Virtual Implementation in Iteratively Undominated Strategies: Incomplete Information

Dilip Abreu
Princeton University
and
Hitoshi Matsushima
The University of Tsukuba

First Version: September 1990
Revised: August 1992

Abstract

We extend our results for complete information to general Bayesian environments. A social choice function now maps from a finite set of profiles of signals or types to lotteries over alternatives. Self-selection is an obvious necessary condition for virtual implementation under any solution concept. We derive an additional necessary condition which we term measurability. Measurability requires that the social choice function be measurable with respect to a particular partition of each player's signal space. These partitions are derived from players 'state-dependent utility functions and their conditional probability distributions over other players' types. The condition will often be trivially satisfied; the relevant players' partitions will be the finest possible, each element of each player's partition containing a single signal. Under weak domain restrictions we show that any social choice function which satisfies self-selections and measurability is virtually implementable in iteratively undominated strategies. This result is essentially complete, and as permissive as one might hope. It applies to an arbitrary number of players and in particular embraces the two player complete information case.

1 Introduction

This paper considers implementation in Bayesian environments with two or more players. It extends our results for complete information in Abreu and Matsushima (hereafter AM) [1] to this more general setting. In terms of motivation and general background the complete and incomplete information problems have much in common. To avoid repetition, this introduction, and more generally the paper, is written to be read in conjunction with our companion piece on complete information.

The setting is the following. There are a collection of alternatives and a group of players, each of whom receives a private signal. The set of possible signals or types is finite[1]. Every player has a distribution over players' signals, conditional on the signal she receives. (In the special case of complete information each player receives the same signal). Players' preferences over lotteries are, in general, determined by the entire profile of players' signals. A social choice function maps from the space of possible signal profiles to a (fixed) set of possible outcomes, indicating the signal-contingent outcome the principal or planner wishes to impose. An essential feature of our formulation is that the set of outcomes is taken to be the space of (simple) lotteries over some (arbitrary) set of pure alternatives. We will assume that players' preferences over lotteries satisfy the expected utility hypothesis.

A game form or mechanism consists of a message space for each player and an outcome function which specifies an outcome for each message profile. A social choice function is implementable via the iterative elimination of strictly dominated strategies if there exists a game form for which the iterative elimination of strictly dominated strategies leads to a unique iteratively undominated strategy profile with the required outcome for all (possible) signal profiles.

A social choice function satisfies self-selection if truthful reporting is a best response given that other players report their signals truthfully. That is, truthfully reporting is a Bayesian Nash equilibrium in the standard revelation mechanism in which players only report their signals and the outcome function is simply the social choice function itself. Self-selection is a necessary condition for implementation (virtual or exact) using any solution concept. We derive a second necessary condition which is new to the literature. We term this condition measurability. It requires that the social choice function be measurable with respect to a particular partition os each player's signal space. The relevant partitions are determined by players' state-dependent utility functions, and their conditional distributions over other players' signals. Measurability will, of course, be trivially satisfied if the partition is the finest possible, each element of the partition being a singleton. This is in fact the case in complete information environments, and we indicate plausible conditions under which this will be true in general incomplete

[1] See section 5 in AM [1] for a discussion of this assumption.

information environments also. Significantly, measurability is a necessary condition for implementation (virtual and exact) in Bayesian Nash equilibrium also, when the implementing game form satisfies a natural regularity condition. Hence in situations in which the measurability condition is non-trivial it cannot be evaded by weakening the solution concept to Bayesian Nash equilibrium.

Under weak domain restrictions on preferences we show that any social choice function which satisfies self-selection and measurability is virtually implementable via the iterative elimination of strictly dominated strategies. This characterization is essentially complete. Given our domain restrictions, self-selection and measurability are necessary and sufficient for virtual implementation. Since these conditions are also necessary for implementation in Bayesian Nash equilibrium it shows that (modulo virtualness) there is no advantage in weakening the solution concept beyond the iterative removal of strictly dominated strategies.

We are aware of no characterizations involving iterative dominance arguments in general Bayesian environments.[2] There is however a fairly large body of work on Bayesian-Nash implementation in incomplete information settings starting with Postlewaite and Schmeidler [17]. Subsequent contributions include Palfrey and Srivastava [15], [16], Mookherjee and Reichelstein [12], Jackson [6], Matsushima [9], [11], and others. For further references see the survey by Palfrey [14]. This literature shows that self-selection and Bayesian version of the monotonicity condition are necessary for Bayesian Nash implementation. As in the complete information case there is no presumption that monotonicity is a non-trivial condition in general.[3,4] All the drawbacks of the Nash literature carry over to the Bayesian Nash case. These include the use of integer games and the arbitrary exclusion of mixed strategies. See Jackson [7] and our companion paper for a critique. In contrast our results are permissive, obtained for a much weaker solution concept, and use finite mechanisms. The unique iteratively undominated profile is a strict Bayesian Nash equilibrium. Moreover it is a unique equilibrium. Furthermore, our construction applies to an arbitrary number of players, and consequently provides a unified treatment of the two and more than two player case. The two player case has not been considered in the Bayesian implementation literature and has needed new methods of proof in the complete information case. This paper is organized as follows:

[2] See however Matsushima [10].

[3] See Matsushima [11], who argues that Bayesian monotonicity is trivial when side payment is permitted.

[4] Palfrey and Srivastava [16] dispense with this condition for the special case of private values by considering implementation in undominated Bayesian Nash equilibrium. Their mechanisms,are, however, suspect in that they involve the elimination of dominated strategies in *unbounded* mechanisms. See Jackson [7]. In such mechanisms a strategy x may be dominated even though there exists *no* undominated strategy y which dominates x. Such perversities appear to play a key role in obtaining the desired results.

Section 2 develops notations, Section 3 develops a special case, Section 4 discusses measurability and Section 5 presents the theorem. Section 6 shows how the theorem applies to complete information and the two player case. Section 7 concludes.

2 Preliminaries

Let $N = (1, ..., n)$ denote the set of individuals (agents, players), and S_i be the finite set of types or signals of individual i. A profile of signal defines a state $s = (s_1, ..., s_n)$ and $s = \prod_{i \in N} s_i$ denotes the set of possible states. Let $p_i(s_{-i}| s_i)$ denote agent i's conditional probability that other agents receive the profile of signals s_{-i} when she receives the signal s_i. As in our companion paper, A denotes the set of simple lotteries over some arbitrary set of pure alternatives. A pair of lotteries $a, b \in A$ is $\varepsilon - closed$ if the distance between them is at most ε in the usual Euclidean metric $\sum_{\in \Gamma} |a(\tau) - b(\tau)| \leq \varepsilon$ where Γ is the support of lotteries a and b and $a(\tau)$ is the probability of pure alternative τ in the simple lottery a. Player i's state dependent von Neumann-Morgensten utility function is denoted $u_i : A \times S \to R$. It is linear in its first argument. Note that u_i in general depends on all players' signals.

A social choice function $x : S \to A$ maps from states to lotteries. We will write $x = \lambda y + (1 - \lambda) z$ if $x(s) = \lambda y(s) + (1 - \lambda) z(s)$ for all $s \in S$. The social choice functions x and y are $\varepsilon - closed$ if for all $s \in S$, the lotteries $x(s)$ and $y(s)$ are $\varepsilon - closed$. Let

$$U_i(x, s_i) = \sum_{s_{-i} \in S_{-i}} u_i(x(s), s)\, p_i(s_{-i}| s_i)$$

denote individual i's conditional expected utility from a social choice function x when she receives the signal s_i. A mechanism or a game form G is an $(n+1) - tuple$ $(M_1, ..., M_n, g)$, where M_i is a message space for agent i, $M = M_1 \times ... \times M_n$, and $g : M \to A$ is an outcome function. Our constructions only use finite M_i's. Let $\sigma_i : S_i \to M_i$ denote a (pure) strategy for agent i and Σ_i her set of pure strategies. Let

$$v_i(G, \sigma, s_i) = \sum_{s_{-i} \in S_{-i}} u_i(g(\sigma(s)), s)\, p_i(s_{-i}| s_i)$$

denote agent i's conditional expected utility from a mechanism G under the strategy profile σ, when she receives the signal s_i. Note that a player's strategy specifies a message for each of her possible types.

Fix a game form $G = (M, g)$ arbitrarily. Let H_i be a subset of Σ_i. A strategy $\sigma_i \in H_i$ is strictly dominated for player i with respect to $H = \prod_{j \in N} H_j$ if there exist $\sigma_i \in H_i$ and $s_i \in S_i$ such that for every $\sigma_i \in H_i$,

$$v_i(G, \sigma/\sigma_i', s_i) > v_i(G, \sigma, s_i)$$

Note that from the point of view of a player i who receives a signal s_i the domination is strict with respect to all possible strategies of other players. Let $Q_i(H)$ denote the set of all undominated strategies for agent i with respect to H. Let $Q_i(H) = \prod_{i \in N} Q_i(H)$. Let $Q_i^0(\Sigma) = \Sigma_i$, $Q^k = \prod_{i \in N} Q_i^k$ where $Q_i^k(\Sigma) = Q_i(Q^{k-1}(\Sigma))$. For simplicity, we write Q^k for $Q^k(\Sigma)$, etc.

Let Q^* denote the intersection of $\{Q^k, k = 0, 1, ...\}$. A strategy profile $\sigma \in \Sigma$ is iteratively undominated if $\sigma \in Q^*$. Since Σ is finite, there exists k such that $Q^k = Q^*$for all $k \geq k$. The mechanism $G = (M, g)$ exactly implements a social choice function x in iteratively undominated strategies if and only if Q^* is a singleton and $g(\sigma^*(s)) = x(s)$ for all $s \in S$, where $Q^* = \{\sigma^*\}$.

As in our earlier paper, the order of elimination of strictly dominated strategies is irrelevant. If σ^* is the unique strategy which survives the iterative elimination of strictly dominated pure strategies, it will also be the unique strategy which survives the iterative elimination of strictly dominated (pure and) mixed strategies. Furthermore, σ^* is a unique mixed Bayesian Nash equilibrium, and is a strict equilibrium for each type of every player.

For every $i \in N$, every $s_i \in S_i$ and every $s_i' \in S_i$, let

$$V_i(x, s_i, s_i') = \sum_{s_{-i} \in S_{-i}} u_i\left(x\left(s/s_i'\right), s\right) p_i\left(s_{-i} \middle| s_i\right)$$

denote the expected utility of the direct mechanism (S, x) for player i conditional on s_i when she announces s_i' and the other players make truthful announcements. Note that

$$V_i(x, s_i, s_i) = U_i(x, s_i)$$

Definition 1 *A social choice function x satisfies self-selection if for every $i \in N$, every $s_i \in S_i$ and every $s_i' \in S_i \backslash \{s_i\}$,*

$$U_i(x, s_i) \geq V_i(x, s_i, s_i').$$

A social choice function x satisfies strict self-selection if these inequalities strictly hold.

Self-selection means that truth-telling is a Bayesian Nash equilibrium in the direct mechanism (S, x). Self-selection is a well-known (and immediate) necessary condition for implementation in Bayesian Nash equilibrium, or in any refinement of Bayesian Nash equilibrium.

3 A Special Case

Before proceeding to our general characterization we consider a special case which we hope will serve as an easy introduction to the incomplete information setting. In this section, we assume:

(1) That small side-payments are possible and that these affect utility additively. Specifically, we denote by $t_i \in [-\varepsilon, \varepsilon]$ the side-payment to payer i and suppose that her total utility is $u_i(a, s) + t_i$.

(2) Private values: Player i's preferences over lotteries depend only on her own type. Since u_i is independent of s_{-i}, we simply write $u_i(a, s_i)$ instead of $u_i(a, s_i, s_{-i})$.

(3) That for every $i \in N$, and $s_i \in S_i$, there exist $a, a' \in A$ such that

$$u_i(a, s_i) > u_i(a', s_i).$$

That is, for each type of player i, we exclude the possibility of universal indifference across lotteries.

(4) That distinct types of player i have distinct orderings over lotteries.

Assumption (4) is not innocuous. In particular it excludes the complete information case. A player's signal represents what she knows, and in the complete information setting this is the entire profile of players' preferences. Hence two distinct signals for player i may represent differences in other players' preference orderings and not her own. See Section 4 for a further discussion of this point.

The preceding assumptions (specifically $(2)-(4)$) imply that for every $i \in N$ there exists a function $f_i : S_i \to A$ such that for every $s_i \in S_i$ and every $s_i' \in S_i \backslash \{s_i\}$,

$$u_i(f_i(s_i), s_i) > u_i(f_i(s_i'), s_i).$$

This straightforward result corresponds to the Lemma in AM [1], and may be proved analogously. We may think of f_i as a social choice function which is independent of s_{-i}. It gives player i's "dictatorial" choice within the set $\{a : a = f_i(s_i) \text{ for some } s_i \in S_i\}$. By construction this choice varies with s_i.

We now construct a mechanism as follows. Every player i makes $(K+1)$ simultaneous announcements, each of which is of her own signal $M_i = M_i^0 \times M_i^1 \times ... \times M_i^K = S_i \times ... \times S_i$. Recall that S_i is finite so that this construction will yield a finite mechanism. Denote

$$\begin{aligned} m_i &= (m_i^0, ..., m_i^K) \in M_i,\ m_i^h \in M_i^h \\ m &= (m^0, ..., m^K) \in M,\ m^h = (m_i^h)_{i\in N} \in M^h = \underset{i\in N}{\times} M_i^h \end{aligned}$$

Given a social choice function x, for any profile of player messages m, the lottery chosen is

$$g(m) = \frac{\varepsilon}{n} \sum_{i \in N} f_i(m_i^0) + \frac{1}{K} \sum_{h=1}^{K} \left(\frac{\varepsilon^2}{n} \sum_{i \in N} f_i(m_i^h) + \left(1 - \varepsilon - \varepsilon^2\right) x(m^h) \right)$$

where ε is small and strictly positive. Equivalently,

$$g(m) = \frac{\varepsilon}{n} \sum_{i \in N} f_i(m_i^0) + \frac{1 - \varepsilon}{K} \sum_{h=1}^{K} x'(m^h)$$

where

$$x'(m^h) = \frac{\alpha}{n} \sum_{i=1}^{n} f_i(m_i^h) + (1 - \alpha)\, x(m^h)$$

and $\alpha \equiv \frac{\varepsilon^2}{1-\varepsilon}$.

Note that given x that satisfies self-selection, x' satisfies strict self-selection. This is because of the addition of the f_i terms. Of course, for small ε, x' is "close" to x.

In the above mechanism, player i's zero-th announcement affects the outcome with probability $\frac{\varepsilon}{n}$. With this probability the outcome is $f_i(m_i^0)$, and this lottery depends only on player i's zero-th announcement. On the other hand, player i's h-th announcement $(h \geq 1)$ affects the outcome with probability $\frac{1-\varepsilon}{K}$ via the $x'(m_i^h, m_{-i}^h)$ which does depend on other players' h-th announcements.

In addition to this lottery players receive small fines according to the following rules. The first deviant from her own zero-th announcement is fined η. That is, player i is fined $t_i : M \to \mathbb{R}$ where

$t_i(m) = -\eta$	if player i is the first deviant, i.e., for some h, $m_i^h \neq m_i^0$ and $m^{h'}$ m^0 for all $h' < h$.
$t_i(m) = 0$	otherwise

Choose η sufficiently small, such that for every $i \in N$,

$$\frac{\varepsilon}{n} \left(u_i \left(f_i(s_i), s_i \right) - u_i \left(f_i(s_i'), s_i \right) \right) > \eta \text{ for all } s_i \in S_i \text{ and all } s_i' \in S_i \backslash \{s_i\}$$

Note that η is smaller than the "direct" expected utility loss from a zero-th message $m_i^0 \neq s_i$.

Let σ be an iteratively undominated strategy profile. Recall that $\sigma_i : S_i \to M_i$. Denote strategies for players by

$$\begin{aligned} \sigma_i &= (\sigma_i^0, ..., \sigma_i^K), \;\; \sigma_i^h : S_i \to M_i^h \\ \sigma &= (\sigma^0, ..., \sigma^K), \;\; \sigma^h : S \to M^h \end{aligned}$$

Since m_i^0 affects player i's utility only through f_i and t_i, it follows directly from the definition of η that if σ_i is iteratively undominated, then

$$\sigma_i^0 (s_i) = s_i.$$

We will now show that if for every $i \in N$, every iteratively undominated strategy profile σ_i and every $s_i \in S_i$, $\sigma_i^h (s_i) = s_i$ for all $h \in \{0, ..., k\}$, then

$$\sigma_i^{k+1} (s_i) = s_i \text{ for all } i \in N \text{ and all } s_i \in S_i.$$

The following inductive step completes the argument: for any (small) $\varepsilon > 0$, the mechanism yields a unique iteratively undominated profile σ with $\sigma_i^k (s_i) = s_i$ for all $s_i \in S_i$, all $i \in N$ and all $k = 0, ..., K$. The resultant income is

$$\left(1 - \varepsilon - \varepsilon^2\right) x (s) + \sum_{i \in N} \left(\frac{\varepsilon}{n} + \frac{\varepsilon^2}{n} \right) f_i (s_i),$$

and there are no fines.

For the inductive step to go through we must choose K sufficiently large, such that for every $i \in N$, every $s_i \in S_i$ and every $s' \in S$,

$$\eta > \frac{1 - \varepsilon}{K} \left(u_i \left(x' \left(s'/s_i \right), s_i \right) - u_i \left(x' (s), s_i \right) \right).$$

Then the fine η is larger than $\frac{1-\varepsilon}{K}$ times the maximal "direct" utility gain from changing the outcome.

suppose that $\sigma_i^{k+1} (s_i) \neq s_i$ for some $i \in N$ and some $s_i \in S_i$. Define $\overline{\sigma}_i$ such that

$$\begin{aligned} &\overline{\sigma}_i^h = \sigma_i^h \text{ for } h \neq k+1, \;\; \overline{\sigma}_i^{k+1} (s_i') = \sigma_i^{k+1} (s_i) \;\text{ for all } s_i' \in S_i \backslash \{s_i\} \\ &\text{and } \overline{\sigma}_i^{k+1} (s_i) = s_i \end{aligned}$$

Under the inductive hypothesis above if $\sigma_j^{k+1} (s_j) = s_j$ for all $j \in N \backslash \{i\}$ and all $s_j \in S_j$, then by strict self-selection $\overline{\sigma}_i$ yields higher payoff than σ_i even if player i is the first to deviate given $(\overline{\sigma}_i, \sigma_{-i})$. On the other hand, if for some $j \in N \backslash \{i\}$, $\sigma_j (s_j) \neq s_j$ for all $s_j \in S_j$, then (by the choice of K above) $\overline{\sigma}_i$, by saving the fine η, yields a higher payoff than σ_i. Thus $\overline{\sigma}_i$ dominates σ_i for player i of type s_i. In fact this argument is not complete in that some types of player y may announce $\sigma'_j(s_j) \neq s_j$, while others make truthful announcements. This possibility is allowed for in the detailed arguments we present below. Assumptions (2) and (4) are the non-innocuous simplifying assumptions used here. These assumptions are relaxed in the next two sections which also provide formal proofs.

4 Measurability

This section introduces the measurability condition, discusses issues of computation, establishes that measurability is necessary for implementation in Bayesian Nash equilibrium also, and finally provides a simple sufficient condition under which measurability is satisfied by any social choice function.

4.1 The condition

As indicated in the Introduction, the measurability condition requires that the social choice function be measurable with respect to a particular partition of each player's signal space. In order to clarify exactly what this condition means it is useful to start with a little notation.

Denote by Ψ_i a partition of S_i, where ψ_i is a generic element of Ψ_i and $\gamma_i(s_i)$ is the element of Ψ_i which includes s_i. Let $\Psi = \times_{i \in N} \Psi_i$ and $\psi = \times_{i \in N} \psi_i$.

Definition 2 *A social choice function x is measurable with respect to Ψ if for every $i \in N$, every $s_i \in S_i$ and $s'_i \in S_i$,*

$$x(s) = x(s/s'_i) \text{ for all } s_{-i} \in S_{-i} \text{ whenever } \gamma_i(s_i) = \gamma_i(s'_i)$$

A social choice function x is strictly measurable with respect to Ψ if for every $i \in N$, every $s_i \in S_i$ and every $s'_i \in S_i$,

$$x(s) = x(s/s'_i) \text{ for all } s_{-i} \in S_{-i} \text{ if and only if } \gamma_i(s_i) = \gamma_i(s'_i)$$

Moreover, a social choice function x is measurable with respect to a social choice function y if for any $i \in N$ and any $s_i, s'_i \in S_i$,

$$x(s) = x(s/s'_i) \text{ for all } s_{-i} \in S_{-i} \text{ whenever } y(s) = y(s/s'_i) \text{ for all } s_{-i} \in S_{-i}$$

Measurability of x with respect to Ψ implies that for any player i, x does not distinguish between any pair of signals in the same element of the partition Ψ_i.

Definition 3 *A strategy σ_i for player i is measurable with respect to Ψ_i if for every $s_i \in S_i$ and every $s_i' \in S_i$,*

$$\sigma_i(s_i) = \sigma_i(s_i') \text{ whenever } \gamma_i(s_i) = \gamma_i(s_i')$$

A strategy profile σ is measurable with respect to Ψ if for every $i \in N$, σ_i is measurable with respect to Ψ_i.

Consider player i, and Y^i, the set of social choice functions which depend only on other players' signals. For any signal s_i player i receives, she can rank elements of Y^i in terms of conditional expected utility. In any game form, and for any tuple of strategies σ_{-i} of other players, a message m_i by player i generates a social choice function in Y^i (y_i such that $y_i(s) = g(m_i, \sigma_{-i}(s_{-i}))$).

If the signals s_i and s_i' generate the same posterior ordering over Y^i then, in any game form, player i's optimal strategies when she is type s_i are identical to her optimal strategies when she is type s_i'. Player i could not possibly have a strict incentive to send a distinct message for each of these signals. Hence if x is implementable it must be that $x(s_i, s_{-i}) = x(s_i', s_{-i})$ for all $s_{-i} \in S_{-i}$. That is, x must be measurable with respect to some partition of S_i with s_i and s_i' in the same cell of this partition. The partition generated by posterior orderings over Y^i is the finest possible partition with respect to which an implementable social choice function must be measurable. In general the relevant partition might be coarser. This is because in distinguishing between any pair of signals s_i and s_i' we are using the richest possible set of social choice functions Y^i which depends only on other players' signals. The same analysis applied to players $j \neq i$ may yield non-trivial partitions Ψ_j of the S_j's. Then in order to discriminate between player i's signals we may only use a set of social choice functions which are measurable with respect to Ψ_{-i}. Proceeding in this way we iteratively obtain coarser and coarser partitions. As will become clear the procedure described below starts at the "other end" and iteratively yields finer and finer partitions.

For every $i \in N$, $s_i \in S_i$, $s_i' \in S_i$ and $(n-1)$-tuple of partitions Ψ_{-i}, s_i is equivalent to s_i' with respect to Ψ_{-i} if for every x and every y which are measurable with respect to $\{S_i\} \times \Psi_{-i}$,

$$U_i(x, s_i) \geq U_i(y, s_i) \text{ if and onlt if } U_i(x, s_i') \geq U_i(y, s_i').$$

Let $\rho_i(s_i, \Psi_{-i})$ be the set of all elements of S_i equivalent to s_i with respect to Ψ_{-i}, and let

$$R^i(\Psi_{-i}) = \{\rho_i(s_i, \Psi_{-i}) : s_i \in S_i\}.$$

We define an infinite sequence of n-tuples of partitions, $\{\Psi^h\}_{h=0}^{\infty}$, $\Psi^h = \Psi_1^h \times ... \times \Psi_n^h$, in the following way. For every $i \in N$,

$$\Psi_i^0 = \{S_i\},$$

and recursively, for every $i \in N$ and every $h = 1, 2, ...,$

$$\Psi_i^h = R^i(\Psi_{-i}^{h-1}).$$

Note that for every $h = 0, 1, ...,$ Ψ_i^{h+1} is the same as, or finer than, Ψ_i^h. Since S_i is finite, there exists a positive integer L such that for every $h \geq L$, $\Psi^h = \Psi^L$. We denote $\Psi^* = \Psi^L$.

We will argue that a necessary condition for a social choice function to be virtually implementable is that it be measurable with respect to Ψ^*.

Definition 4 *A social choice function satisfies the measurability condition if it is measurable with respect to* Ψ^**.*

Consider a game form $G = (M_1, ..., M_n; g)$ which exactly implements a social choice function y in iteratively strictly undominated strategies and let Q_i^k, Q^k, the sets of iteratively undominated strategies at the k-th round of iterative removal, etcetera, be defined as in Section 2.

Consider an arbitrary constant strategy profile $\sigma[0] \in Q^0$ (that is, $\sigma[0]$ is measurable with respect to $\times_{i \in N} \{S_i\}$). By the definition of Ψ^1, it follows that for every $i \in N$, there exists $\sigma_i[1] \in \Sigma_i$ which is a best response to $\sigma[0]$ and is measurable with respect to Ψ^1. Hence, $\sigma_i[1]$ is not strictly dominated for player i with respect to Σ, that is, $\sigma_i[0] \in Q_i^1$.

Fix $k = 2, 3, ...$ arbitrarily, and suppose that there exists a strategy profile $\sigma[k-1] \in Q^{k-1}$ which is measurable with respect to Ψ^{k-1}. Then, it is easy to see that for every $i \in N$, there exists $\sigma_i[k] \in \Sigma_i$ which is a best response to $\sigma[k-1]$ and is measurable with respect to Ψ_i^k. Since $\sigma_i[k]$ is a best response it is not strictly dominated for player i with respect to $\Sigma_i \times Q_{-i}^{k-1}$. As we are eliminating strictly dominated strategies, $Q_i^k = Q\left(\Sigma_i \times Q_{-i}^{k-1}\right)$. Hence for all $k = 0, 1, ...,$ there exists $\sigma[k] \in Q^k$ which is measurable with respect to Ψ^k.

Let σ^* be the unique iteratively undominated strategy profile in the implementing game form G. Then the preceding argument implies that σ^* is measurable with respect to Ψ^*. It follows that $y = g \cdot \sigma^*$ is measurable with respect to Ψ^* also, i.e., satisfies the measurability condition.

Now suppose that $\{y^m\}_{m=1}^{\infty}$ is a sequence of social choice functions such that y^m is exactly implementable, and $(\frac{1}{m})$-close to x. Clearly there exists $\overline{m}$ such that for all $m = \overline{m}, \overline{m}+1, ...,$ x is measurable with respect to y^m (and vice-versa). Thus if x is virtually implementable there exists y which is measurable with respect to x and exactly implementable. Hence we have proved that:

Proposition 1. *If a social choice function is virtually implementable in iteratively undominated strategies, then it satisfies the measurability condition.*

4.2 Computing Ψ^*

There are simple sufficient conditions under which no computation is necessary: The partition is simply the finest possible partition $\overline{\Psi}_i$, each element of Ψ_i^* containing only a single element. See Section 4.4. When computation is required and the set of pure alternatives Γ is finite, Ψ^* may be determined via a finite procedure as follows. Define

$$X^P = \{x : x(s) \text{ is a degenerate lottery for all } s \in S\}.$$

If Γ is finite so also is X^P.

The following definition of equivalence of signals is implied by and implies the definition given earlier: s_i is equivalent to s_i' with respect to Ψ_{-i} if there exist α and $\beta > 0$ such that for any $x \in X^P$ which is measurable with respect to $\{S_i\} \times \Psi_{-i}$,

$$U_i(x, s_i) = \alpha + \beta U_i(x, s_i').$$

We need only consider social choice functions which map to pure outcomes and check the condition above.

4.3 Measurability and Bayesian Nash implementation

We argue here that measurability is a necessary condition for implementation in Bayesian Nash equilibrium. When implementing game forms satisfying a natural regularity condition. Thus measurability is a weak condition; requiring implementation in iteratively undominated strategies does not entail requirements beyond these needed for Bayesian implementation.

For any mechanism $G = (M, g)$, let $BN(G)$ be the set of Bayesian Nash equilibria of G. A social choice function x is implementable in Bayesian Nash equilibrium by the game form G if $BN(G) \neq \emptyset$ and for all $\sigma \in BN(G)$,

$$g(\sigma(s)) = x(s) \text{ for all } s \in S.$$

For every $i \in N$ and every partition Ψ_i, let $\Sigma_i(\Psi_i)$ denote the set of strategies of player i which are measurable with respect to Ψ_i. The profile $\sigma \in \times_{i \in N} \Sigma_i(\Psi_i)$ is a pseudo-Bayesian Nash equilibrium with respect to Ψ in G if for all $i \in N$ and all $\psi_i \in \Psi_i$, there exists some $s_i \in \psi_i$ such that

$$v_i(G, \sigma, s_i) \geq v_i(G, \sigma/\sigma_i', s_i) \text{ for all } \sigma_i' \in \Sigma_i.$$

We will say that G is regular if for any Ψ, a pseudo-Bayesian Nash equilibrium exists. Regularity is a minimal requirement, and will be satisfied, for instance, under any of the standard conditions used to establish the existence of a Bayesian Nash equilibrium via a fixed point argument.[5]

Proposition 2. *If a social choice function is virtually implementable in Bayesian Nash equilibrium by a regular game form, then it satisfies the measurability condition.*

Proof. As in the earlier proof, it suffices to establish the result for exactly implementable social choice functions. Let $G = (M, g)$ exactly implement a social choice function x, and let $\sigma \in \times_{i \in N} \Sigma_i (\Psi_i^*)$ be a pseudo-Bayesian Nash equilibrium with respect to Ψ^*. If $m_i = \sigma_i(s_i)$ is a best response for player i with signal s_i then m_i is also a best response for any $s_i' \in \xi_i(s_i, \Psi_i^*)$, where $R^i(\Psi_{-i}^*) = \left\{\rho_i\left(t_i, \Psi_{-i}^*\right) : t_i \in S_i\right\}$, but $R^i(\Psi_{-i}^*) = \Psi_i^*$. It follows that any pseudo-equilibrium σ which is measurable with respect to Ψ^* is in fact a regular equilibrium. Since $x = g \cdot \sigma$, x must satisfy the measurability condition. ■

4.4 A simple sufficient condition

We now present a simple sufficient condition under which measurability will be automatically satisfied. This condition is based on the possibility of small side payments and based on the idea of "scoring rules" (see Good [5] and Winkler [19]). Recall that the finest possible of partitions is denoted by $\overline{\Psi}_i$. Let $\overline{\Psi} = \times_{i \in N} \overline{\Psi}_i$. Clearly any social choice function is measurable with respect to $\overline{\Psi}$ so that measurability with respect to $\overline{\Psi}$ is trivially satisfied.

Let $t_i \in [-\varepsilon, \varepsilon]$ denote the side payment to player i and suppose that her total utility is $u_i(a, s) + t_i$. Let

$$P_i\left(\psi_{-i} \middle| s_i\right) = \sum_{s_{-i} \in \psi_{-i}} p_i\left(s_{-i} \middle| s_i\right),$$

and suppose that for every $i \in N$, every $s_i \in S_i$, every $s_i' \in S_i \backslash \{s_i\}$, there exists $\psi_{-i}^* \in \Psi_{-i}^*$ such that $p_i\left(\psi_{-i}^* \middle| s_i\right) \neq p_i\left(\psi_{-i}^* \middle| s_i'\right)$. Construct a transfer rule $\widehat{t}_i : S_i \times \Psi_{-i}^* \to [-\varepsilon, \varepsilon]$ such that

$$\widehat{t}_i\left(s_i, \psi_{-i}^*\right) = -\alpha\left(1 - p_i\left(\psi_{-i}^* \middle| s_i\right)\right)^2 - \alpha \sum_{\psi_{-i} \in \Psi_{-i}^* \backslash \left\{\psi_{-i}^*\right\}} p_i\left(\psi_{-i} \middle| s_i\right),$$

where α is a positive real number small enough to satisfy

$$-\varepsilon \leq \widehat{t}_i\left(s_i, \psi_{-i}^*\right) \leq \varepsilon \text{ for all } s_i \in S_i \text{ and all } \psi_{-i}^* \in \Psi_{-i}^*.$$

[5] Most of the mechanisms that have been used in the Nash and Bayesian Nash literature are not regular. We take this to be additional evidence of the unsatisfactory nature of these mechanisms.

Note that for every $s_i \in S_i$ and every $s_i' \in S_i \backslash \{s_i\}$,

$$\sum_{\psi_{-i} \in \Psi^*_{-i} \backslash \{\psi^*_{-i}\}} \left(\widehat{t}_i \left(s_i, \psi^*_{-i}\right) - \widehat{t}_i \left(s_i', \psi^*_{-i}\right) \right) p_i \left(\psi^*_{-i} \middle| s_i \right) > 0.$$

Now consider the "extended" social choice functions which map to $A \times [-\varepsilon, \varepsilon]^n$ so that the social outcome consists of a lottery in A and a transfer payment. Let Ψ^{**} be the analogue of Ψ^* for "extended" social choice functions. It follows from the construction of transfer rule $\widehat{t}_i$ above that $\Psi^{**} = \overline{\Psi}$.

5 The Theorem

Our main theorem asserts that self-selection and measurability are necessary and sufficient for implementation in iteratively undominated strategies. The necessity of self-selection and measurability were argued earlier and it only remains to establish sufficiency. The latter is true under two weak assumptions which are satisfied automatically if arbitrarily small fines may be levied. Before proceeding to details we provide a brief initial discussion premissed on the availability of small fines.

Definition 5 *A social choice function x satisfies set self-selection for player i with respect to Ψ if it is measurable with respect to Ψ and*

$$U_i(x, s_i) \geq V_i(x, s_i, s_i') \text{ for all } s_i \in S_i \text{ and } s_i' \in S_i \backslash \{\gamma_i(s_i)\}.$$

A social choice function satisfies strict set self-selection for player i with respect to Ψ if it satisfies set self-selection for player i with respect to Ψ and the above inequalities strictly hold.

The strategy of proof is as follows. We first show that there exists a social choice function $\widetilde{x}$ which is strictly measurable with respect to Ψ^*, and is moreover exactly implementable in iteratively undominated strategies (this step is discussed below). This function is implementable using a mechanism in which players announce elements of their partition Ψ_i^*. Given this social choice function $\widetilde{x}$, we can virtually implement any social choice function x' which satisfies strict set self-selection for all players and measurable with respect to $\widetilde{x}$, or equivalently measurable with respect to Ψ^*, by mimicking the argument of Section 3 replacing announcements of signals by announcements of cells of the partition Ψ_i^*: if a social choice function x satisfies the measurability condition, we will define a function $x^+ : \Psi^* \to A$ by

$$x^+(\psi) = x(s) \text{ wenever } \psi = \gamma^*(s),$$

where $\gamma_i^*(s_i)$ is the element of Ψ_i^* that includes s_i, and $\gamma^*(s) = (\gamma_i^*(s_i))_{i \in N}$. Let

$$\begin{aligned} M_i &= \Psi_i^* \times \Psi_i^* \times ... \times \Psi_i^*, \\ g(m) &= \varepsilon \widetilde{x}^+ \left(m^0\right) + (1-\varepsilon) \frac{1}{K} \sum_{h=1}^{K} x^+(m^h). \end{aligned}$$

In addition small fines are levied as before. The logic of the argument is now similar to the one presented earlier.

Note that for any social choice function x which satisfies the measurability condition and set self-selection for all players with respect to Ψ^* there exists a nearby social choice function x' which in addition satisfies strict set self-selection for all players with respect to Ψ^*. This social choice function is

$$\eta \widetilde{x} + (1-\eta) x.$$

Thus for virtual implementation the distinction between weak and strict self-selection is unimportant.

How do we prove the existence of $\widetilde{x}$? Recall the "dictatorial" function f_i of Section 3. By an analogous argument, there exists a social choice function x_i^1 which satisfies strict set self-selection for player i with respect to $\Psi_i^1 \times \Psi_{-i}^0$. Hence the social choice function $\frac{1}{n}\sum_{i \in N} x_i^1$ is strictly measurable with respect to Ψ^1, and is implemented via (one round of) elimination of strictly dominated strategies.

Can we go a step further? By exactly the same logic there exists a social choice function x_i^2 which is measurable with respect to $\Psi_i^2 \times \Psi_{-i}^1$ and satisfies strict set self-selection for player i. The for small enough ε the social choice function

$$\frac{1}{\varepsilon + \varepsilon^2} \left(\frac{\varepsilon}{n} \sum_{i \in N} x_i^1 + \frac{\varepsilon^2}{n} \sum_{i \in N} x_i^2 \right)$$

is implementable in iteratively strictly undominated strategies, and is strictly measurable with respect to Ψ^2. We now need two rounds of iterative removal.

Proceeding in this way we inductively obtain the desired $\widetilde{x}$. Now to details and formal proofs.

Assumption 1. *For every $i \in N$, every $s_i \in S_i$, every $h \in (0, ..., L)$ and every $\psi_{-i} \in \Psi_{-i}^h$, if $p_i\left(\psi_{-i} \middle| s_i\right) > 0$, then there exists $a \in A$ and $a' \in A$ such that*

$$\sum_{s_{-i} \in \psi_{-i}} \left(u_i(s,a) - u_i(a',s)\right) p_i\left(s_{-i} \middle| s_i\right) > 0, \textit{ where } p_i\left(\psi_{-i} \middle| s_i\right) = \sum_{s_{-i} \in \psi_{-i}} p_i\left(s_{-i} \middle| s_i\right).$$

It is clear that this assumption is weak; it rules out indifference (in terms of conditional expected utility) across all lotteries. It is satisfied trivially if small transfers of private goods are permitted.

Lemma 1. *For every $i \in N$ and $h = 1, ..., L$, there exists a social choice function x_i^h which satisfies strict set self-selection for player i with respect to $\Psi_i^h \times \Psi_{-i}^{h-1}$, where L is the positive integer introduced in Section 4 such that $\Psi^L = \Psi^*$.*

Proof. Fix $i \in N$ and $h = 1, ..., L$ arbitrarily. Let Y_i^{h-1} be the set of all social choice functions which are measurable with respect to $\{S_i\} \times \Psi_i^{h-1}$. Then s_i and s_i' induce different preference orderings over Y_i^{h-1} if and only if $\gamma_i^h(s_i) \neq \gamma_i^h(s_i')$. By Assumption 1, neither of these preference orderings involves complete indifference. Hence, for all $\psi_i \in \Psi_i^h$ and $\psi_i' \in \Psi_i^h \backslash \{\psi_i\}$, there exist social choice functions x and y which are measurable with respect to $\{S_i\} \times \Psi_{-i}^{h-1}$ such that

$$\begin{aligned} U_i(x, s_i) &> U_i(y, s_i) \text{ if } \gamma_i^h(s_i) = \psi_i, \text{ and} \\ U_i(y, s_i) &> U_i(x, s_i) \text{ if } \gamma_i^h(s_i) = \psi_i'. \end{aligned}$$

Let Z_i^h be a finite subset of Y_i^{h-1} such that for all $\psi_i \in \Psi_i^h$ and $\psi_i' \in \Psi_i^h \backslash \{\psi_i\}$, there exist $x \in Z_i^h$ and $y \in Z_i^h$ with the above properties. Let $J = |Z_i^h|$. For all $\psi_i \in \Psi_i^h$ and $j \in \{1, ..., J\}$ let $x_j^{\psi_i}$ be social choice functions which satisfy $\{x_j^{\psi_i} : j \in \{1, ..., J\}\} = Z_i^h$, and

$$U_i(x_j^{\psi_i}, s_i) \geq U_i(x_{j+1}^{\psi_i}, s_i) \text{ for all } j = 1, 2, ..., J-1 \text{ if } \gamma_i^h(s_i) = \psi_i.$$

Consider the social choice function x_i^h defined by

$$x_i^h(s) = \sum_{j=1}^{J} \alpha_j x_j^{\psi_i}(s) \text{ if } \gamma_i^h(s_i) = \psi_i,$$

where the α_j's are strictly positive, strictly decreasing in j, and sum to one. Then x_i^h is measurable with respect to $\Psi_i^h \times \Psi_{-i}^{h-1}$ and satisfies strict set self-selection for player i. ■

Lemma 2. *There exists a social choice function which is exactly implementable in iteratively undominated strategies and is strictly measurable with respect to Ψ^*.*

Proof. Define a social choice function $\widetilde{x}$ by

$$\widetilde{x}(s) = \alpha \sum_{i \in N} \sum_{h=1}^{L} \varepsilon^h x_i^h(s),$$

where the social choice functions x_i^h are as in Lemma 1, and

$$\alpha = \frac{1}{n(\varepsilon + \varepsilon^2 + ... + \varepsilon^L)}.$$

For small enough $\varepsilon > 0$, $\widetilde{x}$ is strictly measurable with respect to Ψ^*.[6] We will show that $\widetilde{x}$ is exactly implementable.

We define a mechanism $\widetilde{G} = (\widetilde{M}, \widetilde{g})$ by

$$\widetilde{M}_i = \Psi_i^* \text{ for all } i \in N,$$

and

$$\widetilde{g} = \widetilde{x}^+.$$

We can choose a positive real number $\eta > 0$ such that for every $i \in N$, every $s_i \in S_i$, every $s_i' \in S_i$ and every $h \in \{1, ..., L\}$, if $s_i' \notin \gamma_i^h(s_i)$, then

$$U_i(x_i^h, s_i) - V_i(x_i^h, s_i, s_i') > \eta.$$

For every $x \in X$, define

$$F_i(x) = \max_{(s,s')} \{u_i(x(s'), s) - u_i(x(s'/s_i), s)\}.$$

Choose ε small enough such that for every $i \in N$ and every $h \in \{1, ..., L\}$,

$$\eta > \sum_{j \in N} \sum_{k=h+1}^{L} \varepsilon^{k-h} F_i(x_j^k).$$

Let θ_i and θ_i' be strategies for agent i in $\widetilde{G}$.

Let $P(h)$ be the statement: "if θ is iteratively undominated, then for every $i \in N$ and every $s_i \in S_i$, $\theta_i(s_i)$ is a subset of $\gamma_i^h(s_i)$." Recall that $\gamma_i^0(s_i) = S_i$ for all $s_i \in S_i$. Then $P(0)$ holds trivially. We will show that for all $h = 1, ..., L$, $P(h-1) \Rightarrow P(h)$.

Suppose that $P(h-1)$ and consider an iteratively undominated strategy θ_i for player i. Fix $s_i \in S_i$ arbitrarily. Suppose $\theta_i(s_i)$ is a subset of $\gamma_i^h(s_i)$, and $\theta_i'(s_i)$ is a subset of $\gamma_i^{h-1}(s_i)$ and also a subset of $\gamma_i^h(s_i')$ for some $s_i' \in S_i \backslash \{\gamma_i^h(s_i)\}$. Then

$$\begin{aligned} v_i(\widetilde{G}, \theta, s_i) - v_i(\widetilde{G}, \theta/\theta_i', s_i) &\geq \varepsilon^h \alpha \left(U_i(x_i^h, s_i) - V_i(x_i^h, s_i, s_i') - \sum_{j \in N} \sum_{k=h+1}^{L} \varepsilon^{k-h} F_i(x_j^k) \right) \\ &\geq \varepsilon^h \alpha \left(U_i(x_i^h, s_i) - V_i(x_i^h, s_i, s_i') - \eta \right) \\ &> 0. \end{aligned}$$

[6] Suppose that x is strictly measurable with respect to some partition Ψ and y is measurable with respect to some coarser partition Ψ'. Then $(1-\lambda)x + \lambda y$ need not be strictly measurable with respect to Ψ for arbitrary λ, but will be for small enough λ.

This means that if θ_i is iteratively undominated, $\theta_i(s_i)$ is a subset of $\gamma_i^h(s_i)$, and therefore, $P(h)$ must hold.

Hence $\theta = \gamma^*$. Finally recall that

$$\widetilde{g}(\gamma^*(s)) = \widetilde{x}(s) \text{ for all } s \in S. \quad \blacksquare$$

Assumption 2. *For every $i \in N$ and every $\psi \in \Psi^*$, there exist $\overline{a}(i,\psi) \in A$ and $\underline{a}(i,\psi) \in A$ such that for every $s \in \psi$, if $p_i(s_{-i}|\, s_i)$, then*

$$u_i(\overline{a}(i,\psi), s) - u_i(\underline{a}(i,\psi), s) > 0,$$

and

$$u_j(\underline{a}(i,\psi), s) - u_j(\overline{a}(i,\psi), s) \geq 0 \text{ for all } j \in N \backslash \{i\}.$$

Assumption 2 corresponds to the single assumption used in our earlier paper on complete information. Like Assumption 1, Assumption 2 is trivially satisfied if strictly positive (though possibly, arbitrarily small) transfers of private goods are possible. It requires that for every player i and every signal profile, there exists a pair of lotteries which are strictly ranked for player i and for which other players have the (weakly) opposite ranking. In addition this pair of lotteries must be chosen measurably with respect to Ψ^*.

Theorem. *A social choice function is virtually implementable in iteratively undominated strategies if and only if it satisfies self-selection and is measurable with respect to Ψ^*.*

Proof. As noted above self-selection is obviously necessary. By Proposition 1, so is measurability. Hence, it only remains to establish the 'if' part of the theorem.

By Lemma 2 there exists a social choice function $\widetilde{x}$ which is strictly measurable with respect to Ψ^* and exactly implementable in iteratively undominated strategies. Let $\widetilde{G} = (\widetilde{M}, \widetilde{g})$ be the implementing game form constructed in the proof of Lemma 2, where $\widetilde{M}_i = \Psi_i^*$ for all $i \in N$, $\widetilde{g} = \widetilde{x}^+$, and γ^* is the unique iteratively undominated strategy profile of $\widetilde{G}$. Since γ^* is a strict Bayesian Nash equilibrium in $\widetilde{G}$, $\widetilde{x}$ satisfies strict self-selection for all players with respect to Ψ^*. Let

$$y(s) = (1-\alpha)\, x(s) + \alpha \widetilde{x}(s).$$

Since x satisfies self-selection and $\widetilde{x}$ satisfies strict self-selection for all players with respect to Ψ^*, y^+ satisfies strict self-selection for all players with respect to Ψ^* also, for any $\alpha \in (0,1]$. We will assume without loss of generality that x^+ satisfies strict self-selection for all players with respect to Ψ^* since there exist arbitrarily close social choice functions which do.

The game form $G = (M, g)$ we construct here is similar to the one for complete information. Specifically,

$$M_i = M_i^0 \times M_i^1 \times ... \times M_i^K = \Psi_i^* \times ... \times \Psi_i^*$$

for an integer K to be defined.

Define the function $\xi : N \times M \to A$ as follows:

$$\begin{array}{ll} \xi(i,m) = \underline{a}(i,m^0) & \text{if there exists } k \in \{1,...,K\} \text{ such that } m_i^K \neq m_i^0, \\ & \text{and } m^h = m^0 \text{ for all } h \in \{1,...,K-1\}, \\ \xi(i,m) = \overline{a}(i,m^0) & \text{otherwise,} \end{array}$$

where $\overline{a}$ and $\underline{a}$ are as in Assumption 2.

The outcome function $g : M \to A$ is

$$g(m) = \varepsilon \widetilde{x}^+ \left(m^0\right) + \frac{\varepsilon^2}{n} \sum_{i \in N} \xi(i,m) + \left(1 - \varepsilon - \varepsilon^2\right) \frac{1}{K} \sum_{h=1}^{K} x^+(m^h).$$

Define $\sigma_i^* = (\gamma_i^*, ..., \gamma_i^*)$, and let $\sigma^* = (\sigma_i^*)_{i \in N}$. We show that for small enough ε, σ^* is the unique iteratively undominated strategy profile in the game form G.

We first argue that if σ is iteratively undominated in $G = (M, g)$, then $\sigma^0 = \gamma^*$. For every $i \in N$, every σ and every σ_i', if $\sigma_i'^h = \sigma_i^h$ for all $h \in \{1, ..., K\}$, then

$$\begin{aligned} v_i(G, \sigma/\sigma_i', s_i) - v_i(G, \sigma, s_i) &= \varepsilon \left(v_i(\widetilde{G}, \sigma^0/\sigma_i'^0, s_i) - v_i(\widetilde{G}, \sigma^0, s_i) \right) \\ &\quad + \frac{\varepsilon^2}{n} \sum_{s_{-i} \in S_{-i}} \sum_{j \in N} \left(u_i\left(\xi\left(j, \sigma/\sigma_i'(s)\right), s\right) - u_i\left(\xi\left(j, \sigma(s)\right), s\right)\right) p_i\left(s_{-i} | s_i\right) \\ &\geq \varepsilon \left(v_i(\widetilde{G}, \sigma^0/\sigma_i'^0, s_i) - v_i(\widetilde{G}, \sigma^0, s_i) \right) - \frac{2\varepsilon^2}{n} E_i. \end{aligned}$$

For the mechanism $\widetilde{G}$, let $(\widetilde{Q}_i^h)_{i \in N'}$, $h = 1, 2, ...$, be a sequence of sets of iteratively undominated strategies as defined in Section 3, and let $(Q_i^h)_{i \in N'}$, $h = 1, 2, ...$, be the corresponding sequence for G. It follows from definitions that there exists a positive number $\eta > 0$ such that for every $i \in N$, $h \in \{0, 1, ...\}$ and $\theta_i \in \widetilde{Q}_i^h$, if θ_i is dominated with respect to $\widetilde{Q}^h$, then there exists $s_i \in S_i$ and $\theta_i' \in \widetilde{Q}_i^h$ such that for all $\theta_{-i} \in \widetilde{Q}_{-i}^h$,

$$v_i(\widetilde{G}, \theta/\theta_i', s_i) - v_i(\widetilde{G}, \theta, s_i) > \eta.$$

For every $i \in N$, let

$$E_i = \max_{s \in S, m \in M} \left\{ \sum_{jseN} \left| u_i\left(\xi\left(j, m\right), s\right) \right| \right\}.$$

Choose $\overline{\varepsilon} > 0$ such that

$$\eta > 2\overline{\varepsilon} E_i \text{ for all } i \in N, \tag{$*$}$$

and assume below that $0 < \varepsilon \leq \overline{\varepsilon}$. Then,

$$v_i(G, \sigma/\sigma_i', s_i) - v_i(G, \sigma, s_i) \geq \varepsilon \left(v_i(\widetilde{G}, \sigma^0/\sigma_i'^0, s_i) - v_i(\widetilde{G}, \sigma^0, s_i) - \eta \right).$$

It follows that if σ is an element of Q^1, then σ^0 must be an element of $\widetilde{Q}^1$, and recursively, for every $h \in \{2, ...\}$, if σ is an element of Q^h, then σ^0 must be an element of $\widetilde{Q}^h$. Consequently, if σ is iteratively undominated in G, then $\sigma^0 = \gamma^*$.

For every $i \in N$ and every $s \in S$, define

$$\begin{aligned} B_i(s) &= u_i(\overline{a}(i, \gamma^*(s)), s) - u_i(\underline{a}(i, \gamma^*(s)), s), \\ D_i(s) &= \max_{s' \in S} \{u_i(x(s), s) - u_i(x(s/s_i'), s) + u_i(x(s'), s) - u_i(x(s'/s_i), s)\}. \end{aligned}$$

By Assumption 2, for every $i \in N$ and every $s \in S$,

$$B_i(s) > 0 \text{ if } p_i(s_{-i} | s_i) > 0.$$

Hence, there exists a positive integer K such that for every $i \in N$ and every $s \in S$,

$$K \frac{\varepsilon^2}{n} B_i(s) > \left(1 - \varepsilon - \varepsilon^2\right) D_i(s) \text{ if } p_i(s_{-i} | s_i) > 0. \tag{$**$}$$

Let $P(h)$ be the statement: "if σ is iteratively undominated in G, then for every $i \in N$,

$$\sigma_i^q = \gamma_i^* \text{ for all } q \in \{0, ..., h\}."$$

We have established $P(0)$. We now show that for all $h = 1, ..., K$, $P(h-1) \Rightarrow P(h)$.

Suppose $P(h-1)$, and consider an iteratively undominated strategy σ_i for player i. Then, by $P(h-1)$,

$$\sigma_i^q = \gamma_i^* \text{ for all } q \in \{0, ..., h-1\}.$$

We need to show that $\sigma_i^h = \gamma_i^*$ also. Suppose not, and let $\sigma_i^h(s_i) = \gamma_i^*(s_i') \neq \gamma_i^*(s_i)$ for some $s_i' \in S_i$. Let σ_i' be the strategy for player i such that

$$\sigma_i'^q = \sigma_i^q \text{ for all } q \neq h, \text{ and } \sigma_i'^h(s_i) = \gamma_i^*(s_i).$$

Consider any iteratively undominated strategy profile σ_{-i} for other players. Then $\sigma_j^q = \gamma_j^*$ for all $q \in \{0, ..., h-1\}$ and all $j \in N \backslash \{i\}$. Let

$$S_{-i}^* = \left\{s_{-i} \in S_{-i} : \sigma_{-i}^h(s_{-i}) \neq \gamma_{-i}^*(s_{-i})\right\}.$$

Then, by (**),

$$
\begin{aligned}
v_i(G,\sigma/\sigma_i',s_i)-v_i(G,\sigma,s_i) \ \ge\ & \sum_{s_{-i}\notin S^*_{-i}} \frac{1}{K}\left(1-\varepsilon-\varepsilon^2\right)\left(u_i\left(x\left(s\right),s\right)-u_i\left(x\left(s/s_i'\right),s\right)\right)p_i\left(s_{-i}\middle|\, s_i\right) \\
& + \sum_{s_{-i}\notin S^*_{-i}} \frac{1}{K}\left(1-\varepsilon-\varepsilon^2\right)\Big[u_i(x^+(\sigma^h\left(s\right)/\gamma_i^*\left(s_i\right)),s) \\
& -u_i(x^+(\sigma^h\left(s\right)),s)+\frac{\varepsilon^2}{n}B_i\left(s\right)\Big]\,p_i\left(s_{-i}\middle|\, s_i\right) \\
>\ & \frac{1}{K}\left(1-\varepsilon-\varepsilon^2\right)\left(U_i\left(x,s_i\right)-V_i\left(x,s_i,s_i'\right)\right) \\
& +\frac{1}{K}\sum_{s_{-i}\in S^*_{-i}}\left(K\frac{\varepsilon^2}{n}B_i\left(s\right)-\left(1-\varepsilon-\varepsilon^2\right)D_i\left(s\right)\right)p_i\left(s_{-i}\middle|\, s_i\right) \\
>\ & 0.
\end{aligned}
$$

The extra subtlety here, which does not appear in the complete information case, is that some types of other players (this is the set S^*_{-i}) may misrepresent m_i^h while others do not.

This inequality implies that σ_i is not iteratively undominated in G, a contradiction. Hence, $\sigma_i^h\left(s_i\right)=\gamma_i^*\left(s_i\right)$, $\sigma^h=\gamma^*$, and the unique iteratively undominated strategy profile is $\sigma^*=(\gamma^*,...,\gamma^*)$.

Since

$$
g\left(\sigma^*\left(s\right)\right)=\left(1-\varepsilon-\varepsilon^2\right)x^+\left(\gamma^*\left(s\right)\right)+\varepsilon\widetilde{x}^+\left(\gamma^*\left(s\right)\right)+\frac{\varepsilon^2}{n}\sum_{i\in N}\overline{a}\left(i,\gamma^*\left(s\right)\right),
$$

and ε can be taken to be arbitrarily small, the proof is complete. ■

6 Complete information and the two-player case

Our incomplete information framework is very general and in particular incorporates the special case of complete information. Furthermore out result is independent of the number of players and in particular covers the two player case, which traditionally has been treated separately (or not at all) with equilibrium-based implementation literature. We spell out here how our theorem specializes to the complete information case with two or more players.

In general, some signals $s=(s_1,...,s_n)\in S$ may occur with zero probability. This is, for instance, the case in the complete information setting where all players receive the same signal – that is, all probability mass lies on the "diagonal" of S (i.e., $s \in S$ such that $s_i = s_1$ for all $i \in N$).

The self-selection condition is expressed in terms of social choice functions which are defined for all $s \in S$. Of course, from the point of view of final realized equilibrium outcomes, any pair of social choice functions which agree on $\overline{S}$, the support of S, are equivalent. Starting from a social choice function defined on $\overline{S}$ we would require that the self-selection condition be satisfied for some extension of the function to S, the Cartesian product of the individual signal space S_i.

In our paper on complete information AM [1], it was convenient to define the social choice function on $\overline{S}$, i.e., the diagonal of S. With three or more players such a social choice function may be extended to S in a manner in which the self-selection condition is automatically satisfied; the extension simply ignores a single player deviation from an otherwise unanimous announcement.

In terms of the notation of this paper the complete information case may be described as follows.

$$
\begin{array}{ll}
S_i = S_1 & \text{for all } i \in N, \\
p_i\left(s_{-i} \mid s_i\right) = 1 & \text{if } s_j = s_i \text{ for all } j \in N \backslash \{i\} \\
p_i\left(s_{-i} \mid s_i\right) = 0 & \text{otherwise} \\
\overline{S} = \{(s_1, ..., s_n) \in S : s_i = s_1 \text{ for all } i \in N\} &
\end{array}
$$

For any $s_i \in S_i$, let $s_i \cdot e$ denote $t \in \overline{S}$ for which $t_j = s_i$ for all $j \in N$. Each $s_i \in S_i$ corresponds to a preference profile, one for each player. Each S_i may be partitioned into subsets $\phi_i(s_i)$ where each element of $\phi_i(s_i)$ yields the same preferences over lotteries for player i. Furthermore, for all $s_i \in S_i$, $\{s_i\} = \cap_{j \in N} \left\{\phi_j(s_i)\right\}$.

We first argue that measurability is trivial in complete information environments. In fact, this follows from our earlier paper (at least when there are three or more players) which shows that any social choice function is virtually implementable in the complete information case. Since measurability is a necessary condition, it follows that with complete information any (three or more player) social choice function is measurable. We sketch a direct argument below (which also covers the two-player case).

Specialized to the complete information case Assumption 1 amounts to ruling out player types who are indifferent over all lotteries. As we show in AM [1] this implies that there exists a function $f_i : S_i \to A$ such that for every $s_i \in S_i$,

$$f_i(s_i) = f_i(s_i') \text{ for all } s_i' \in \phi_i(s_i),$$

and

$$u_i(f_i(s_i), s_i) > u_i(f_i(s_i'), s_i) \text{ for all } s_i' \notin \phi_i(s_i).$$

By definition,

$$\Psi_i^0 = \{S_i\}.$$

By considering constant social choice function $x_{t_i}^1 : S \to A$ where

$$x_{t_i}^1(z) = f_i(t_i) \text{ for all } z \in S,$$

the reader may check that $R^i(\Psi_{-i}^0) = \{\phi_i(s_i) : s_i \in S_i\}$ (for any $w_i \notin \phi_i(s_i)$, consider the pair of constant social choice functions $x_{s_i}^1$ and $x_{w_i}^1$). Hence

$$\Psi_i^1 = \{\phi_i(s_i) : s_i \in S_i\}.$$

Finally the second round of the iterative procedure yields

$$\Psi_i^2 = \overline{\Psi}_i = \{\{s_i\} : s_i \in S_i\},$$

the finest possible partition. Hence $\Psi_i^* = \Psi_i^2$, and the iterative procedure terminates in the second round. Let $\overline{b}_i(\phi_i(s_i))$ and $\underline{b}_i(\phi_i(s_i))$ satisfy

$$u_i(\overline{b}_i(\phi_i(s_i)), s_i) > u_i(\underline{b}_i(\phi_i(s_i)), s_i).$$

The existence of such lotteries is guaranteed by Assumption 1 (non-indifference over all lotteries). In distinguishing between any pair of signals w_i and s_i we may now use social choice functions which are measurable with respect to $\{S_i\} \times \Psi_{-i}^1$. Define $g_i : S \to A$ by

$$\begin{aligned} g_i(s) &= \overline{b}_i(\phi_i(s_i)) \text{ if } \phi_j(s_j) = \phi_j(s_i) \text{ for all } j \in N, \\ &= \underline{b}_i(\phi_i(s_i)) \text{ otherwise.} \end{aligned}$$

Let $x_{t_i}^2 : S_i \to A$ be defined by

$$x_{t_i}^2(s) = g_i(t_i, s_{-i}) \text{ for all } s \in S.$$

Now to distinguish between w_i and s_i where $w_i \neq s_i$ and $w_i \in \phi_i(s_i)$, consider the social choice functions $x_{s_i}^2$ and $x_{w_i}^2$:

$$U_i(x_{s_i}^2, s_i) = u_i(\overline{b}_i(s_i), s_i) > u_i(\underline{b}_i(s_i), s_i) = U_i(x_{w_i}^2, s_i),$$

and conversely when the signal is w_i.[7]

[7] We remark that since measurability is trivially satisfied in complete information environments whereas not all social choice functions are monotonic in such environments, it is clear that the measurability condition does not imply Bayesian monotonicity.

Let us turn to the self-selection condition. As noted above this condition is satisfied trivially when there are three or more players. Let $\overline{x} : \overline{S} \rightarrow A$ be a social choice function as in the earlier paper on complete information. It may be extended to yield an equivalent (on the support $\overline{S}$ of S) social choice function $x : S \rightarrow A$ which satisfies self-selection as follows:

$$\begin{array}{ll} x(s) = \overline{x}(s_j \cdot e) & \text{if there exists a subset } J \text{ of } N \text{ such that} \\ & |J| \geq n-1,\ j \in J \text{ and } s_i = s_j \text{ for all } i \in J, \\ x(s) = b & \text{otherwise, where } b \text{ is an arbitrary element of } A. \end{array}$$

This extension is not well defined if $n < 3$. Indeed in the two-player case, self-selection is a non-trivial requirement, and is equivalent to a condition termed the <u>intersection</u> property.

Definition 6 *A two-person social choice function $\overline{x} : \overline{S} \rightarrow A$ satisfies the <u>intersection</u> property if for every $s_1 \in S_1$ and $s_2 \in S_2$ there exists $b \in A$ such that*

$$u_1(\overline{x}(s_2 \cdot e), s_2 \cdot e) \geq u_1(b, s_2 \cdot e),$$

and

$$u_2(\overline{x}(s_1 \cdot e), s_1 \cdot e) \geq u_2(b, s_1 \cdot e).$$

When this property is satisfied the set of lotteries which is (weakly) worse than $\overline{x}(s_j \cdot e)$ for player i with preferences $\gamma_i(s_j)$ has a non-empty <u>intersection</u> with the corresponding set for player j. For a social choice function $\overline{x}$ which satisfies the intersection property and each $s_1 \in S_1$ and $s_2 \in S_2$ let $b(s_1, s_2)$ satisfy the inequalities of Definition 6. Define an equivalent extended social choice function $x : S \rightarrow A$ by

$$\begin{array}{ll} x(s_1, s_2) = \overline{x}(s_1) & \text{if } s_1 = s_2, \\ x(s_1, s_2) = b(s_1, s_2) & \text{if } s_1 \neq s_2. \end{array}$$

Then x satisfies self-selection. It is also clear that $\overline{x}$ must satisfy the intersection property if some extension of it satisfies self-selection. We remark that the existence of a "holocaust" outcome is a crude sufficient condition for the intersection property to hold.

It follows immediately from out theorem and the preceding discussion that in the complete information case, any two-person social choice function which satisfies the intersection property is virtually implementable in iteratively undominated strategies.

Abreu and Sen [2] showed that the intersection property is necessary (and sufficient) for virtual implementation in <u>Nash equilibrium</u>. We note that the intersection property is much weaker than the complex necessary and sufficient conditions for <u>exact</u> two-person implementation in Nash equilibrium provided by Dutta and Sen [3] and Moore and Repullo [13].

7 Conclusion

A large literature following the Gibbard-Satterthwaite theorem has sought to characterize implementable social choice functions for a variety of concepts weaker than dominant strategy implementation. Under general informational assumptions, and in the context of virtual implementation and the iterative elimination of strictly dominated strategies this paper provides a characterization which is essentially complete. We present two necessary conditions, self-selection and measurability, which (under weak domain restrictions) are also sufficient. The former is obvious and well-known, and the latter is introduced in the present paper. Measurability is frequently automatically satisfied. Modulo measurability, our result is therefore as permissive as one could possibly expect, and furthermore obtained for a solution concept which is even weaker than rationalizability.

We are aware of no characterizations involving non-equilibrium solution concepts in general Bayesian environments. Earlier results for implementation in Bayesian Nash equilibrium are limited by the stringency of the necessary condition (Bayesian monotonicity), the restriction to three or more players, the (undesirable) equilibrium basis of the solution concept, and above all by the unsatisfactory nature of the implementing mechanisms.

References

[1] Abreu, D. and H. Matsushima, "Virtual implementation in Iteratively Undominated Strategies: Complete Information". *Econometrica* 60(5), September 1992, pp. 993-1008.

[2] Abreu, D., and A. Sen, "Virtual Implementation in Nash Equilibrium", *Econometrica* 59(4), July 1991, pp. 997-1021.

[3] Dutta, B. and A. Sen, "Necessary and Sufficient Conditions for 2-person Nash Implementation". *Review of Economic Studies* 58(1), January 1991, pp. 121-128.

[4] Gibbard, A., "Manipulation of Voting Schemes: A General Result". *Econometrica* 41(4), July 1973, pp. 587-601.

[5] Good, I. J., "Rational Decisions", *Journal of the Royal Statistical Society*, *Series B* 14(1), 1952, pp. 107-114.

[6] Jackson, M., "Bayesian Implementation". *Econometrica* 59(2), March 1991, pp. 461-477.

[7] Jackson, M., "Implementation in Undominated Strategies: A Look at Bounded Mechanisms". *Review of Economic Studies* 59(4), October 1992, pp. 757-775.

[8] Matsushima, H., "A New Approach to the Implementation Problem". *Journal of Economic Theory* 45, 1988, pp. 128-144.

[9] Matsushima, H., "Characterization of Full Bayesian Implementation", mimeo, Stanford University, 1990.

[10] Matsushima, H., "Incentive Compatible Mechanisms with Full Transferability". *Journal of Economic Theory* 50, 1991.

[11] Matsushima, H., "Bayesian Monotonicity with Side Payment". *Journal of Economic Theory 59*, 1993.

[12] Mookherjee, D. and S. Reichelstein, "Implementation via Augmented Revelation Mechanisms". *Review of Economic Studies* 57(3), July 1990, pp. 453-475.

[13] Moore, J., and R. Repullo, "Subgame Perfect Implementation". *Econometrica* 56(5), September 1988, pp. 1191-1220.

[14] Palfrey, T., "Implementation in Bayesian Equilibrium: The Multiple Equilibrium Problem in Mechanism Design". In *Advances in Economic Theory Vol. 1*, J.-J. Laffont ed., Cambridge: Cambridge University Press, 1992.

[15] Palfrey, T. and S. Srivastava, "Implementation with Incomplete Information in Exchange Economies". *Econometrica* 57(1), January 1989, pp. 115-134.

[16] Palfrey, T. and S. Srivastava, "Mechanism Design with Incomplete Information: A Solution to the Implementation Problem". *Journal of Political Economy* 97(3), June 1989, pp. 668-691.

[17] Postlewaite, A., and D. Schmeidler, "Implementation in Differential Information Economies". *Journal of Economic Theory* 39(1), June 1986, pp. 14-33.

[18] Satterthwaite, M., "Strategy-proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Choice Functions". *Journal of Economic Theory* 10, April 1975, pp. 187-217.

[19] Winkler, R. L., "The Quantification of Judgement: Some Methodological Suggestions". *Journal of the American Statistical Association* 62(320), December 1967, pp. 1105-1120.