Nash Implementation with Partially Honest Individuals

Bhaskar Dutta and Arunava Sen*

August 22, 2009

PRELIMINARY AND INCOMPLETE. PLEASE DO NOT QUOTE

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 (according to preferences in the true state) than that which obtains when lying. We show that when there are at least three individuals, the presence of even a single partially honest individual (whose identity is not known to the planner) can lead to a dramatic increase in the class of Nash implementable social choice correspondences. In particular, all social choice correspondences satisfying No Veto Power can be implemented. We also provide necessary and sufficient conditions for implementation in the two-person case when there is exactly one partially honest individual and when both individuals are partially honest. We provide examples which illustrate the possibilities for implementation in these cases.

*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 are fully strategic in their behavior; moreover their decision regarding the messages that they send to the principal is based entirely on their preferences over the *outcomes* that they believe will obtain as a result of the messages they send.

In this paper we investigate the consequences of assuming that some agents have preferences not just on the outcomes but also directly on the *messages* themselves. In particular we assume that there are some agents who have a "small" intrinsic preference for honesty. Specifically, we assume the following: there are some agents who when asked to report the state of the world, 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) according to her preferences in the true state. 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 suppose that the message profiles $(m_i, m_{_i})$ and (m'_i, m_{-i}) leads to the same outcome in the mechanism, say a. Then this agent will *strictly* prefer to report m_i rather than m'_i while 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.[1] 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. We consider such behaviour quite plausible at least for some agents.

We investigate the theory of Nash implementation pioneered by Maskin [5][2] 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

[1]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.

[2]See Jackson [4] for a comprehensive survey of the literature.

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 also investigate the important case of two agents. 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 of strict order preferences over outcomes. We show that though non-trivial restrictions remain, the possibilities for implementation increase substantially at least in the case where both individuals are partially honest. In particular, the negative result of Hurwicz and Schmeidler [3] and Maskin [5]no longer applies.

Some recent papers very similar in spirit to ours, though different in substance are Matsushima [6] and Matsushima[7]. We discuss his work in greater detail in Section 4.

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 while Section 6 concludes.

2 The Background

Consider an environment with a *finite* set $N = \{1, 2, ..., n\}$ of *agents* or *individuals* and a set X of feasible *outcomes*. Each individual i has a preference ordering R_i over X. A preference profile $(R_1, \ldots, R_n)$ specifies a preference ordering for each $i \in N$. Letting $\mathcal{R}$ be the set of all orderings over X, $\mathcal{R}^n$ will denote the set of all preference profiles. Let $\mathcal{P}$ denote the set of *strict* orderings over X, and $\mathcal{P}^n$ the set of profiles consisting of strict preference orderings. A *domain* is a set $\mathcal{D} \subset \mathcal{R}^n$. An *admissible* preference profile will be represented by $R, R' \in \mathcal{D}$. In case these admissible profiles involve only strict orders, they will be denoted by $\mathcal{P}, \mathcal{P}'$ etc. We will also refer to an admissible preference profile as a *state of the world*.

We assume that each agent observes R, 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 $f(R) \subseteq X$ for each profile $R \in \mathcal{D}$. 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 (S, π), where S is the product of individual strategy sets S_i and π is the outcome function mapping each vector of individual strategies into an outcome in X.*

A mechanism g together with any state of the world R induces a game with player set N, strategy sets S_i for each player i, and payoffs given by the composition of the outcome function π and the preference orderings R_i. Let $N(g, R)$ represent 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 $B \subseteq X$ and preference R_i, $M(R_i, B) = \{a \in B | aR_ib \; \forall b \in B\}$, is the set of *maximal elements* in B according to R_i.

The *lower contour set* of $a \in X$ for individual i and preference ordering R_i is $L(R_i, a) = \{b \in X | aR_ib\}$.

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 [5] that there is no loss of generality in restricting ourselves to mechanisms of this kind. Therefore, consider a mechanism g in which for each $i \in N$, $M_i = \mathcal{R}^n \times A_i$, where A_i denotes the other components of the message space. For each i and $R \in \mathcal{D}$, let $T_i(R) = \{R\} \times A_i$. 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 X 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.

Definition 4 *Let $g = (M, \pi)$ be a mechanism where $M_i = \mathcal{D} \times A_i$. 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.[3] We omit formal definitions.

4 Many Person Implementation

The seminal paper of Maskin [5] 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 the dictatorial single-valued social choice correspondence can satisfy this condition if there is no restriction on the domain of preferences. 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*

[3] We denote this set as $N(g, \succeq^R)$.

$a \in X$, if $a \in f(R) \setminus f(R')$, then there is $i \in N$ and $b \in X$ such that aR_ib and bP'_ia.

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

In this section we make the following assumption.

Assumption A: There exists at least one partially honest individual. This fact is known to the planner. However, the identity of this individual is not known to her.

We show here that the presence of even one partially honest individual - even when the identity of the individual is not known - results in a dramatically different result. In particular, we show that Monotonicity is no longer required, 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 $i \in N$, $M_i = \mathcal{D} \times X \times \{1, \ldots, n\}$. Hence, for each agent i, a typical message or strategy consists of a state of world R, an outcome a, and an integer in $\{1, \ldots, n\}$. 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 $a \in f(R)$, then the outcome is a.

(R.2) : In all other cases, the outcome is the one announced by i^*, where $i^* \equiv \sum_{j \in N} k_j \bmod n$ where k_j is the integer announced by j.

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 $k_i \in \{1, \ldots, n\}$. 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) and the same $a \in f(R')$. Let the outcome be some $b \in X$. Then, any one of $(n-1)$ individuals can deviate, precipitate the modulo game, *and* be the winner of the modulo 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' and $b \in f(R')$, 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 $m_i(R) \in T_i(R)$. The outcome will still remain b, but i gains from telling the truth. ■

Remark 1 Matsushima [7] also 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. 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 [6] focuses on the incomplete information framework and proves a similar permissive result for Bayesian implementation when players suffer a small utility loss from lying.

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 [2] and Moore and Repullo [8] (see also Busetto and Codognato [1]). 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 and Matsushima's result show that when there are at least three individuals, the presence of a partially honest player even though her identity is not known to the planner 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 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. $\mathcal{D} \subset \mathcal{P}^n$. Later we shall discuss some of the complications which arise when indifference in individual orderings is permitted.

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 implementable. 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 [3] and Maskin [5] 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 and the planner knows this fact.

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 β^2 if there exists a set B which contains the range of f, and for each $i \in N$, $R \in \mathcal{D}$ and $a \in f(R)$, there exists a set $C(R_i, a) \subseteq B$ with $a \in C(R_i, a) \subseteq L(R_i, a)$ such that*

(i) $C(R_i, a) \cap C(R_j^1, b) \neq \emptyset$ *for all* $R^1 \in \mathcal{D}$ *and* $b \in f(R^1)$.

(ii) $[a \in M(R_i, B) \cap M(R_j, B)] \Rightarrow [a \in f(R)]$.

Condition β^2 comprises two parts. The first is an *intersection property* which requires appropriate lower contour sets to have a non-empty interesction. 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 β^2.*

Proof. We first show if a scc f is implementable, it satisfies Condition β^2.[4]

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

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

[4]This argument is quite close to that in Dutta and Sen [2] and Moore and Repullo [8].

Now fix a state $R \in \mathcal{D}$ and let $a \in A$ be such that $a \in M(R_i, B) \cap M(R_j, B)$. Since $a \in B$, there exists a message profile m such that $\pi(m) = a$. If m is a Nash equilibrium in state R, then $a \in f(R)$. If m is not a Nash equilibrium, then there is an individual, say i and $\hat{m}_i \in T_i(R)$ such that $\pi(\hat{m}_i, m_j) R_i a$. However, $a \in M(R_i, B)$ and $\pi(\hat{m}_i, m_j) \in B$ implies that $\pi(\hat{m}_i, m_j) R_i a$. Since R_i is a strict order, we must have $\pi(\hat{m}_i, m_j) = a$. If $(\hat{m}_i, m_j)$ is a Nash equilibrium, then again $a \in f(R)$. Otherwise, $j \neq i$ deviates to some $\hat{m}_j \in T_j(R)$ 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 $a \in f(R)$. Hence f satisfies (ii) of Condition β^2.

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

For each i, let $M_i = \mathcal{D} \times X \times X \times \{T, F\} \times \{1, 2\}$. The outcome function π is defined as follows.

(i) If $m_i = (R, a, b, T, k^i)$ and $m_j = (R, a, c, T, k^j)$ where $a \in f(R)$, then $\pi(m) = a$.

(ii) If $m_i = (R, a, c, T, k^i)$ and $m_j = (R^1, b, d, T, k^j)$ where $a \in f(R)$ and $b \in f(R^1)$ with $(a, R) \neq (b, R^1)$, then $\pi(m) = x$ where $x \in C(R^1_i, b) \cap C(R_j, a)$.

(iii) If $m_i = (R, a, c, F, k^i)$ and $m_j = (R^1, b, d, T, k^j)$, with $a \in f(R)$ and $b \in f(R^1)$, then $\pi(m) = c$ if $c \in C(R^1_i, b)$ 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 modulo game.

The mechanism is similar, but not identical to that used by Dutta and Sen [2]. Essentially, when both individuals announce the same (R, a) where $a \in f(R)$ *and* T, then the mechanism identifies this as the "equilibrium" message for (R, 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 (R, 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 $C(R_i, a)$. Finally, in all other cases, the modulo game is employed.

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

Consider any $a \in f(R)$. Let $m_i^* = (R, a, ., T, k^i)$ 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 $L(R_i, a)$, and so m^* must be a Nash equilibrium.

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

Case 1: Suppose $m_i = (R^1, a, ., T, k^i)$ for all i, where $a \in f(R^1)$.[5] Then, $\pi(m) = a$. If $R = R^1$, there is nothing to prove, since $a \in f(R)$. Assume therefore that $R \neq R^1$. Let i deviate to $m_i' = (R, b, a, F, k^i)$, where $b \in f(R)$. Then $\pi(m_i^{'}, m_j) = a$. But, this remains a profitable deviation for i since $m_i' \in T_i(R)$. Hence m is not a Nash equilibrium.

Case 2: Suppose $m_i = (R^1, a, c, T, k^i)$ and $m_j = (R^2, b, d, T, k^j)$ where $a \in f(R^1)$ and $b \in f(R^2)$ with $(a, R) \neq (b, R^1)$. Then $\pi(m) = x$ where $x \in C(R_i^2, b) \cap C(R_j^1, a)$. Suppose that $R = R^1 = R^2$ does not hold. So, either $R^1 \neq R$ or $R^2 \neq R$. Suppose w.l.o.g that $R^1 \neq R$. Then, i can deviate to $m_i' \in T_i(R)$ such that $m_i' = (R, b, x, F, k^i)$. Since $x \in C(R^2, b)$ by assumption, $\pi(m_i^{'}, m_j) = x$. However, $m_i^{'} \in T_i(R)$ so that i gains by deviating. Hence m is not a Nash Equilibrium. Suppose instead that $R = R^1 = R^2$ holds and m is a Nash equilibrium. Then it must be the case that $x \in M(R_i, C(R_i, b))$ (Recall that $R = R^2$), i.e. xR_ib. However bR_ix since $x \in C(R_i, b)$ by assumption. Since R_i is a strict order, we have $b = x$. Since $b \in f(R)$, it follows that $\pi(m) \in f(R)$.

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

[5]That is, both individuals "coordinate" on a lie about the state of the world.

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

Case 4: The remaining possibility is that m is such that the modulo game decides the outcome. In this case, if m is a Nash equilibrium, then $\pi(m) \in M(R_i, B) \cap M(R_j, B)$. From (ii) of Condition β^2, we have $\pi(m) \in f(R)$. ■

As we have seen above, assuming that all admissible preferences are strict orders leads to a simple characterization of implementable sccs. Matters are more complicated and subtle when we allow for indifference. For instance, even Unanimity is no longer necessary. Let R be a state of the world where the maximal element 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. The complications associated with weak orders extend beyond that arising from Unanimity. However, we can establish weaker necessary conditions and prove a version of Theorem 2 with a "gap" between the necessary and sufficient conditions. This requires considerable investment in notation, which we do not consider worthwhile at this point since the main points of this exercise are brought out by Theorem 2 in its present form.

5.2 Exactly One Partially Honest Individual

Here we make the following informational assumption.

Assumption A1: There is exactly one partially honest individual. The planner knows this fact but does not know the identity of the 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 β^1 if there is a set B which* *contains the range of f, and for each $i \in N$, $R \in \mathcal{D}$ and $a \in f(R)$, there exists a set $C(R_i, a) \subseteq B$ with $a \in C(R_i, a) \subseteq L(R_i, a)$ such that*

(i) $C(R_i, a) \cap C(R^1_j, b) \neq \emptyset$ *for all* $R^1 \in \mathcal{D}$ *and for all* $b \in f(R^1)$.

(ii) for all $R^2 \in \mathcal{D}$, *if* $b \in C(R_i, a)$ *and* $b \in M(R^2_i, C(R_i, a)) \cap M(R^2_j, B)$, *then* $b \in f(R^2)$.

(iii) for all $R^2 \in \mathcal{D}$, $[a \in M(R^2_i, B) \cap M(R^2_j, B)] \Rightarrow [a \in f(R^2)]$.

The only difference between Conditions β^1 and β^2 is the extra requirement (ii) in the former. Our next result shows that Condition β^1 is the exact counterpart of Condition β^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 β^1.*

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

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

The proof of part (iii) of Condition β^1 is similar though not identical to the proof of its counterpart in β^2. Let $R^2 \in \mathcal{D}$ and consider a such that $a \in M(R^2_i, B) \cap M(R^2_j, B)$. 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 $\hat{m}_i \in T_i(R^2)$ such that $\pi(\hat{m}_i, m_j) R^2_i a$. However, $a \in M(R^2_i, B)$

and $\pi(\hat{m}_i, m_j) \in B$ implies that $\pi(\hat{m}_i, m_j) R_i^2 a$. Since R_i^2 is a strict order, we must have $\pi(\hat{m}_i, m_j) = a$. Since $a \in M(R_j^2, B)$, it must be the case that $(\hat{m}_i, m_j)$ is a Nash equilibrium and $a \in f(R^2)$.

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

Let R be the true state of the world. The proof that every $a \in f(R)$ 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 $f(R)$. 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 truthtelling strategy without changing the outcome.

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

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

Case 3: Suppose $m_i = (R^1, a, c, F, k^i)$ and $m_j = (R^2, b, d, T, k^j)$ where $a \in f(R^1)$ and $b \in f(R^2)$. Let $\pi(m) = x$. We know that $x \in C(R_i^2, b)$. Suppose $R \neq R^1$ and $R \neq R^2$ 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 $R = R^1$, i.e i is the partially honest individual. Note that individual j can trigger the modulo game and obtain any alternative in B by unilateral deviation from m while i can obtain any alternative in $C(R_i^2, b)$ by unilateral deviation from m. Since we have assumed that m is a Nash equilibrium in state R, it must be the case that $x \in M(R_i, C(R_i^2, b)) \cap M(R_j, B)$. Then by part (ii) of Condition β^1, we have $x \in f(R)$.

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

Case 4: The remaining possibility is that m is such that the modulo 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 $\pi(m) \in M(R_i, B) \cap M(R_j, B)$ and applying (iii) of Condition β^1 to conclude that $\pi(m) \in f(R)$. ■

5.3 Implications

In this section, we briefly discuss the implications of our results in the two-player case. It is easy to verify by inspection that Condition β^1 implies Condition β^2. We first show that Condition β^2 imposes non-trivial restrictions on the class of implementable sccs.

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

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

(i) aP_idP_ib for all $d \notin \{a, b\}$.

(i) bP_jdP_Ja for all $d \notin \{a, b\}$.

Then, $L(P_i, b) \cap L(P_j, b) = \emptyset$, and so Condition β^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. 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 by found in Moulin [9].

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

Definition 10 *Let* $\mu : A \to A$ *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* $i \in N$, $[aP_ib \Leftrightarrow \mu(a)P_i^\mu\mu(b)]$. *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 $A = \{a_1, ..., a_m\}$ where m is even. Consider a preference profile $P \in \mathcal{P}^2$ such that $a_1P_ia_2....P_ia_{m-1}P_ia_m$ and $a_mP_ja_{m-1}....P_ja_2P_ja_1$. 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, $\mu : A \to A$ where $\mu(a_k) = a_{m-k+1}$, $k = 1, ..., m$. Since f is neutral, $a_{m-r+1} \in f(P'^\mu)$. However $P'^\mu = P$, so that $a_{m-r+1} \in f(P)$. Now observe that $L(P_i, a_r) = \{a_r, a_{r+1}, ..., a_m\}$ while $L(P_j, a_{m-r+1}) = \{a_1, ..., a_{m-r+1}\}$. 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 β^2. ■

We now demonstrate the existence of a class of well-behaved social choice functions (singleton-valued sccs) which can be implemented when both individuals are partially honest but not when exactly one individual is partially honest. The possibility result stands in contrast to the negative result of Hurwicz and Schmeidler [3] who showed that there does not exist any two-person, Pareto efficient, non-dictatorial, implementable social choice correspondence.

Let $|A| = m$. Choose integers m_i and m_j such that $m_i + m_j = m - 1$. If m is odd, then we can choose $m_i = m_j = (m-1)/2$. Consider the following "voting by veto" social choice function f defined on $\mathcal{P}^n$.[6] For any $P \in \mathcal{P}^n$, individual i vetoes or eliminates the worst m_i elements in A according to P_i. Denoting this set as V_i, individual j then vetoes the worst m_j elements in $A - V_i$ according to P_j. Denoting this set as V_j, $f^v(P) = A - V_i - V_j$. In order to avoid including dictatorial social choice functions within this class, we assume that $0 < m_i, m_j < m - 1$. Observe that f^v has been defined over the set of all strict order profiles. It is also easy to verify that it is Pareto efficient.

[6]See Moulin [9] for a discussion of Voting by Veto rules.

Proposition 3 *The social choice function f^v is implementable under Assumption A2 but not under A1.*

Proof. For all $P \in \mathcal{P}^2$ and $a = f^v(P)$, let $C(P_i, a) = L(P_i, a)$. Observe that $|L(P_i, a)| \geq m_i + 1$ since individual i is vetoing m_i alternatives.

Pick an arbitrary pair $P, P^1 \in \mathcal{P}^2$ and let $a = f^v(P)$ and $b = f^v(P^1)$. Since $|C(P_i, a)| \geq m_i + 1$ and $|C(P_j^1, b)| \geq m_j + 1$, $m_i + m_j = m - 1$ and $|A| = m$, the intersection of the two sets must be non-empty. Hence part (i) of Condition β^2 is satisfied. Part (ii) of β^2 follows from the fact that f^v is Pareto efficient. Applying Theorem 2, we conclude that f^v is implementable under Assumption A2.

We now show that f^v violates part (ii) of Condition β^2. Let $A = \{a_1, ..., a_m\}$ and let P^1 be the profile where $a_m P_i^1 a_{m-1} ... P_i^1 a_2 P_i^1 a_1$ and $a_1 P_j^1 a_2 ... P_j^1 a_{m-1} P_j^1 a_m$. Clearly $a_{m_i+1} = f^v(P^1)$. Note that $a_1 \notin L(P_j^1, a_{m_i+1})$. Now let P be the profile where $a_{m_i+1} P_i a_1 P_i a_2 ... P_i a_m$ and $a_1 P_j a_{m_i+1} P_j a_2 ... P_j a_m$, i.e P_i and P_j rank alternatives $a_2, ..., a_m$, third through m^{th} while they switch a_1 and a_{m_i+i} between first and second places. Note that $a_{m_i+1} = M(P_j, L(P_j, a_{m_i+1}))$ so that $a_{m_i+1} = M(P_j, C(P_j, a_{m_i+1}))$ for any $C(P_j, a_{m_i+1}) \subset L(P_j, a_{m_i+1})$. Also $a_{m_i+1} = M(P_i, A)$. Hence part (ii) of β^1 requires $a_{m_i+1} = f^v(P)$. However $f^v(P) = a_1$. Applying Theorem 3, we conclude that f^v is not implementable under Assumption A1. ■

Condition β^1 is clearly weaker than the conditions required for two-person implementation without partially honest players. Most obviously, it does not impose Monotonicity which is otherwise a necessary condition. This raises the question of whether it is possible to avoid the Hurwicz-Schmeidler and Maskin impossibility result. We show below that the impossibility result does not hold by constructing a non-dictatorial sub-correspondence of the Pareto correspondence which is implementable under Assumption A1.

Example 1 *Consider the following scc f on domain $\mathcal{P}^n$. Choose some $x^* \in X$. For any $P \in \mathcal{P}^n$,*

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

where $Q(P)$ is the Pareto correspondence.

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 given its bias in favour of x^*. However, it does satisfy Condition β^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 $z \in M(P_i^1, L(P_i, x)) \cap M(P_j^2, X)$. If $z = x^*$, then clearly x^* in $f(P^2)$. Otherwise, since $x^* \in L(P_i, x)$ and $z \in M(P_i^1, L(P_i, x))$, we must have zP_ix^*. Also, if $z \in M(P_i^1, L(P_i, x))$, then $zP_i^2 \in Q(P)$. It is also easy to check that $zP_j^2x^*$. Hence, it follows that $z \in f(P)$.

6 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 are at least three individuals, all social choice correspondences satisfying the weak No Veto Power condition can be implemented. In the two-person case, the results are more subtle but are nevertheless similar in spirit. 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] Busetto, F. and G. Codognato. (2009), "Reconsidering Two-agent Nash Implementation", *Social Choice and Welfare*, 22, 171-179.

[2] Dutta, B. and A. Sen (1991), "A Necessary and Sufficient Condition for Two-Person Nash Implemention", *Review of Economic Studies*, 58, 121-128.

[3] Hurwicz, L. and D. Schmeidler (1978), "Outcome Functions that Guarantee the Existence and Pareto Optimality of Nash Equilibria", *Econometerica*, 46, 144-174.

[4] Jackson, M.O. (2001), “A Crash Course in Implementation Theory”, *Social Choice and Welfare*, 18, 655-708.

[5] Maskin, E. (1999), “Nash Equilibrium and Welfare Optimality”, *Review of Economic Studies*, 66, 23-38.

[6] Matsushima, H. (2008a), “Role of Honesty in Full Implementation”, *Journal of Economic Theory*, 139, 353-359.

[7] Matsushima, H. (2008b), “Behavioral Aspects of Implementation Theory”, *Economic Letters*, 100, 161-164.

[8] Moore, J. and R. Repullo (1990), “Nash Implementation: A Full Characterization”, *Economterica*, 58, 1083-1099.

[9] Moulin, H. (1983), *The Strategy of Social Choice*, Advanced Texbooks in Economics, Editors: C.J. Bliss and M.D. Intrilligator, North-Holland, Amsterdam.