satisfied in environments where there is some good such as money which all individuals “like”. Even in voting environments where preferences are unrestricted, most well-behaved social choice correspondences such as those which select majority winners when they exist, scoring correspondences and so on, satisfy the No Veto Power condition.

These two conditions are defined formally below.

**Definition 5** *The scc  $f$  satisfies Monotonicity if for all  $R, R' \in \mathcal{D}$ , for all  $a \in A$ , if  $a \in f(R) \setminus f(R')$ , then there is  $i \in N$  and  $b \in A$  such that  $aR_i b$  and  $bP'_i a$ .*

---

<sup>5</sup>We denote this set as  $N(g, \succeq^R)$ .

**Definition 6** A scc  $f$  satisfies No Veto Power if for all  $a \in A$ , for all  $R \in \mathcal{D}$ , if  $|\{i \in N \mid aR_i b \text{ for all } b \in A\}| \geq n - 1$ , then  $a \in f(R)$ .

In this section we make the following assumption.

**Assumption A:** There exists at least one partially honest individual.

We show below that the presence a single partially honest individual renders the Monotonicity requirement redundant so that any social choice correspondence satisfying No Veto Power can now be implemented.

**Theorem 1** Let  $n \geq 3$  and suppose Assumption A holds. Then, every scc satisfying No Veto Power can be implemented.

**Proof.** Let  $f$  be any scc satisfying No Veto Power.

We prove the theorem by using a mechanism which is similar to the canonical mechanisms used in the context of Nash implementation. In particular, the message sets are identical, although there is a slight difference in the outcome function.

For each  $i \in N$ ,  $M_i = \mathcal{D} \times A \times Z_+$ , where  $Z_+$  is the set of positive integers. Hence, for each agent  $i$ , a typical message or strategy consists of a state of world  $R$ , an outcome  $a$ , and a positive integer. The outcome function is specified by the following rules :

- (R.1) : If at least  $(n - 1)$  agents announce the same state  $R$ , together with the same outcome  $a$  where  $a \in f(R)$ , then the outcome is  $a$ .
- (R.2) : In all other cases, the outcome is the one announced by  $i^*$ , where  $k_{i^*} > k_j$  for all  $j \neq i^*$ . A tie in the highest integer announced is broken in favour of the individual with the lowest index.

Let us check that this mechanism implements  $f$ .

Suppose the “true” state of the world is  $R \in \mathcal{D}$ . Let  $a \in f(R)$ . Suppose for each  $i \in N$ ,  $m_i = (R, a, k_i)$  where  $k_i \in \{1, \dots, n\}$ . Then, from (R.1), the outcome is  $a$ . No unilateral deviation can change the outcome. Moreover, each individual is announcing the truth. Hence, this unanimous announcement must constitute a Nash equilibrium, and so  $f(R) \subseteq N(g, \succeq^R)$ .

We now show that  $N(g, \succeq^R) \subseteq f(R)$ . Consider any  $n$ -tuple of messages  $m$ . Suppose no more than  $(n - 1)$  individuals announce the same state of the world  $R'$  (where  $R'$  may be distinct from  $R$ ), the same  $a \in f(R')$ . Let the

outcome be some  $b \in A$ . Then, any one of  $(n - 1)$  individuals can deviate, precipitate the integer game, *and* be the winner of the integer game. Clearly, if the original announcement is to be a Nash equilibrium, then it must be the case that  $b$  is  $R_i$ -maximal for  $(n - 1)$  individuals. But, then since  $f$  satisfies No Veto Power,  $b \in f(R)$ .

Suppose now that *all* individuals unanimously announce  $R'$ ,  $b \in f(R')$ , where  $R' \neq R$ . Then, the outcome is  $b$ . However, this  $n$ -tuple of announcements cannot constitute a Nash equilibrium. For, let  $i$  be a partially honest individual. Then,  $i$  can deviate to the truthful announcement of  $R$ , that is to some  $m_i \in T_i(R)$ . The outcome will still remain  $b$ , but  $i$  gains from telling the truth. ■

We discuss the result and related literature in the next subsection.

## 4.1 Discussion and Related Literature

A critical idea behind the construction of the mechanism used in Theorem 1 is that it is relatively “less responsive” to variations in agent strategies. For instance, according to R.1 an agent cannot change the outcome unilaterally from a unanimous announcement. Although preferences for truthful announcements of the state of the world is lexicographic, this feature of the mechanism allows these “second order” preferences to come into play and eliminate unwanted equilibria. Variants of the same idea are employed in all the results in the paper.

Another feature of the mechanism worth commenting on is that it works perfectly well irrespective of who the partially honest individuals are. This is an attractive feature for the following reason. Since this is a model of complete information, agents know the state of the world, and hence have this information (a model with incomplete information is explored in Section 7). However, the planner does not know the state of the world and hence does not know the identity of the partially honest individual(s), and so cannot construct a mechanism which uses information about the identity of the partially honest individual(s).

It is possible to interpret our permissive result in terms of Monotonicity. In our model where individuals have preferences over state announcements and alternatives, preference reversals must be defined over state, alternative *pairs*. Suppose  $a \in f(R) \setminus f(R')$ . Then, a partially honest individual has a preference reversal over the pair  $(a, R)$  and  $(a, R')$  in the sense that  $(a, R) \succ_i^R$

$(a, R')$  and  $(a, R') \succ_i^{R'} (a, R)$ . Thus, one may think of Monotonicity being trivially satisfied in this model.<sup>6</sup>

Matsushima (2008b) focuses on Nash implementation with honest players. However, there are several differences between his framework and ours. In his framework, the social choice function selects a *lottery* over the basic set of outcomes. Individuals have VNM preferences over lotteries. He also assumes that *all* players have an intrinsic preference for honesty, suffering a small utility loss from lying. In his framework, the planner can also impose small fines on the individuals.<sup>7</sup> In this setting, he shows that when there are three or more individuals, every social choice function is implementable in the iterative elimination of strictly dominated strategies, and hence in Nash equilibrium when there are *three* or more individuals. Matsushima (2008a) focuses on the incomplete information framework and proves a similar permissive result for Bayesian implementation when players suffer a small utility loss from lying.

Another paper related to ours is Kartik and Tercieux (2010). They analyze a generalization of the classical complete information implementation problem where agents can send evidence regarding the state of the world. They provide a necessary condition for implementation in these environments; the condition together with the No Veto Power condition, is also sufficient for implementation. They also consider a model of costly evidence fabrication and establish a permissive result for implementation (in the  $n \geq 3$  case) which is similar in spirit to our Theorem 1.

An earlier paper that provides an alternative behavioral approach to implementation is Corchón and Herrero (2004). Instead of assuming preferences on state announcements that depend on the true state, they propose restrictions on the set of admissible announcements (or strategies) that depend on the true state. They refer to these strategies as *decent* strategies and investigate the class of sccs that can be Nash implemented in decent strategies. They obtain a relatively permissive result (without Monotonicity but a stronger NVP condition) for a particular formulation of decent strategies.

---

<sup>6</sup>We are grateful to a referee for this observation. See also Kartik and Tercieux (2010) for a similar observation.

<sup>7</sup>Notice that our model is purely ordinal and so also accommodates the important context of voting problems.

## 5 Two-person Implementation

The two-person implementation problem is an important one theoretically. However it is well-known that analytically, it has to be treated differently from the “more than two” or many-person case. The general necessary and sufficient condition for the two-person case are due to Dutta and Sen (1991) and Moore and Repullo (1990).<sup>8</sup> These conditions are more stringent than those required for implementation in the many-person case. Monotonicity remains necessary; in addition, some non-trivial conditions specific to the two-person case also become necessary.

Theorem 1, as well as the results of Matsushima (2008b) and Kartik and Tercieux (2010) show that when there are at least three individuals, the presence of a partially honest player allows for a very permissive result since monotonicity is no longer a necessary condition for implementation. In this section, we investigate social choice correspondences which are implementable under two alternative scenarios - when there is exactly one partially honest individual, as well as when both individuals are partially honest.

In order to simplify notation and analysis, we shall assume throughout this subsection that the admissible domain consists of *strict orders*, i.e. indifference is not permitted. Later we shall discuss some of the complications which arise when indifference in individual orderings is permitted. We will also discuss briefly why the essential message(s) derived from this section will not change if individual preferences are not restricted to be strict orders. We shall write individual  $i$ 's preference ordering as  $P_i$  with the interpretation that if  $xP_iy$ , then “ $x$  is strictly preferred to  $y$  under  $P_i$ ”. The set of all strict preference orderings is denoted by  $\mathcal{P}$ . A preference profile or state of the world will be denoted by  $P \equiv (P_1, \dots, P_n)$ . The set of states will continue to be denoted by  $\mathcal{D} \subset \mathcal{P}^n$ . For any set  $B \subseteq A$  and preference  $P_i$ , the set of maximal elements in  $B$  according to  $P_i$  is  $M(P_i, B) = \{a \in B \mid \text{there does not exist } b \text{ such that } bP_ia\}$ . The *lower contour set* of  $a \in A$  for individual  $i$  and  $P_i \in \mathcal{D}$ , is  $L(P_i, a) = \{b \in A \mid aP_ib\} \cup \{a\}$ . Other definitions carry over to this setting with appropriate notational changes.

Our results establish two general facts. The first is that the necessary conditions for implementation are restrictive in the two-person case even when both individuals are partially honest. For instance, no correspondence which contains the union of maximal elements of the two individuals, is im-

---

<sup>8</sup>See also Busetto and Codognato (2009).

plementable. We also show that if the number of alternatives is *even*, then no anonymous and neutral social choice correspondence is implementable. The second fact is that the presence of partially honest individuals makes implementation easier relative to the case when individuals are not partially honest. Consider, for instance, a classic result due to Hurwicz and Schmeidler (1972) and Maskin (1999) which states that if a two-person, Pareto efficient, social choice correspondence defined on the domain of all possible strict orderings is implementable, then it must be dictatorial. We show that this result no longer holds when both individuals are partially honest. To summarize: the presence of partially honest individuals, ameliorates the difficulties involved in two-person implementation relative to the case where individuals are not partially honest; however, unlike the case of many-person implementation with partially honest individuals, it does not remove these difficulties completely.

We now proceed to the analysis of the two cases.

## 5.1 Both Individuals Partially Honest

In this subsection, we make the following informational assumption.

**Assumption A2:** Both individuals are partially honest.

A fundamental condition for implementation in this case is stated below. In what follows, we shall refer to the players as  $i$  and  $j$ .

**Definition 7** A scc  $f$  satisfies Condition  $\beta^2$  if there exists a set  $B$  which contains the range of  $f$ , and for each  $i \in N$ ,  $P \in \mathcal{D}$  and  $a \in f(P)$ , there exists a set  $C(P_i, a) \subseteq B$  with  $a \in C(P_i, a) \subseteq L(P_i, a)$  such that

$$(i) \ C(P_i, a) \cap C(P_j^1, b) \neq \emptyset \text{ for all } P^1 \in \mathcal{D} \text{ and } b \in f(P^1).$$

$$(ii) \ [a \in M(P_i, B) \cap M(P_j, B)] \Rightarrow [a \in f(P)].$$

Condition  $\beta^2$  comprises two parts. The first is an *intersection property* which requires appropriate lower contour sets to have a non-empty intersection. The second is a *unanimity condition* which requires alternatives which are maximal in an appropriate set for both individuals, to be included in the value of the scc at that state. Conditions of this sort are familiar in the literature on two-person Nash implementation.

**Theorem 2** *Assume  $n = 2$  and suppose Assumption A2 holds. Let  $f$  be a SCC defined on a domain of strict orders. Then  $f$  is implementable if and only if it satisfies Condition  $\beta^2$ .*

**Proof.** We first show if a scc  $f$  is implementable, it satisfies Condition  $\beta^2$ .

Let  $f$  be an implementable scc and let  $g = (M, \pi)$  be the mechanism which implements it. Let  $B = \{a \in A \mid \pi(m) = a \text{ for some } m \in M\}$ . For each  $P \in \mathcal{D}$  and  $a \in f(P)$ , let  $m^*(P, a)$  be the Nash equilibrium strategy profile with  $\pi(m^*(P, a)) = a$ . Such a strategy profile must exist if  $f$  is implementable. For each  $i$ , let  $C(P_i, a) = \{c \in A \mid \pi(m_i, m_j^*(P, a)) = c \text{ for some } m_i \in M_i\}$ . It follows that  $a \in C(P_i, a) \subseteq L(P_i, a)$  and that  $C(P_i, a) \subseteq B$ .

Take any  $P^1 \in \mathcal{D}$  and  $b \in f(P^1)$ , and suppose  $x = \pi(m_i^*(P, b), m_j^*(P^1, a))$ . Then,  $x \in C(P_i^1, a) \cap C(P_j, b)$ . Hence,  $f$  satisfies (i) of Condition  $\beta^2$ .

Now fix a state  $P \in \mathcal{D}$  and let  $a \in A$  be such that  $a \in M(P_i, B) \cap M(P_j, B)$ . Since  $a \in B$ , there exists a message profile  $m$  such that  $\pi(m) = a$ . If  $m$  is a Nash equilibrium in state  $P$ , then  $a \in f(P)$ . If  $m$  is not a Nash equilibrium, then there is an individual, say  $i$  and  $\hat{m}_i \in T_i(P)$  such that  $\pi(\hat{m}_i, m_j) \succ_i^P a$ . However, since  $P_i$  is a strict order,  $a \in M(P_i, B)$  and  $\pi(\hat{m}_i, m_j) \in B$  implies that either  $a P_i \pi(\hat{m}_i, m_j)$  or  $a = \pi(\hat{m}_i, m_j)$ . So,  $\pi(\hat{m}_i, m_j) \succ_i^P a$  implies that  $\pi(\hat{m}_i, m_j) = a$  must hold. If  $(\hat{m}_i, m_j)$  is a Nash equilibrium, then again  $a \in f(P)$ . Otherwise,  $j \neq i$  deviates to some  $\hat{m}_j \in T_j(P)$  with  $\pi(\hat{m}_i, \hat{m}_j) = a$  (using the same argument as before). But there cannot be any further deviation from  $(\hat{m}_i, \hat{m}_j)$ , and so  $a \in f(P)$ . Hence  $f$  satisfies (ii) of Condition  $\beta^2$ .

To prove sufficiency, let  $f$  be any scc satisfying Condition  $\beta^2$ . For all  $P \in \mathcal{D}$  and  $a \in f(P)$ , let  $C(P_i, a)$  and  $B$  be the sets specified in Condition  $\beta^2$ .

For each  $i$ , let  $M_i = \mathcal{D} \times A \times A \times \{T, F\} \times Z_+$ . The outcome function  $\pi$  is defined as follows.

- (i) If  $m_i = (P, a, b, T, k^i)$  and  $m_j = (P, a, c, T, k^j)$  where  $a \in f(P)$ , then  $\pi(m) = a$ .
- (ii) If  $m_i = (P, a, c, T, k^i)$  and  $m_j = (P^1, b, d, T, k^j)$  where  $a \in f(P)$  and  $b \in f(P^1)$  with  $(a, P) \neq (b, P^1)$ , then  $\pi(m) = x$  where  $x \in C(P_i^1, b) \cap C(P_j, a)$ .
- (iii) If  $m_i = (P, a, c, F, k^i)$  and  $m_j = (P^1, b, d, T, k^j)$ , with  $a \in f(P)$  and  $b \in f(P^1)$ , then  $\pi(m) = c$  if  $c \in C(P_i^1, b)$  and  $\pi(m) = b$  otherwise.

- (iv) In all other cases, the outcome is the alternative figuring as the third component of  $m_{i^*}$ , where  $i^*$  is the winner of the integer game.

The mechanism is similar, but not identical to that used by Dutta and Sen (1991). Essentially, when both individuals announce the same  $(P, a)$  where  $a \in f(P)$  and  $T$ , then the mechanism identifies this as the “equilibrium” message for  $(P, a)$ . However, if the two individuals send these equilibrium messages corresponding to *different* states of the world, then the planner cannot identify which individual is telling the truth, and so the outcome corresponding to these conflicting messages has to be in the intersection of the appropriate lower contour sets. If one individual appears to be sending the equilibrium message corresponding to  $(P, a)$ , while the other individual  $i$  announces  $F$  instead of  $T$  (even if the other components correspond to some equilibrium), then the latter individual is allowed to select any outcome in  $C(P_i, a)$ . Finally, in all other cases, the integer game is employed.

Let us check that this mechanism implements  $f$  in Nash equilibrium. Throughout the remaining proof, let the true state be  $P$ .

Consider any  $a \in f(P)$ . Let  $m_i^* = (P, a, ., T, k^i)$  for both  $i$ . where  $k^i$  is any positive integer. Then,  $\pi(m^*) = a$ . Any deviation by  $i$  can only result in an outcome in  $L(P_i, a)$ , and so  $m^*$  must be a Nash equilibrium.

We complete the proof of Sufficiency by showing that *all* Nash equilibrium outcomes are in  $f(P)$ . Consider a message profile  $m$  and suppose that it is a Nash equilibrium in state  $P$ . We consider all possibilities below.

**Case 1:** Suppose  $m_i = (P^1, a, ., T, k^i)$  for all  $i$ , where  $a \in f(P^1)$ . Then,  $\pi(m) = a$ . If  $P = P^1$ , there is nothing to prove, since  $a \in f(P)$ . Assume therefore that  $P \neq P^1$ .<sup>9</sup> Let  $i$  deviate to  $m'_i = (P, b, a, F, k^i)$ , where  $b \in f(P)$ . Then  $\pi(m'_i, m_j) = a$ . But, this remains a profitable deviation for  $i$  since  $m'_i \in T_i(P)$ . Hence  $m$  is not a Nash equilibrium.

**Case 2:** Suppose  $m_i = (P^1, a, c, T, k^i)$  and  $m_j = (P^2, b, d, T, k^j)$  where  $a \in f(P^1)$  and  $b \in f(P^2)$  with  $(a, P^1) \neq (b, P^2)$ . Then  $\pi(m) = x$  where  $x \in C(P_i^2, b) \cap C(P_j^1, a)$ . Suppose that  $P = P^1 = P^2$  does not hold. So, either  $P^1 \neq P$  or  $P^2 \neq P$ . Suppose w.l.o.g that  $P^1 \neq P$ . Then,  $i$  can deviate to  $m'_i \in T_i(P)$  such that  $m'_i = (P, b, x, F, k^i)$ . Since  $x \in C(P_i^2, b)$  by assumption,  $\pi(m'_i, m_j) = x$ . However,  $m'_i \in T_i(P)$  so that  $i$  gains by deviating. Hence  $m$  is not a Nash Equilibrium.

---

<sup>9</sup>That is, both individuals “coordinate” on a lie about the state of the world.

Suppose instead that  $P = P^1 = P^2$  holds and  $m$  is a Nash equilibrium. Then it must be the case that  $x \in M(P_i, C(P_i, b))$  (Recall that  $P = P^i$ ), i.e.  $xP_ib$ . However  $bP_ix$  since  $x \in C(P_i, b)$  by assumption. Since  $P_i$  is a strict order, we have  $b = x$ . Since  $b \in f(P)$ , it follows that  $\pi(m) \in f(P)$ .

**Case 3:** Suppose  $m_i = (P^1, a, c, F, k^i)$  and  $m_j = (P^2, b, d, T, k^j)$  where  $a \in f(P^1)$  and  $b \in f(P^2)$ . Then  $\pi(m) = x$  where  $x \in C(P_i^2, b)$ . Suppose that  $P = P^1 = P^2$  does not hold. So, either  $P^1 \neq P$  or  $P^2 \neq P$ . Suppose first, that  $P^1 \neq P$ . Then, replicating the argument in Case 2 above, it follows that  $i$  can profitably deviate to  $m'_i \in T_i(P)$  such that  $m'_i = (P, b, x, F, k^i)$  establishing that  $m$  is not a Nash Equilibrium. Suppose then that  $P^2 \neq P$ . Then,  $j$  can deviate to  $m'_j \in T_j(P)$  such that  $m'_j = (P, b, x, F, k^j)$  and win the integer game (by a suitable choice of  $k^j$ ). Then  $\pi(m_i, m'_j) = x$  and  $m'_j$  is a profitable deviation since  $m'_j \in T_j(P)$ . Hence  $m$  is not a Nash Equilibrium.

The only remaining case is  $P = P^1 = P^2$ . Observe that since  $m$  is a Nash equilibrium,  $x \in M(P_i, C(P_i, b))$ , i.e.  $xP_ib$ . Since  $bP_ix$  as well, we have  $b = x$  since  $P_i$  is a strict order. Since  $b \in f(P)$  by hypothesis, we conclude that  $\pi(m) \in f(P)$ .

**Case 4:** The remaining possibility is that  $m$  is such that the integer game decides the outcome. In this case, if  $m$  is a Nash equilibrium, then  $\pi(m) \in M(P_i, B) \cap M(P_j, B)$ . From (ii) of Condition  $\beta^2$ , we have  $\pi(m) \in f(P)$ . ■

## 5.2 Exactly One Partially Honest Individual

Here we make the following informational assumption.

**Assumption A1:** There is exactly one partially honest individual.

The condition which is necessary and sufficient for implementation under Assumption A1 (assuming strict orders) is slightly more complicated than the earlier case.

**Definition 8** *The scc  $f$  satisfies Condition  $\beta^1$  if there is a set  $B$  which contains the range of  $f$ , and for each  $i \in N$ ,  $P \in \mathcal{D}$  and  $a \in f(P)$ , there exists a set  $C(P_i, a) \subseteq B$  with  $a \in C(P_i, a) \subseteq L(P_i, a)$  such that*

(i)  $C(P_i, a) \cap C(P_j^1, b) \neq \emptyset$  for all  $P^1 \in \mathcal{D}$  and for all  $b \in f(P^1)$ .

(ii) for all  $P^2 \in \mathcal{D}$ ,  $[a \in M(P_i^2, B) \cap M(P_j^2, B)] \Rightarrow [a \in f(P^2)]$ .

(iii) for all  $P^2 \in \mathcal{D}$ , if  $b \in C(P_i, a)$  and  $b \in M(P_i^2, C(P_i, a)) \cap M(P_j^2, B)$ , then  $b \in f(P^2)$ .

The only difference between Conditions  $\beta^1$  and  $\beta^2$  is the extra requirement (iii) in the former. Our next result shows that Condition  $\beta^1$  is the exact counterpart of Condition  $\beta^2$  in the case where Assumption A1 holds.

**Theorem 3** *Assume  $n = 2$  and suppose Assumption A1 holds. Let  $f$  be a SCC defined on a domain of strict orders. Then  $f$  is implementable if and only if it satisfies Condition  $\beta^1$ .*

**Proof.** Again, we start with the proof of necessity. Let  $(M, \pi)$  be the mechanism which implements  $f$ . Consider part(i) of Condition  $\beta^1$ . Clearly, the intersection condition remains necessary.

The proof of part (ii) of Condition  $\beta^1$  is similar though not identical to the proof of its counterpart in  $\beta^2$ . Let  $P^2 \in \mathcal{D}$  and consider  $a$  such that  $a \in M(P_i^2, B) \cap M(P_j^2, B)$ . Since  $a \in B$ , there exists a message profile  $m$  such that  $\pi(m) = a$ . Suppose w.l.o.g. that  $i$  is the partially honest individual. If  $m$  is not a Nash equilibrium, then it must be the case that there exists  $\hat{m}_i \in T_i(P^2)$  such that  $\pi(\hat{m}_i, m_j) P_i^2 a$ . However,  $a \in M(P_i^2, B)$  and  $\pi(\hat{m}_i, m_j) \in B$  implies that  $\pi(\hat{m}_i, m_j) P_i^2 a$ . Since  $P_i^2$  is a strict order, we must have  $\pi(\hat{m}_i, m_j) = a$ . Since  $a \in M(P_j^2, B)$ , it must be the case that  $(\hat{m}_i, m_j)$  is a Nash equilibrium and  $a \in f(P^2)$ .

Consider part (iii) of Condition  $\beta^1$ . Let  $P^2 \in \mathcal{D}$ . We need to show that if  $b \in C(P_i, a)$  and  $b \in M(P_i^2, C(P_i, a)) \cap M(P_j^2, B)$ , then  $b \in f(P^2)$ . Let  $\pi(m) = b$  with  $m_j = m_j^*(P, a)$  being the equilibrium message of  $j$  supporting  $a$  as a Nash equilibrium when the state is  $P$ . Suppose the state is  $P^2$ . Let  $i$  be the partially honest individual. Since  $b \in M(P_i^2, C(P_i, a))$ ,  $i$  can have a profitable deviation from  $m_i$  only if  $m_i \notin T_i(P^2)$  and there is  $m'_i \in T(P^2)$  such that  $\pi(m'_i, m_j) = b$ , the last fact following from our assumption that  $P_i^2$  is a strict order. But, now consider  $(m'_i, m_j)$ . Individual  $i$  cannot have a profitable deviation since  $m'_i \in T(P^2)$  and  $b$  is  $P_i^2$ -maximal in  $C(P_i, a)$ . Neither can  $j$  since  $b$  is  $P_j^2$ -maximal in  $B$  and  $j$  is not partially honest. So,  $(m'_i, m_j)$  must be a Nash equilibrium corresponding to  $P^2$ , and hence  $b \in f(P^2)$ .

We now turn to the proof of sufficiency. Let  $f$  be any scc satisfying Condition  $\beta^1$ . Consider the same mechanism used in the proof of Theorem 2.

Let  $P$  be the true state of the world. The proof that every  $a \in f(P)$  is supported as a Nash equilibrium is identical to that of Theorem 2.

We need to show that every outcome corresponding to a Nash equilibrium is in  $f(P)$ . Let  $m$  be any candidate Nash equilibrium strategy profile. Once again, we consider all possibilities exhaustively. Suppose  $m$  is covered by Case 1 of Theorem 2. Then, the proof is identical to that of Theorem 2 since the partially honest individual can deviate to a truth-telling strategy without changing the outcome.

**Case 2:** Suppose  $m_i = (P^1, a, c, T, k^i)$  and  $m_j = (P^2, b, d, T, k^j)$  where  $a \in f(P^1)$  and  $b \in f(P^2)$  with  $(a, P^1) \neq (b, P^2)$ . Let  $\pi(m) = x \in C(P_i^2, b) \cap C(P_j^1, a)$ . Suppose w.l.o.g that  $i$  is the partially honest individual. We claim that  $P^1 = P$ . Otherwise  $i$  can deviate to  $m'_i = (P, z, x, F, k^i)$  so that  $\pi(m'_i, m_j) = x$  since  $x \in C(P_i^2, b)$ . Since  $m_i \notin T(P)$  while  $m'_i \in T(P)$ , it follows that the deviation is profitable and  $m$  is not a Nash equilibrium.

Suppose therefore that  $P^1 = P$ . Since  $m$  is a Nash equilibrium, it must be true that  $x \in M(P_j, C(P_j, a))$ , i.e  $xP_ja$ . However, since  $x \in C(P_j, a)$  and  $P_j$  is a strict order, we must have  $x = a$ . Since  $a \in f(P)$  by assumption, we have shown  $\pi(m) \in f(P)$  as required.

**Case 3:** Suppose  $m_i = (P^1, a, c, F, k^i)$  and  $m_j = (P^2, b, d, T, k^j)$  where  $a \in f(P^1)$  and  $b \in f(P^2)$ . Let  $\pi(m) = x$ . We know that  $x \in C(P_i^2, b)$ . Suppose  $P \neq P^1$  and  $P \neq P^2$  hold. As we have seen in the proof of Case 3 in Theorem 2, both individuals can unilaterally deviate to a truth-telling strategy without changing the outcome. The partially honest individual will find this deviation profitable contradicting our hypothesis that  $m$  is a Nash equilibrium.

Suppose  $P = P^1$ , i.e  $i$  is the partially honest individual. Note that individual  $j$  can trigger the integer game and obtain any alternative in  $B$  by unilateral deviation from  $m$  while  $i$  can obtain any alternative in  $C(P_i^2, b)$  by unilateral deviation from  $m$ . Since we have assumed that  $m$  is a Nash equilibrium in state  $P$ , it must be the case that  $x \in M(P_i, C(P_i^2, b)) \cap M(P_j, B)$ . Then by part (iii) of Condition  $\beta^1$ , we have  $x \in f(P)$ .

Suppose  $P = P^2$ , i.e  $j$  is the partially honest individual. By the same argument as in the previous paragraph, we have  $x \in M(P_i, C(P_i, b))$ , i.e.  $xP_ib$ . But  $x \in C(P_i, b)$  implies  $bP_ix$ . Since  $P_i$  is a strict order, we have  $b = x$ . Since  $b \in f(P)$ , we have  $\pi(m) \in f(P)$  as required.

**Case 4:** The remaining possibility is that  $m$  is such that the integer game decides the outcome. We use the same argument as in Case 4, in Theorem

2, i.e if  $m$  is a Nash equilibrium, then  $\pi(m) \in M(P_i, B) \cap M(P_j, B)$  and applying (ii) of Condition  $\beta^1$  to conclude that  $\pi(m) \in f(P)$ . ■

Theorems 2 and 3 show that a simple characterization of implementable sccs is possible under the assumption that all admissible preferences are strict orders. Matters are more complicated and subtle when we allow for indifference. For instance, even Unanimity is no longer a necessary condition. Let  $R$  be a state of the world where the maximal elements for  $i$  and  $j$  are  $\{a, b\}$  and  $\{a, c\}$ . We can no longer argue that  $a$  is  $f$ -optimal at this state for the following reason. Suppose that the message profile  $m$  which leads to  $a$  involves individual  $i$  announcing a non-truthful state of the world. However a truthful message from  $i$  (holding  $j$ 's message constant) may lead to the outcome  $b$  which is not maximal for  $j$ . If this is the case, then  $m$  is no longer a Nash equilibrium. It is not difficult to show that a weaker Unanimity condition is necessary.<sup>10</sup>

However, the main point brought out by our characterization results do not change even if the domain of individual preference profiles is extended to allow for indifference. The two theorems seek to establish the following point - even when players are partially honest, one cannot hope to get the kind of permissive results in the two-person case that one gets in the many-person case. The reason is that while partial honesty helps to dispense with Monotonicity, the Intersection condition continues to be necessary in the two-person case.<sup>11</sup> We show in the next subsection that it is the intersection condition which severely restricts the class of implementable sccs. Hence, the negative implications (Propositions 1 and 2) remain true even when the domain is enlarged to allow for indifference. So, by restricting attention to strict preferences, we are able to establish the difference between the two-person and many-person cases in a relatively simple way, without providing a full-blown analysis which would entail cumbersome notation and technicalities without providing much additional insight.

### 5.3 Implications

In this section, we briefly discuss the implications of our results in the two-player case.

---

<sup>10</sup>A recent paper which extends Theorems 2 and 3 to allow for indifference is Lombardi and Yoshinara (2011).

<sup>11</sup>If indifference is permitted, redefine  $L(R_i, a) = \{b \in A | aR_i b\}$ .

Define a *worst outcome* relative to the scc  $f$  to be some  $w \in A$  such that for all  $i \in N$ , for all  $R \in D$ , for all  $a \in f(R)$ ,  $a P_i w$ . That is, in all states of the world, a worst outcome is ranked below any outcome in  $f$  by all individuals. Notice that the existence of a worst outcome relative to  $f$  ensures that the intersection condition (part (i) of Assumption A2 or Assumption A1) is satisfied. The following is an easy corollary of Theorem 2.

**Corollary 1** *Let  $n = 2$ , and suppose Assumption A2 is satisfied. Suppose  $w$  is a worst outcome relative to a scc  $f$ , which satisfies part (ii) of Condition  $\beta^2$ . Then,  $f$  is implementable in Nash equilibrium.*

Of course, an analogous corollary follows if Assumption A1 is satisfied.

We now present some other less obvious implications of Theorems 2 and 3. It is clear that Condition  $\beta^1$  implies Condition  $\beta^2$ . We first show that even the weaker Condition  $\beta^2$  imposes non-trivial restrictions on the class of implementable sccs.

**Proposition 1** : *For all  $P \in \mathcal{P}^n$ , let  $M(P_i, A) \cup M(P_j, A) \subseteq f(P)$ . Then,  $f$  is not implementable under Assumption A2.*

**Proof.** : Consider the profiles  $P \in \mathcal{P}^2$  given below.

- (i)  $a P_i d P_i b$  for all  $d \notin \{a, b\}$ .
- (i)  $b P_j d P_j a$  for all  $d \notin \{a, b\}$ .

Then,  $L(P_i, b) \cap L(P_j, a) = \emptyset$ , and so Condition  $\beta^2$  is not satisfied. ■

According to Proposition 1, no scc which is a superset of the correspondence which consists of the union of the best-ranked alternatives of the two players, is implementable even when both individuals are partially honest. An immediate consequence of this result is that the Pareto correspondence is not implementable.

The next proposition is another impossibility result, showing that no anonymous and neutral scc can be implemented even when both individuals are partially honest provided the number of alternatives is even. Anonymity and Neutrality are symmetry requirements for sccs with respect to individuals and alternatives respectively. They are pervasive in the literature but we include formal definitions for completeness. An extensive discussion of these properties can be found in Moulin (1983).

**Definition 9** Let  $\sigma : N \rightarrow N$  be a permutation. The scc  $f$  is anonymous if for every profile  $P \in \mathcal{P}^n$ , we have  $f(P) = f(P_{\sigma(1)}, P_{\sigma(2)}, \dots, P_{\sigma(n)})$ .

**Definition 10** Let  $\mu : A \rightarrow A$  be a permutation. Let  $P \in \mathcal{P}^n$ . Let  $P^\mu \in \mathcal{P}^n$  be defined as follows. For all  $a, b \in A$  and  $i \in N$ ,  $[aP_i b \Leftrightarrow \mu(a)P_i^\mu \mu(b)]$ . The scc  $f$  is neutral if for every profile  $P \in \mathcal{P}^n$ , we have  $[a \in f(P)] \Leftrightarrow [\mu(a) \in f(P^\mu)]$ .

**Proposition 2** Let the number of alternatives in  $A$  be even. Then, no anonymous and neutral scc is implementable under Assumption A2.

**Proof.** Let  $f$  be an anonymous, neutral and implementable scc. Without loss of generality, let  $A = \{a_1, \dots, a_m\}$  where  $m$  is even. Consider a preference profile  $P \in \mathcal{P}^2$  such that  $a_1 P_1 a_2 \dots P_1 a_{m-1} P_1 a_m$  and  $a_m P_j a_{m-1} \dots P_j a_2 P_j a_1$ . Suppose  $a_r \in f(P)$  for some integer  $r$  lying between 1 and  $m$ . We claim that  $a_{m-r+1} \in f(P)$ . We first note that  $a_r$  is distinct from  $a_{m-r+1}$ . Otherwise  $m = 2r - 1$  contradicting our assumption that  $m$  is even. Let  $P'$  denote the profile where individual  $i$ 's preferences are  $P_j$  and individual  $j$ 's preferences are  $P_i$ . Since  $f$  is anonymous,  $a_r \in f(P')$ . Now consider the permutation  $\mu : A \rightarrow A$  where  $\mu(a_k) = a_{m-k+1}$ ,  $k = 1, \dots, m$ . Since  $f$  is neutral,  $a_{m-r+1} \in f(P'^\mu)$ . However  $P'^\mu = P$ , so that  $a_{m-r+1} \in f(P)$ . Observe that  $L(P_i, a_r) = \{a_r, a_{r+1}, \dots, a_m\}$  while  $L(P_j, a_{m-r+1}) = \{a_1, \dots, a_{m-r+1}\}$ . Since  $m$  is even, it is easy to verify that  $L(P_i, a_r) \cap L(P_j, a_{m-r+1}) = \emptyset$  contradicting part (i) of Condition  $\beta^2$ . ■

In the many-person case, we have shown that the *absence* of veto power is sufficient for implementation. In the two-person case, No Veto Power essentially means that both individuals can get their best alternatives into the choice set. Notice that Proposition 1 suggests that a social choice correspondence may need to endow individuals with “some” Veto power in order to be implementable. For simplicity, we explore this possibility below for the case of a finite set of alternatives. Suppose now that  $A$  has cardinality  $m \geq 3$ .

Choose non-negative integers  $v_i$  and  $v_j$  such that  $v_i + v_j \leq m$ . We will say that individual  $i$  has *veto power*  $v_i$  if for all  $P \in \mathcal{P}^n$ , individual  $i$  can “veto” or eliminate the worst  $v_i$  elements in  $A$  according to  $P_i$ . Say that a scc  $f$  gives individual  $i$  positive veto power if  $v_i > 0$ .<sup>12</sup>

<sup>12</sup>See Moulin (1983) for an illuminating discussion of Voting by Veto rules. The concept of veto power can be extended to the case when  $A$  is infinite. See Abdou and Keiding (1991).

**Proposition 3** *Let  $f$  be any non-dictatorial scc which satisfies Neutrality. If  $f$  is implementable under Assumption A2, then  $f$  gives each individual positive veto power.*

**Proof.** Choose any  $f$  which is neutral. Suppose that  $v_i = 0$ . Then, there is some  $P \in \mathcal{P}^n$  such that  $aP_i x$  for all  $a \in A \setminus \{x\}$ , but  $x \in f(P)$ .

There are two possibilities. Either there exists some  $y$  such that  $yP_j x$  or  $xP_j a$  for all  $a \in A \setminus \{x\}$ .

Suppose there is  $y$  such that  $yP_j x$ . Let  $\sigma$  be a permutation such that  $\sigma(x) = y, \sigma(y) = x$  and  $\sigma(z) = z$  for all other  $z$  in  $A$ . Since  $f$  satisfies Neutrality,  $y \in f(P^\mu)$ . But,  $L(P_i, x) \cap L(P_j^\mu, y) = \emptyset$  and so Condition  $\beta^2$  cannot be satisfied.

Suppose  $xP_j a$  for all  $a \in A \setminus \{x\}$ . Since  $f$  is non-dictatorial, there is  $P' \in \mathcal{P}$  such that  $yP'_j a$  for all  $a \neq y$ , but  $z \in f(P')$  where  $z \neq y$ . If  $x \notin L(P'_j, z)$ , then again  $L(P_i, x) \cap L(P'_j, z)$  is empty and  $\beta^2$  is violated. So, assume that  $zP'_j x$ .

Now, consider the permutation  $\mu$  such that  $\mu(y) = x, \mu(x) = y$  and  $\mu(a) = a$  for all other  $a$ . Since  $x \in f(P)$ , neutrality implies that  $y \in f(P^\mu)$ . Also, note that  $L(P_i^\mu, y) = \{y\}$ . So,  $L(P_i^\mu, y) \cap L(P'_j, z)$  is empty and Condition  $\beta^2$  is violated.

This shows that  $v_i > 0$  for both  $i$  if any non-dictatorial and neutral  $f$  is to be implemented. ■

The assumption of Neutrality cannot be dispensed with in the last proposition. In the next example, we construct a non-neutral social choice correspondence which satisfies the stronger condition  $\beta^1$ , but which does not endow any individual with veto power.

**Example 1** Consider the following scc  $f$  on domain of all strict orderings. Let  $Q(P) = \{a \in A \mid \text{there does not exist } b \text{ such that } bP_i a \forall i \in N\}$ . Choose some  $x^* \in A$ . For any  $P \in \mathcal{P}^n$ ,

$$f(P) = \begin{cases} \{x^*\} & \text{if } x^* \in Q(P) \\ \{y \in A \mid yP_i x^* \forall i \in N\} \cap Q(P) & \text{otherwise} \end{cases}$$

So,  $f$  is the correspondence which chooses a distinguished alternative  $x^*$  whenever this is Pareto optimal. Otherwise, it selects those alternatives from the Pareto correspondence which Pareto dominate  $x^*$ . Notice that this is not a very “nice” social choice correspondence since it is biased in favour of  $x^*$ . However, it does satisfy Condition  $\beta^1$ .

To see this, first note that for all  $P \in \mathcal{P}^n$ , and  $x \in f(P)$ ,  $x^* \in L(P_i, x)$ . Hence,  $x^* \in L(P_i, x) \cap L(P'_j, z)$  where  $z \in f(P')$ . So, the intersection condition is satisfied for  $C(P_i, x) = L(P_i, x)$ .

Next, suppose  $z \in M(P_i^2, L(P_i, x)) \cap M(P_j^2, A)$ . If  $z = x^*$ , then clearly  $x^*$  is in  $f(P^2)$ . Otherwise, since  $x^* \in L(P_i, x)$  and  $z \in M(P_i^2, L(P_i, x))$ , we must have  $z P_i^2 x^*$ . Also, if  $z \in M(P_i^2, L(P_i, x))$ , then  $z \in Q(P^2)$ . It is also easy to check that  $z P_j^2 x^*$ . It follows that  $z \in f(P^2)$ . Part (ii) of Condition  $\beta^1$  holds trivially. Therefore  $f$  satisfies Condition  $\beta^1$ .

Notice that in this example, everyone can veto all outcomes other than  $x^*$ . However, no one has positive veto power. At the same time, the condition of No Veto Power is not satisfied since neither individual can guarantee that her best outcome is in  $f(P)$ . In other words, the *absence* of positive veto power does not mean that No Veto Power is satisfied. The absence of positive veto power and No Veto Power coincide only in the presence of neutrality.

We now demonstrate the existence of a class of well-behaved social choice correspondences which can be implemented even when there is just one partially honest individual. The possibility result stands in contrast to the negative result of Hurwicz and Schmeidler (1978) and Maskin (1999) who showed that there does not exist any two-person, Pareto efficient, non-dictatorial, implementable social choice correspondence.

Choose integers  $v_i$  and  $v_j$  such that  $v_i + v_j = m - 1$ .<sup>13</sup> Let  $V_i(P)$  denote the set of  $v_i$ -worst elements according  $P_i$ . Then, letting  $v$  denote the vector  $(v_1, v_2)$ , define

$$f^v(P) = Q(P) \setminus (V_1(P) \cup V_2(P))$$

The scc  $f^v$  corresponds to what Moulin (1983) calls the *veto core* correspondence given the veto vector  $v$  and profile  $P$ .

**Proposition 4** *The scc  $f^v$  is implementable under Assumption A1.*

**Proof.** For all  $P \in \mathcal{P}^2$  and  $a \in f^v(P)$ , let  $C(P_i, a) = L(P_i, a)$ . Observe that  $|L(P_i, a)| \geq v_i + 1$  since individual  $i$  is vetoing  $m_i$  alternatives. Also, set  $B = A$ . We will show that  $f^v$  satisfies Condition  $\beta^1$  under these specifications.

Pick an arbitrary pair  $P, P^1 \in \mathcal{P}^2$  and let  $a \in f^v(P)$  and  $b \in f^v(P^1)$ . Since  $|L(P_i, a)| \geq v_i + 1$  and  $|L(P_j^1, b)| \geq v_j + 1$ ,  $v_i + v_j = m - 1$  and  $|A| = m$ , the intersection of the two sets must be non-empty. Hence part (i)

<sup>13</sup>If  $m$  is odd, then we can choose  $v_i = v_j = (m - 1)/2$ . The scc  $f$  to be constructed would then satisfy Anonymity and Neutrality.

of Condition  $\beta^1$  is satisfied. Part (ii) of  $\beta^1$  follows from the fact that  $f^v$  is Pareto efficient.

We check for part (iii) of Condition  $\beta^1$ . Let  $P^2$  be any arbitrary profile in  $\mathcal{P}^2$ , and suppose  $c = M(P_i^2, L(P_i, a)) \cap M(P_j^2, A)$ . Then,  $c \in Q(P^2)$ . Since  $|L(P_i, a)| = v_i + 1$ , and  $c$  is  $P_i^2$ -maximal in  $L(P_i, a)$ ,  $c \notin V_i(P^2)$ . Similarly,  $c \notin V_2(P^2)$ , since  $c$  is  $P_j^2$ -maximal in  $A$ . Hence,  $c \in f^v(P^2)$ . So,  $f^v$  satisfies Condition  $\beta^1$ .

Applying Theorem 3, we conclude that  $f^v$  is implementable under Assumption A1. ■

Condition  $\beta^1$  is of course stronger than Condition  $\beta^2$ . We conclude this section by constructing an example of a single-valued selection of  $f^v$  which satisfies Condition  $\beta^2$  but not  $\beta^1$ .

Consider the following “voting by veto” social choice function  $\bar{f}^v$ . Given the vector  $v = (v_1, v_2)$  with  $v_1 + v_2 = m - 1$  and any profile  $P$ , individual 1 first vetoes alternatives in  $V_1(P)$ . Next, individual 2 vetoes the worst  $v_2$  elements in  $A - V_1(P)$  according to  $P_2$ . We denote this set as  $\bar{V}_2(P)$  and define

$$\bar{f}^v(P) = A - V_1(P) - \bar{V}_2(P)$$

Notice that at any profile  $P$ , if  $V_1(P) \cap V_2(P) = \emptyset$ , then  $f^v(P) = \bar{f}^v(P)$ . Otherwise,  $\bar{f}^v$  selects some element of  $f^v$ .

**Proposition 5** *The social choice function  $\bar{f}^v$  is implementable under Assumption A2 but not under A1.*

**Proof.** Clearly,  $\bar{f}^v$  satisfies Condition  $\beta^2$ . We show that it violates part (iii) of Condition  $\beta^1$ .

Let  $A = \{a_1, \dots, a_m\}$  and let  $P^1$  be the profile where  $a_m P_1^1 a_{m-1} \dots P_1^1 a_2 P_1^1 a_1$  and  $a_1 P_2^1 a_2 \dots P_2^1 a_{m-1} P_2^1 a_m$ . Clearly  $a_{v_1+1} = \bar{f}^v(P^1)$ . Also,  $a_1 \notin L(P_2^1, a_{v_1+1})$ . Now let  $P$  be the profile where  $a_{v_1+1} P_1 a_1 P_1 a$  and  $a_1 P_2 a_{v_1+1} P_2 a$  for all  $a \notin \{a_1, a_{v_1+1}\}$ . Note that  $a_{v_1+1} = M(P_2, L(P_2^1, a_{v_1+1}))$ . Hence,  $a_{v_1+1} = M(P_2, C(P_2^1, a_{v_1+1}))$  for any  $C(P_2^1, a_{v_1+1}) \subset L(P_2^1, a_{v_1+1})$ . Also  $a_{v_1+1} = M(P_1, A)$ . Hence part (iii) of  $\beta^1$  requires  $a_{v_1+1} = \bar{f}^v(P)$ . However  $\bar{f}^v(P) = a_1$ . Applying Theorem 3, we conclude that  $\bar{f}^v$  is not implementable under Assumption A1. ■

## 6 Separable Environments

The proof of Theorem 1 relies on the “integer game”, a construction that has been criticized in the literature (see Jackson (1992)). In such games, each player has an infinite sequence of strategies, each strategy in the sequence weakly dominating the preceding strategy. Thus, a player does not have a best-response in a weakly undominated strategy. An alternative construction is the so-called “modulo game” which is finite and avoids the difficulty described above. However mixed strategy Nash equilibria are very natural in such games and these are ruled out only by assumption. In this section we show that it is possible to sidestep both the issues of non-existence of appropriate best-responses and unaccounted mixed-strategy Nash equilibria in the case of  $n \geq 3$  agents if (i) an additional assumption on the domain of preferences is introduced and (ii) all agents are assumed to be partially honest.

**Definition 11** *The domain  $\mathcal{D}$  is separable if there exists an alternative  $w \in A$  satisfying the following property: for all  $a \in A$  and  $J \subseteq N$ , there exists  $a^J \in A$  such that  $a^J I_j w$  for all  $j \in J$  and  $a^J I_i a$  for all  $i \notin J$  for all  $R \in \mathcal{D}$ .*

Separability requires the existence of a reference alternative  $w$  with the property that for any alternative  $a \in A$  and subset of agents  $J$ , we can find an alternative  $a^J$  where agents in  $J$  are indifferent to  $w$  while agents not in  $J$  are indifferent to  $a$  in all states of the world. Note that we do not require the alternative  $w$  to be the “worst” outcome relative to outcomes in the range of the scc. The most prominent example of a separable domain is the pure exchange economy.

**Example 2** Consider a pure exchange economy with  $l$  commodities with  $l \geq 2$ . There is an aggregate endowment vector  $\Omega^l \in \mathfrak{R}_{++}^l$ . An alternative  $a$  is an allocation  $(a_1, \dots, a_n) \in \mathfrak{R}^{ln}$  such that  $a_i \geq 0$  and  $\sum_{i \in N} a_i \leq \Omega^l$ . A state of the world  $R \equiv (R_1, \dots, R_n)$  is an  $n$ -tuple of preference orderings over  $\mathfrak{R}^l$  where each  $R_i$  is strictly increasing. We claim that this is a separable domain. The alternative  $w$  is the allocation  $(0, \dots, 0)$  where all agents get zero amounts of all goods. For any other allocation  $a$  and set of agents  $J$ , the allocation  $a^J$  is the one where agents in  $J$  get zero amounts of all goods while giving those not in  $J$  exactly what they got in  $a$ , i.e  $a_i^J = 0$  if  $i \in J$  and  $a_i^J = a_i$  if  $i \notin J$ .

Related domain assumptions are made in Jackson et al. (1994) and Sjöström (1994) while proving permissible results in the context of undominated Nash implementation with bounded mechanisms. For instance, Jackson et al. (1994) also define separable environments and it is instructive to compare their notion with ours. At a formal level the definitions are not logically comparable. Their definition is stronger than ours in the following respects (i) they require the alternative  $w$  to be the worst alternative relative to outcomes in the scc and (ii) they have a condition which requires strict reversal of preferences for some pair of alternatives for every pair of states. On the other hand the nature of the “decomposition” of an alternative relative to  $w$  is much weaker in their notion of separability than in ours. In our definition,  $a^J$  cannot depend on preferences since it has to be indifferent to  $w$  for all agents in  $J$  and indifferent to  $a$  for all agents outside  $J$  for all states. This implies that our separability condition (unlike theirs) does not apply to environments with public goods with standard quasi-linear preferences.

Our informational assumption regarding the partial honesty of agents is the following.

**Assumption B:** All individuals are partially honest.

Assumption B is of course, significantly stronger than Assumption A. However, we still regard it as being reasonable since the requirement of partial honesty is relatively mild.

We now define a stronger notion of implementation.

**Definition 12** *Let  $g = (M, \pi)$  be a mechanism. A message  $m_i^* \in M_i$  is strictly dominant for agent  $i$  in state  $R \in \mathcal{D}$  if  $\pi(m_i^*, m_{-i}) \succ_i^R \pi(m_i, m_{-i})$  for all  $m_i \in M_i \setminus m_i^*$  and  $m_{-i} \in M_{-i}$ .*

Let  $S_i(g, R)$  denote the set of strictly dominant messages for agent  $i$  in mechanism  $g$ . It is clear that if  $S_i(g, R) \neq \emptyset$ , it consists of a singleton. Let  $S(g, R) = S_1(g, R) \times S_2(g, R) \times \dots \times S_n(g, R)$ . We say that  $g$  implements the scf  $f$  in strictly dominant strategies if  $f(R) = S(g, R)$  for all  $R \in \mathcal{D}$ .

Let  $\mathcal{D}$  be a separable domain and let  $f$  be a scf defined on  $\mathcal{D}$ . Consider the direct mechanism  $g_f = (M, \pi)$  where  $M_i = \mathcal{D}$  and the outcome function  $\pi$  is described below.

R.3 If  $m_i = R$  for all  $i \in N$ , then  $\pi(m) = f(R)$ .

R.4 If there exists  $j \in N$  such that  $m_i = R$ , for all  $i \neq j$  and  $m_j = R'$ , then  $\pi(m) = [f(R)]^{N-j}$  where  $[f(R)]^{N-j} \bar{I}_i w$  for all  $i \neq j$  and  $[f(R)]^{N-j} \bar{I}_j f(R)$  for all  $\bar{R} \in \mathcal{D}$ .

R.5 In all other cases  $\pi(m) = w$ .

The mechanism  $g_f$  can be informally described as follows. If all agents unanimously announce a state  $R$ , then the outcome is  $f(R)$ . Unilateral deviations from common state announcements (say,  $R$ ) are dealt with by picking an alternative which makes the deviant agent indifferent between the alternative and the alternative which would have obtained if the deviant agent had announced the same state as the others, while making the other agents indifferent to the alternative  $w$ . In all other cases, the outcome is  $w$ .

**Theorem 4**<sup>14</sup> *Let  $n \geq 3$  and suppose Assumption B holds. Let  $f$  be a scf defined over a separable domain. Then  $g_f$  implements  $f$  in strictly dominant strategies.*

**Proof.** Suppose that the true state of the world is  $R \in \mathcal{D}$ . We will show that  $R = S_i(g_f, R)$  for all  $i \in N$ . Consider the following different possibilities for messages  $m_{-i} = (R^j)_{j \neq i}$  of agents other than  $i$ .

**Case 1:**  $R^j = R$  for all  $j \neq i$ , i.e. all agents other than  $i$  are announcing the true state. If  $i$  announces  $R^i \neq R$ , then the outcome is indifferent to  $f(R)$  under  $R_i$ . Announcing  $R^i = R$  yields  $f(R)$ . However, since  $i$  is partially honest, she is strictly better-off announcing  $R$ .

**Case 2:**  $R^j = R'$  for all  $i \neq j$  and  $R \neq R'$ . Note that for all announcements  $R^i$  of agent  $i$ , the outcome is  $[f(R')]^{N-i}$  where  $[f(R')]^{N-i} I_i [f(R')]$ . Since  $i$  is partially honest, announcing  $R^i = R$  strictly dominates any other announcement.

**Case 3:** There exists  $k \neq i$  such that  $R^j = R'$  for all  $j \neq k$  and  $R^k = R'' \neq R'$ . An announcement  $R^i$  by agent  $i$  leads to one of following alternatives: (i)

---

<sup>14</sup>In an earlier version of the paper, we showed that the mechanism  $g_f$  Nash-implements  $f$ . We are particularly indebted to a referee who drew our attention to the fact that  $g_f$  implements  $f$  in strictly dominant strategies. The theorem remains true for a slightly weaker definition of separability. However, we have not been able to find an interesting example of a domain which satisfies the weaker definition but not the stronger definition of separability.

$[f(R^l)]^{N-k}$  (ii)  $[f(R'')]^{N-j}$  (if  $n = 3$ ) or (iii)  $w$ . Now,  $[f(R'')]^{N-j}I_iw$  and  $[f(R^l)]^{N-k}I_iw$ . Since  $i$  is partially honest,  $R^i = R$  is a strictly dominant strategy.

**Case 4:** There exist distinct agents  $j, k, l \neq i$  such that  $R^j, R^k$  and  $R^l$  are all distinct. In this case all announcements  $R^i$  of agent  $i$  yield  $w$ . Since  $i$  is partially honest,  $R^i = R$  is a strictly dominant strategy.

Cases 1-4 exhaust all possibilities completing the proof. ■

We make several remarks below regarding the interpretation and implications of Theorem 4.

**Remark 1** The mechanism  $g_f$  is extremely simple and has no integer/modulo game features. Moreover each agent has a strategy (truthtelling) which is strictly dominant. Consequently, issues regarding mixed-strategy Nash equilibria also do not arise.

**Remark 2** Our permissive result has no bearing on the negative results obtained in exchange economies for strategy-proof (i.e. truthtelling is a weakly dominant strategy) social choice functions (for example Barberà and Jackson (1995) and Serizawa and Weymark (2003)). This is because the results pertain to completely different models. The models where strategy-proofness is investigated are incomplete information models where agents only know their own preference orderings; the negative results arise due to the strong implications of the incentive-compatibility constraints. In our model, on the other hand, agents know each other's preferences and report them to the planner. It is well-known that there are no incentive-compatibility requirements in our model provided that there are at least three agents.

**Remark 3** If the partial honesty requirement is dropped, truth-telling is still a weakly dominant strategy for each agent in the mechanism  $g_f$ . However, implementation fails because of the existence of other Nash equilibria. For instance, suppose that there are three agents all of whom announce distinct states of the world. The outcome is then  $w$ . Observe that these strategies constitute a Nash equilibrium for every state of the world. Hence  $g_f$  implements a social choice correspondence whose value is every state of the world includes  $w$ .

**Remark 4** The classic paper of Hurwicz (1972) has inspired a large literature that investigates social choice correspondences which can be implemented in exchange economies with complete information. A focus of this

literature has been the design of “natural” and “simple” mechanisms that implement the Walrasian correspondence.<sup>15</sup> Recently, Giraud and Stahn (2010) discuss mechanisms that can implement the Walrasian correspondence that satisfy the additional requirements of *feasibility* (i.e. outcomes following all possible message-tuples are feasible) and *boundedness* in the sense of Jackson (1992). Example 2 shows that the classical exchange economy constitutes a separable domain. In this context, the mechanism  $g_f$  is feasible; moreover since there exists a strictly dominant strategy at every state of the world,  $g_f$  is also bounded. Our result therefore implies that if agents are partially honest, *every* social choice function (in particular, arbitrary selections of the Walrasian correspondence) can be implemented by a “simple” mechanism that is feasible and bounded.

It would be interesting to find a counterpart to Theorem 4 in the case of two agents. A natural question is whether a strengthened version of separability where the alternative  $w$  is the worst outcome in all states, is sufficient for implementation by a mechanism similar to  $g_f$ . However, we conjecture that the answer to this question must be negative for the following reason. As the preceding proof makes clear, Separability “helps” in the multi-agent case because it allows for selective punishment. For instance, if all agents are lying unanimously and getting the outcome  $f(R')$  when the true state is  $R$ , then some agent can deviate to truth-telling without (essentially) changing the outcome as far as she is concerned. So, the deviation is profitable for her since she benefits from telling the truth. But, in the two-person case, the planner has to punish *both* players when the announcement is of the form  $(R, R')$  - this follows from necessity of the Intersection condition.

## 7 A Model with Incomplete Information on Honesty

In this section we depart from the informational assumptions made in the previous sections. In particular, we no longer assume that there exists some partially honest individual with probability one. Instead we assume that there exists a particular agent, say  $j$  who is partially honest with probability

---

<sup>15</sup>See for instance, Schmeidler (1980), Postlewaite (1985), Postlewaite and Wettstein (1989), Dutta et al (1995), Giraud (2003).

$\epsilon > 0$  and self-interested with probability  $1 - \epsilon$ . All other players are self-interested. A mechanism  $g$  as specified in Section 2 is a pair  $(M, \pi)$  where  $M$  is a product message set and  $\pi : M \rightarrow A$ . As before we shall assume without loss of generality that  $M_i = \mathcal{D} \times S_i$  where  $S_i$  denotes other components of agent  $i$ 's message space.

Let  $R \in \mathcal{D}$  be a state of the world and let  $g$  be a mechanism. A game of incomplete information is induced as follows. Individual  $j$  has two types, truthful and self-interested denoted by  $t$  and  $s$  respectively. All individuals other than  $j$  have a single type  $s$ . The action set for individual  $i$  is  $M_i$ . For individual  $j$  of type  $t$ , preferences over outcomes are induced by the order  $\succeq^R$  as in Definition 4. For all individuals of type  $s$ , preferences over lotteries with outcomes in  $A$  must be considered. Let  $i$  be an arbitrary individual of type  $s$ . Let  $v$  be a utility function which represents  $R_i$ , i.e  $v$  is a mapping  $v : A \rightarrow \mathfrak{R}$  which satisfies the requirement that for all  $x, y \in A$ ,  $[xPy \Leftrightarrow v(x) > v(y)]$  and  $[xIy \Leftrightarrow v(x) = v(y)]$ . Let  $p = \{p_x\}$ ,  $x \in A$  be a lottery over elements of  $A$ , i.e.  $p(x) \geq 0$  and  $\sum_{x \in A} p_x = 1$ . We say that lottery  $p$  is at least as good as lottery  $p'$  according to cardinalization  $v$ , denoted by  $pR_i^v p'$ , if  $\sum_{x \in A} v(x)p_x \geq \sum_{x \in A} v(x)p'_x$ .

Fix a mechanism  $g$ ,  $R \in \mathcal{D}$  and let  $v_i$  be a cardinalization  $R_i$  for all  $i \in N$ . We now have a game of incomplete information. A strategy for individual  $j$  is a pair of messages  $m_j^t, m_j^s \in M_j$ . A strategy for an individual  $i \neq j$  is a message  $m_i \in M_i$ . A strategy profile  $((\bar{m}_j^t, \bar{m}_j^s), \bar{m}_{-j})$  is a Bayes-Nash equilibrium (BNE) if

1.  $g(\bar{m}_j^t, \bar{m}_{-j}) \succeq_j^R g(m_j, \bar{m}_{-j})$  for all  $m_j \in M_j$ .
2.  $g(\bar{m}_j^s, \bar{m}_{-j}) R_j g(m_j, \bar{m}_{-j})$  for all  $m_j \in M_j$ .
3.  $v_i(g(\bar{m}_j^t, \bar{m}_i, \bar{m}_{-i,j}))\epsilon + v_i(g(\bar{m}_j^s, \bar{m}_i, \bar{m}_{-i,j}))(1-\epsilon) \geq v_i(g(\bar{m}_j^t, m_i, \bar{m}_{-i,j}))\epsilon + v_i(g(\bar{m}_j^s, m_i, \bar{m}_{-i,j}))(1-\epsilon)$  for all  $m_i \in M_i$  and all  $i \neq j$ .

In other words, no individual whether truthful or self-interested has a unilateral incentive to deviate. It is of course, evident that whether or not a strategy profile is an equilibrium will depend on the cardinalizations chosen. We will however consider a restricted class of mechanisms, called *ordinal mechanisms* (see Abreu and Sen(1991) for a further discussion of this notion) where the set of equilibria do not depend on the cardinalizations chosen. Thus, Part 3 of the condition above holds for *all* cardinalizations  $v_i$  of  $R_i$ .

Let  $g$  be an ordinal mechanism. Note that by assumption, the pair  $(g, R)$  defines a game of incomplete information for any  $R \in \mathcal{D}$ . We say that  $g$  implements a scc  $f$  if, for all  $R \in \mathcal{D}$ , the following conditions hold:

1. For all  $a \in f(R)$ , there exists a BNE of the game  $(g, R)$  denoted by  $((\bar{m}_j^t, \bar{m}_j^s), \bar{m}_{-j})$  such that  $\pi(\bar{m}_j^t, \bar{m}_{-j}) = \pi(\bar{m}_j^s, \bar{m}_{-j}) = a$ .
2. Let  $((\bar{m}_j^t, \bar{m}_j^s), \bar{m}_{-j})$  be an arbitrary BNE of  $(g, R)$ . Then  $\pi(\bar{m}_j^t, \bar{m}_{-j}), \pi(\bar{m}_j^s, \bar{m}_{-j}) \in f(R)$ .

A mechanism implements  $f$  if for every state  $R$  and every  $a \in f(R)$ , there exists a Bayes-Nash equilibria whose outcome is  $a$  irrespective of the realization of individual  $j$ 's type. In addition, all Bayes-Nash equilibria in the game  $(g, R)$  are  $f$ -optimal for all realizations of  $j$ 's type.

We are able to show the following permissive result.

**Theorem 5** *Assume  $n \geq 3$ . Then, every scc satisfying No Veto Power is implementable.*

**Proof.** We use the same mechanism used in the proof of Theorem 1 and show that it implements  $f$  where  $f$  is an arbitrary scc satisfying No Veto Power.

Let the “true” state of the world be  $R$  and let  $a \in f(R)$ . Consider the strategy profile  $((\bar{m}_j^t, \bar{m}_j^s), \bar{m}_{-j})$  where  $\bar{m}_j^t = \bar{m}_j^s = \bar{m}_i = (R, a, k_i)$  for all  $i \neq j$ . Observe that no individual of any type can deviate and change the outcome. Hence the strategy profile is a BNE and the outcomes under  $(\bar{m}_j^t, \bar{m}_{-j})$  and  $(\bar{m}_j^s, \bar{m}_{-j})$  are both  $a$ .

Now let  $((\bar{m}_j^t, \bar{m}_j^s), \bar{m}_{-j})$  be an arbitrary BNE of  $(g, R)$ . We will show that the outcomes generated by the message profiles  $(\bar{m}_j^t, \bar{m}_{-j})$  and  $(\bar{m}_j^s, \bar{m}_{-j})$  belong to  $f(R)$ . We consider several mutually exhaustive cases.

**Case 1.** At most  $n - 1$  messages in the  $n$ -tuples  $(\bar{m}_j^t, \bar{m}_{-j})$  and  $(\bar{m}_j^s, \bar{m}_{-j})$  are of the form  $(R', a, -)$  for some  $a \in f(R')$ . Then, either all  $n - 1$  individuals,  $j \neq i$  or  $n - 2$  individuals from the set  $N - \{j\}$  and either individual  $j$  of type  $t$  or type  $s$  can deviate, precipitate and win the integer game. Thus the outcomes generated by the  $n$ -tuples  $(\bar{m}_j^t, \bar{m}_{-j})$  and  $(\bar{m}_j^s, \bar{m}_{-j})$  must be  $R$ -maximal for at least  $n - 1$  agents. By No Veto Power, the outcomes under  $(\bar{m}_j^t, \bar{m}_{-j})$  and  $(\bar{m}_j^s, \bar{m}_{-j})$  must belong to  $f(R)$ .

**Case 2.**  $\bar{m}_j^t = \bar{m}_i = (R', a, -)$  where  $a \in f(R')$  for all  $i \neq j$ . According to the mechanism, the outcome is  $a$  when individual  $j$  is of type  $t$ . We will show

that  $a \in f(R)$ . If  $R' = R$ , there is nothing to prove. Assume therefore that  $R \neq R'$ . Then individual  $j$  of type  $t$  can deviate to a message  $m_j^t(R) \in T_j(R)$  still obtain the outcome  $a$  but be strictly better off. Hence an equilibrium of the type specified cannot exist.

**Case 3.**  $\bar{m}_j^s = \bar{m}_i = (R', a, -)$  where  $a \in f(R')$  for all  $i \neq j$ . As in Case 2, the only case which needs to be dealt with is the one where  $R' \neq R$ . Note that under the specification of the mechanism, the outcome must be  $a$  when individual  $j$  is of type  $t$ . Since this individual can always deviate to  $m_j^t \in T_j(R)$  without changing the outcome from  $a$ , it must be the case that  $\bar{m}_j^t = (R, b, -)$  where  $b \in f(R)$ . Now consider a deviation by individual  $i \neq j$  to  $(\hat{R}, c, k_i)$  where (i)  $(\hat{R}, d) \neq (R', a)$ , (ii)  $k_i$  is strictly greater than the integer components of the messages  $\bar{m}_j^t, \bar{m}_j^s$  and  $\bar{m}_l$  for all  $l \neq i, j$ , and  $c$  is  $R_i$ -maximal. If individual  $j$  is of type  $t$ , the integer game is triggered and won by  $i$  to get say outcome  $c$  which by specification is an  $R_i$ -maximal alternative. If  $j$  is of type  $s$ , the outcome is  $a$ . Hence the outcome is a lottery where  $c$  is obtained with probability  $\epsilon$  and  $a$  with probability  $1 - \epsilon$ . Observe that under the strategy-tuple  $((\bar{m}_j^t, \bar{m}_j^s), \bar{m}_{-j})$ , the outcome is  $a$  irrespective of  $j$ 's type, i.e the outcome is  $a$  with probability 1. Since  $((\bar{m}_j^t, \bar{m}_j^s), \bar{m}_{-j})$  is a BNE, it must be true that the deviation is not profitable, i.e  $a$  is  $R_i$ -maximal.<sup>16</sup> Since  $i$  was chosen arbitrarily from the set  $N - \{j\}$ , No Veto Power implies that  $a \in f(R)$ . ■

A fundamental aspect of Theorem 5 is that its validity does not depend on the value of  $\epsilon$ . In other words, if it is known that there exists an individual  $j$  who has a “tiny” probability of being partially honest, the possibilities for implementation are dramatically increased in the case where there are at least three individuals. A similar analysis can also be extended to the case of two individuals. We do not however pursue these matters further in this paper.

## 8 Conclusion

This paper has investigated the consequence of assuming that players in the Nash implementation problem are “minimally” honest. Our conclusion is that this dramatically increases the scope for implementation. In the case where there are at least three individuals, all social choice correspondences

---

<sup>16</sup>Note that this conclusion holds for all cardinalizations  $v_i$  of  $R_i$ .

satisfying the weak No Veto Power condition can be implemented. If it is further assumed that all individuals are partially honest and the domain satisfies certain restrictions, then all social choice correspondences are implementable by mechanisms which do not require the use of “integer games”. In the two-person case, the results are more subtle but are nevertheless similar in spirit. We also show that the many-person result (Theorem 1) extends to the case where there exists a single individual with an arbitrarily small probability of being partially honest.

We believe that the notion that players are not driven purely by strategic concerns based on their preferences over outcomes, is a natural one. This has an important bearing on mechanism design theory. However, the exact nature of the departure from standard preferences can be modelled in multiple ways. It is a fruitful area for future research.

## References

- [1] Abreu, D. and A. Sen (1991), “Virtual Implementation in Nash Equilibrium”, *Econometrica*, 59, 997-1021.
- [2] Abdou, J. and H. Keiding (1991), *Effectivity Functions in Social Choice*, Kluwer Academic Publishers.
- [3] Barberà, S and M. Jackson (1995), “Strategy-Proof Exchange”, *Econometrica*, 63, 51-87.
- [4] Busetto, F. and G. Codognato (2009), “Reconsidering Two-agent Nash Implementation”, *Social Choice and Welfare*, 22, 171-179.
- [5] Corchón and C. Herrero (2004), “A Decent Proposal”, *Spanish Economic Review*, 6(2), 107-125.
- [6] Dutta, B. and A. Sen (1991), “A Necessary and Sufficient Condition for Two-Person Nash Implementation”, *Review of Economic Studies*, 58, 121-128.
- [7] Dutta, B., A.Sen and R.Vohra (1995), “Nash Implementation through Elementary Mechanisms in Economic Environments”, *Economic Design*, 1, 173-203.

- [8] Fehr, E. and K.Schmidt (1999), “A Theory of Fairness, Competition and Cooperation”, *Quarterly Journal of Economics*, 114, 817-868.
- [9] Fehr, E., U. Fischbacher and S. Gächter (2002), “Strong Reciprocity, Human Cooperation and the enforcement of Social Norms”, *Human Nature*, 13, 1-25.
- [10] Giraud, G. (2003), “Strategic Market Games: An Introduction”, *Journal of Mathematical Economics*, 355-375.
- [11] Giraud, G and H. Stahn (2010), “Nash Implementation of Competitive Equilibria via a Bounded Mechanism”, *mimeo*.
- [12] Hurwicz, L.(1972) “The Design of Mechanisms for Resource Allocation”, in Radner, R. and McGuire, C. (eds) *Decision and Organization*, Amsterdam, North Holland.
- [13] Hurwicz, L. and D. Schmeidler (1978), “Outcome Functions that Guarantee the Existence and Pareto Optimality of Nash Equilibria”, *Econometrica*, 46, 144-174.
- [14] Jackson, M.O. (1992), “Implementation in Undominated Strategies: A Look at Bounded Mechanisms”, *Review of Economic Studies*, 59, 757-775.
- [15] Jackson, M.O., T. Palfrey and S. Srivastava (1994), “Undominated Nash Implementation in Bounded Mechanisms”, *Games and Economic Behavior*, 6, 474-501.
- [16] Jackson, M.O. (2001), “A Crash Course in Implementation Theory”, *Social Choice and Welfare*, 18, 655-708.
- [17] Kahneman, D., J.L. Knetsch and R. Thaler (1986), “Fairness as a Constraint on Profit-seeking: Entitlements in the Market”, *American Economic Review*, 76,728-741.
- [18] Kartik, N and O. Tercieux (2010), “Implementation with Evidence: Complete Information”, *mimeo*.
- [19] Lombardi, M and N. Yoshinara (2011) “Partially Honest Nash Implementation: Characterization Results”, *mimeo*.

- [20] Maskin, E. (1999), “Nash Equilibrium and Welfare Optimality”, *Review of Economic Studies*, 66, 23-38.
- [21] Maskin, E. and T. Sjöström (2002), “Implementation Theory”, in *Handbook of Social Choice Theory, Vol.1*, edited by K.J. Arrow, A.K. Sen and K.Suzumura, North Holland.
- [22] Matsushima, H. (2008a), “Role of Honesty in Full Implementation”, *Journal of Economic Theory*, 139, 353-359.
- [23] Matsushima, H. (2008b), “Behavioral Aspects of Implementation Theory”, *Economic Letters*, 100, 161-164.
- [24] Moore, J. and R. Repullo (1990), “Nash Implementation: A Full Characterization”, *Econometrica*, 58, 1083-1099.
- [25] Moulin, H. (1983), *The Strategy of Social Choice*, Advanced Textbooks in Economics, Editors: C.J. Bliss and M.D. Intrilligator, North-Holland, Amsterdam.
- [26] Postlewaite, A. (1985), “Implementation via Nash Equilibria in Economic Environments”, in Hurwicz, L., Schmeidler, D. and H.Sonnenschein (eds.) *Social Goals and Social Organisation*, Cambridge: Cambridge University Press.
- [27] Postlewaite, A. and D.Wettstein (1989), “Feasible and Continuous Implementation”, *Review of Economic Studies*, 56, 603-611.
- [28] Rabin, M. (1993), “Incorporating Fairness into Game Theory and Economics”, *American Economic Review*, 83, 1281-1320.
- [29] Roth, A. (1995), “Bargaining Experiments” in J.Kagel and A.Roth (edited) *Handbook of Experimental Economics*, Princeton University Press.
- [30] Saijo, T. (1987), “On Constant Maskin Monotonic Social Choice Functions”, *Journal of Economic Theory*, 42, 382-386.
- [31] Schmeidler, D. (1980), “Walrasian Analysis via Strategic Outcome Functions”, *Econometrica*, 48, 1585-1593.

- [32] Serizawa, S. and J. Weymark (2003), “Efficient Strategy-Proof Exchange and Minimal Consumption Guarantees”, *Journal of Economic Theory*, 109, 246-263.
- [33] Sjöström, T. (1994), “Implementation in Undominated Equilibria without Integer Games”, *Games and Economic Behavior*, 6, 502-511.