Type Diversity and Virtual Bayesian Implementation*

Roberto Serrano† and Rajiv Vohra‡

This version: August 2001

http://econ.pstc.brown.edu/~rvohra/papers/wp00-16.pdf

Abstract. It is well known that a social choice function is truthfully implementable in Bayesian Nash equilibrium if and only if it is incentive compatible. However, in general it is not possible to rule out other equilibrium outcomes, and additional conditions, e.g., Bayesian monotonicity, are needed to ensure full implementation. We argue that this multiplicity problem is not very severe for virtual Bayesian implementation. We show that any incentive compatible social choice function is virtually (and fully) Bayesian implementable in an environment satisfying a condition we term *type diversity*. If there are at least three elements in the set of social alternatives, this condition holds generically. Type diversity is a condition with a simple economic interpretation compared to Bayesian monotonicity, and the mechanism we construct is also simple. We also provide a necessary and sufficient condition – *type diversity with respect to deceptions* – for virtual Bayesian implementation in economic environments.

JEL Classification: C72, D78, D82.

Key Words: Type diversity, virtual Bayesian implementation, Bayesian implementation, incentive compatibility, incomplete information.

*We acknowledge numerous conversations with O. Volij on the subject. We also thank S. Baliga, J. Bergin, L. Corchon, M. Jackson, G. Mailath, T. Palfrey and P. Reny for helpful comments. Seminar audiences at Venice, Pompeu Fabra, Stanford, Arizona, Carlos III, Brown, Jerusalem, Minnesota, Caltech and Maryland also had useful suggestions. Both authors acknowledge Salomon research awards from Brown University. Serrano also acknowledges research support from the Alfred P. Sloan Foundation.

†Department of Economics, Brown University, Providence, RI 02912, U.S.A. roberto_serrano@brown.edu; http://econ.pstc.brown.edu/faculty/serrano.

‡Department of Economics, Brown University, Providence, RI 02912, U.S.A. and Department of Economics, New York University, 269 Mercer Street, NY 10003; rajiv_vohra@brown.edu; http://econ.pstc.brown.edu/~rvohra.

1 Introduction

There is now an extensive literature, beginning with Postlewaite and Schmeidler (1986), which studies the problem of Bayesian implementation in environments with incomplete information; see, for example, Palfrey and Srivastava (1987, 1989a), Mookherjee and Reichelstein (1990), and Jackson (1991). The importance of Bayesian incentive compatibility in this context is well known. This condition is, by the revelation principle, both necessary and sufficient for truthful Bayesian implementation, i.e., finding a mechanism where truth-telling is an equilibrium.[1] It is for *full implementation*, requiring the set of equilibrium outcomes of the mechanism to coincide with the social choice set, that additional conditions become important. This is the well-known problem of multiplicity of equilibria, and resolving it is our main concern in this paper. We will argue that Bayesian incentive compatibility is the only important restriction on a social choice function for full implementation provided one accepts two kinds of approximations: (a) requiring virtual instead of exact implementation, and (b) perturbing the environment, if necessary, to ensure a condition we term *type diversity*.

Postlewaite and Schmeidler (1986) showed that a necessary condition (in addition to incentive compatibility) for a social choice set to be Bayesian implementable is Bayesian monotonicity. As the term Bayesian monotonicity suggests, this condition can be seen as an analog of Maskin monotonicity (Maskin (1977)) in the presence of incomplete information.[2] Palfrey and Srivastava (1989a) found a weakening of incentive compatibility and a variant of Bayesian monotonicity that turned out to be sufficient for implementation in exchange economies. The gap after their work between necessary and sufficient conditions was closed by Jackson (1991) with a strengthening of Bayesian monotonicity.[3] Unfortunately, Bayesian monotonicity is not satisfied by any of the well-known social choice functions (SCFs) for exchange

[1] see, for example, Myerson (1987) and the references therein.

[2] Recall that Maskin monotonicity is a necessary condition for Nash implementation. It also turns out to be sufficient in environments where there is a private good and at least three agents. This condition is satisfied by many correspondences of interest in exchange economies (such as the Pareto, core and constrained Walrasian correspondences). However, it may be quite restrictive in other domains (see, for example, Mueller and Satterthwaite (1977) and Saijo (1987)).

[3] Jackson (1991) also provides sufficient conditions that guarantee implementation outside of economic environments. He identifies a condition that he terms "monotonicity no veto" that serves this purpose.

economies with incomplete information; see Palfrey and Srivastava (1987), Chakravorti (1992) and Serrano and Vohra (2001).

There is another sense in which the complete information environment seems to yield more permissive implementation results. Remarkably, the Maskin monotonicity condition can be entirely dispensed with by slightly weakening the notion of implementation. This is the main insight of Abreu and Sen (1991) and Matsushima (1988), who show that under very mild conditions, any social choice correspondence can be virtually Nash implemented in the sense that, making use of lotteries over social alternatives, it is possible to exactly implement an SCF that is arbitrarily close to the given correspondence. See Figure 1, drawn for the case of three social alternatives: since lotteries are allowed in the mechanism, any random outcome x in the interior of the probability simplex over alternatives satisfies Maskin monotonicity because the lower contour sets for u and u' through x are not nested.[4] Moreover, Abreu and Matsushima (1992a) provide a significant improvement of these results by showing that under very weak conditions any SCF can be virtually implemented in the more attractive notion of iteratively undominated strategies, and this is possible without the use of mechanisms involving integer games.

Given the power of the virtual approach in the complete information case, and given that Bayesian monotonicity is often a very strong condition, it is natural to ask if one can find simpler and/or weaker conditions for virtual implementation in the presence of incomplete information. That some condition (in addition to incentive compatibility) is needed even for virtual Bayesian implementation is clear from Example 1 in Serrano and Vohra (2001); there are environments with incomplete information in which only constant SCFs can be virtually implemented. Thus, in contrast to the complete information results, even virtual implementation requires non-trivial restrictions either on the environment or the SCF. Our aim here is to identify a simple, weak, and readily interpretable condition for virtual Bayesian implementation of any incentive compatible SCF.

[4] Recall that Maskin monotonicity requires the following. Suppose that outcome x is socially desirable when the true profile of utilities is u, and that in going to another profile u', the lower contour set through x of every agent at u' contains the one when utilities were u. Then, x should remain socially desirable when utilities are u'.

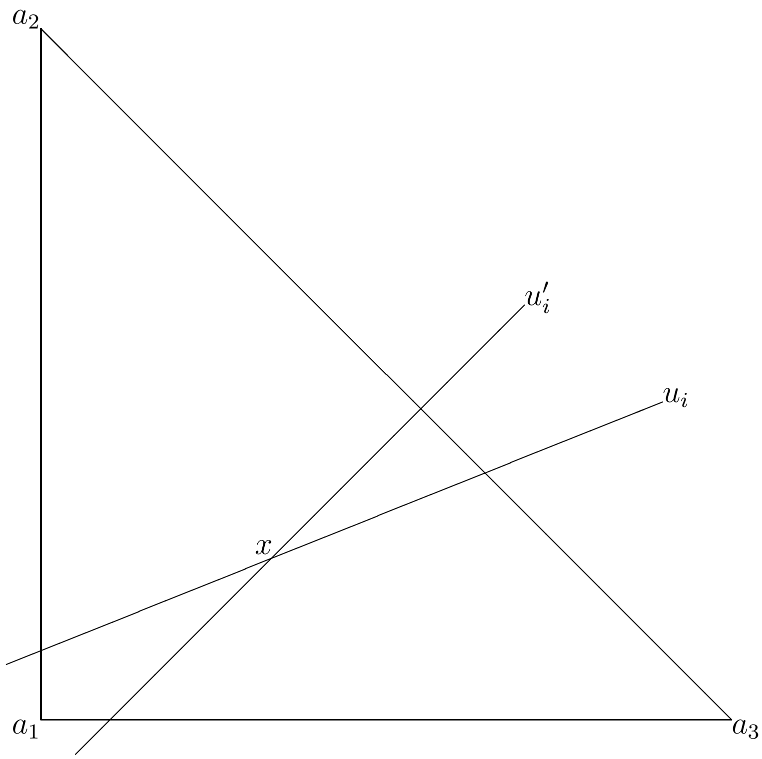

Figure 1: Virtual Nash Implementation

Abreu and Matsushima (1992b) dispense with Bayesian monotonicity and introduce a new condition termed *measurability* (henceforth A-M measurability) which, under other weak assumptions, is shown to be necessary and sufficient for virtual implementation in iteratively undominated strategies. Their sufficiency result applies, a fortiori, to the notion of virtual implementation in Bayesian Nash equilibrium. Duggan (1997) uses the condition of *incentive consistency* and presents a sufficiency result for environments with at least three agents and "best-element private values."[5] Serrano and Vohra (2001) criticize A-M measurability and incentive consistency, by showing them to be strong conditions in certain environments (sometimes, even

[5]In the same endeavor of attempting to dispense with Bayesian monotonicity, Matsushima (1993) shows that this can be done if side payments are allowed.

stronger than Bayesian monotonicity). Indeed, there are environments where every SCF is virtually Bayesian implementable, but only constant SCFs satisfy A-M measurability or incentive consistency.

The first contribution of this paper is the identification of a generic condition on environments, *type diversity,* that makes full implementation with incomplete information rely *solely* on incentive compatibility. Indeed, we shall argue that the difficulty with virtual Bayesian implementation is rare. Type diversity turns out to be generic in the set of all environments with at least three alternatives, and in its presence, virtual implementation is as successful as it can be. Our first main result shows that, in environments satisfying type diversity, *any* incentive compatible SCF is virtually Bayesian implementable (even with two agents). While type diversity is a condition on the domain, and A-M measurability and incentive consistency are conditions on the SCF, type diversity is considerably easier to check and its interpretation is straightforward: it yields a simple single crossing property by requiring that the interim (cardinal) preferences over pure alternatives of different types of an agent be different. In a private values model, it reduces to the condition of value-distinguished types introduced in Palfrey and Srivastava (1989b). Furthermore, in environments satisfying type diversity, every SCF satisfies A-M measurability and incentive consistency.[6] Thus, in environments satisfying type diversity, there is no need to impose measurability or incentive consistency on the SCF. In environments violating type diversity, the two conditions are, however, stronger than necessary.

The main advantage of assuming type diversity is simplicity. At the expense of some simplicity, and inspired by the Bayesian monotonicity condition stated in Jackson (1991), we also provide a characterization of virtual Bayesian implementation for economic environments. This characterization also includes the case of two agents and can be expressed in two forms. That is, the condition of *type diversity with respect to deceptions*, a weakening of the simple type diversity condition, can be imposed on the SCF or the economic environment to obtain a full characterization of virtual Bayesian implementation in these settings.[7] For each deception, the condition assumes the existence of a certain preference reversal for at least one type of

[6] However, our result is not implied by those of Abreu and Matsushima (1992b) or Duggan (1997), since they rely on extra conditions.

[7] Note that type diversity with respect to deceptions is a weaker condition than Bayesian monotonicity, A-M measurability or incentive consistency; these three conditions are not necessary for virtual implementation.

one agent. Thus, without appealing to genericity arguments, we show that in economic environments an SCF is virtually Bayesian implementable if and only if it satisfies incentive compatibility and type diversity with respect to deceptions. Equivalently, the same proof demonstrates that (1) if an economic environment does not satisfy type diversity with respect to deceptions, virtual Bayesian implementation is severely limited; and (2) if it does, any incentive compatible SCF is virtually Bayesian implementable.

In sum, this paper shows that the problem of multiplicity of equilibrium in mechanism design under incomplete information can be completely solved if one takes two degrees of approximation: (a) in the solution concept, by requiring virtual instead of exact implementation, and (b) in the environments, by perturbing them if necessary to ensure type diversity. In doing so, Bayesian incentive compatibility remains the only important restriction on an SCF for full implementation. The simplicity of the type diversity condition also gives us hope for its fruitful applications.

2 The Model and Definitions

We shall consider implementation in the context of a general environment with asymmetric information. Let $N = \{1, \ldots, n\}$ be the finite set of agents. Let T_i denote the (finite) set of agent i's types. The interpretation is that $t_i \in T_i$ denotes the *private information* possessed by agent i. We refer to a profile of types $t = (t_1, \ldots, t_n)$ as a state. Let $T = \prod_{i \in N} T_i$ be the set of states. We will use the notation t_{-i} to denote $(t_j)_{j \neq i}$. Similarly $T_{-i} = \prod_{j \neq i} T_j$.

Each agent has a prior probability distribution q_i defined on T. We assume that for every $i \in N$ and $t_i \in T_i$, there exists $t_{-i} \in T_{-i}$ such that $q_i(t) > 0$. For each $i \in N$ and $\bar{t}_i \in T_i$, the conditional probability of $t_{-i} \in T_{-i}$, given $\bar{t}_i$ is denoted $q_i(t_{-i} \mid \bar{t}_i)$. We shall assume that all agents agree on zero probability states. Let $T^* \subseteq T$ be the set of states with positive probability.

Let A denote the set of social alternatives, which are assumed to be independent of the information state. For example, in the context of an exchange economy, the set A refers to all possible redistributions of a (constant across states) aggregate endowment. Let $\mathcal{A}$ denote the Borel σ-algebra on A and $\triangle$ denote the set of probability measures on $(A, \mathcal{A})$.

A *social choice function* (SCF) is a function $f : T \mapsto \triangle$. Two SCFs, f and h are equivalent if $f(s) = h(s)$ for every $s \in T^*$ (see Jackson (1991) for a discussion on equivalent SCFs). We shall concentrate on SCFs rather than social choice sets because our main interest lies in virtual implementation making use of lotteries over A; a social choice set can be understood as a random function that puts positive measure only on the functions that it includes.

The Bernoulli utility of agent i for alternative a in state s is $u_i(a, s)$. Abusing notation slightly, given an SCF f, $u_i(f, s)$ will refer to agent i's expected utility evaluation of lottery $f(s)$ in state s. The (interim/conditional) expected utility of agent i of type s_i corresponding to an SCF f is defined as:

$$U_i(f|s_i) \equiv \sum_{s'_{-i} \in T_{-i}} q_i(s'_{-i}|s_i) u_i(f, (s'_{-i}, s_i)).$$

We can now define an *environment* as $\mathcal{E} = \{A, (u_i, T_i, q_i)_{i \in N}\}$.

A *mechanism* $G = ((M_i)_{i \in N}, g)$ describes a message space M_i for agent i and an outcome function $g : \prod_{i \in N} M_i \mapsto \triangle$.

A *Bayesian equilibrium* of G is a profile of messages, $(\bar{m}_i)$ where $m_i : T_i \mapsto M_i$ such that $\forall i \in N$, $\forall s_i \in T_i$,

$$U_i(g(\bar{m}(s))|s_i) \geq U_i(g(\bar{m}_{-i}(s_{-i}), m_i)|s_i) \quad \forall m_i \in M_i.$$

Denote by $B(G)$ the set of Bayesian equilibria of the mechanism G. An SCF f is *exactly* Bayesian implementable if there exists a mechanism G such that $g(B(G)) = f$.[8]

A *direct mechanism* is one with $M_i = T_i$ for all $i \in N$.

Consider the following metric on SCFs:

$$d(f, h) = \sup\{|f(B \mid s) - h(B \mid s)| \mid s \in T^*, B \in \mathcal{A}\}.$$

An SCF f is *virtually Bayesian implementable* if $\forall \epsilon > 0$ there exists a mechanism whose (unique) Bayesian equilibrium outcome coincides with an SCF h_ϵ such that $d(f, h_\epsilon) < \epsilon$.

A *deception* is a profile of functions, $\alpha = (\alpha_i)_{i \in N}$, where $\alpha_i : T_i \mapsto T_i$, $\alpha_i(t_i) \neq t_i$ for some $t_i \in T_i$ for some $i \in N$. (Note that the identity function on T will not be treated as a deception.) A deception is said to be *compatible*

[8]Exact implementation in environments with incomplete information has also been defined with respect to solution concepts other than Bayesian equilibrium, such as undominated Bayesian equilibrium (Palfrey and Srivastava (1989b)), perfect Bayesian equilibrium (Brusco (1995)), sequential equilibrium (Baliga (1999), Bergin and Sen (1998)). In each case, the definition of exact implementation requires the set of outcomes selected by the chosen solution concept in the mechanism to coincide with the social choice set.

if $\alpha(s) \in T^*$ for all $s \in T^*$. For an SCF f and a deception α, $f \circ \alpha$ denotes the SCF such that for each $s \in T$, $[f \circ \alpha](s) = f(\alpha(s))$. For an SCF f, a deception α and a type $s_i \in T_i$, let $f_{\alpha_i(s_i)}(s') = f(s'_{-i}, \alpha_i(s_i))$ for all $s' \in T$.

The next condition is necessary for exact Bayesian implementation (see Jackson (1991)).[9]

An SCF f satisfies *Bayesian monotonicity* if for any deception α, whenever $f \circ \alpha$ is not equivalent to f, there exist $i \in N$, $s_i \in T_i$ and an SCF h such that

$$U_i(h \circ \alpha \mid s_i) > U_i(f \circ \alpha \mid s_i) \text{ while } U_i(f \mid s'_i) \geq U_i(h_{\alpha_i(s_i)} \mid s'_i), \forall s'_i \in T_i.$$

An SCF, f, is said to be *incentive compatible* if for all $i \in N$, $s_i \in T_i$ and all deceptions α,

$$U_i(f \mid s_i) \geq U_i(f_{\alpha_i(s_i)} \mid s_i).$$

3 Type Diversity

We shall find it convenient in this Section to assume that the set of alternatives is finite; the reader is referred to Section 6 of Abreu and Sen (1991) for extensions to the case where A is an arbitrary subset of an abstract separable space.

Let $A = \{a_1, \ldots, a_K\}$ be the finite set of alternatives. Henceforth, we will find it convenient to identify a lottery, $x \in \triangle$, as a point in the unit simplex in R^K, i.e., x_k denotes the probability assigned by lottery x to alternative k.

Define $U_i^k(t_i)$ to be the interim utility of agent i of type t_i for the constant SCF which assigns a_k in each state, i.e.,

$$U_i^k(t_i) = \sum_{t_{-i} \in T_{-i}} q_i(t_{-i} \mid t_i) u_i(a_k, t).$$

Let $U_i(t_i) = (U_i^k(t_i))_{k=1,\ldots,K}$.

We will show that any incentive compatible SCF is virtually implementable in Bayesian Nash equilibrium if the environment satisfies the following condition:

An environment $\mathcal{E}$ satisfies *type diversity* if

[9] There is an extra condition termed *closure*, that requires the social choice set to be closed under concatenation of common knowledge events, but this is not too demanding. In any case, this condition will not be relevant in what follows since we will be dealing with social choice functions.

(i) there do not exist $i \in N$, $t_i, t_i' \in T_i$, $t_i \neq t_i'$ such that

$$U_i(t_i) = \alpha U_i(t_i') + \beta$$

for $\alpha > 0$.

(ii) for every $i \in N$ and $t_i \in T_i$, there exist $a_k, a_{k'}$ such that $U_i^k(t_i) \neq U_i^{k'}(t_i)$.

This condition has a simple interpretation: it requires (i) that the interim (cardinal) preferences over pure alternatives of different types of an agent be different, and (ii) that no type of any agent i is indifferent to *all* alternatives. Note that condition (i) does not require ordinal preferences over pure alternatives to differ across types unless $|A| = 2$. Moreover, the condition only concerns constant SCFs. In a private values model, this reduces to the condition that Palfrey and Srivastava (1989b) call *value-distinguished types*, but unlike their condition, it is fully operative regardless of the information structure, including environments with correlated and common values. Type diversity has the obvious virtue of being simple and easy to check, especially compared to other conditions in the literature, such as Bayesian monotonicity, A-M measurability or incentive consistency. Importantly, it is easy to see that in the space of preferences over pure alternatives, type diversity is satisfied *generically* if $|A| \geq 3$. In this sense type diversity is indeed a very weak condition if $|A| \geq 3$.[10] Yet it is remarkable that it suffices for virtual implementation.

The following Lemma provides a useful implication of type diversity from the point of view of implementation.

Lemma 1 *Suppose an environment $\mathcal{E}$ satisfies type diversity. Then there exist constant SCFs $((l_i(t_i))_{t_i \in T_i})_{i \in N}$ such that for every $i \in N$, $t_i, t_i' \in T_i$, $t_i \neq t_i'$,*

$$U_i(l_i(t_i) \mid t_i) > U_i(l_i(t_i') \mid t_i).$$

Proof. Consider the constant SCF $\bar{x}$, which prescribes in each state the lottery $\bar{x}$, assigning equal probability to each alternative in A, i.e., $\bar{x}(t) =$

[10] There is another reason why the weakness of condition (i) relies on there being at least 3 alternatives: if there are only 2 alternatives and an agent has more than 2 types then this condition cannot hold.

$(1/K, \ldots, 1/K))$ for all $t \in T$. We will show that for $i \in N$, $t_i, t_i' \in T_i$, $t_i \neq t_i'$, there exist constant SCFs x and x', close to $\bar{x}$, such that

$$U_i(x \mid t_i) > U_i(x' \mid t_i) \text{ and } U_i(x' \mid t_i) < U_i(x' \mid t_i'). \qquad (1)$$

The (interim) indifference curve of agent i of type t_i through $\bar{x}$ is described by a hyperplane, H, in R_+^{K-1}:

$$H = \{(x_1, \ldots x_{K-1}) \in R_+^{K-1} \mid \sum_{k=1}^{K-1} p_k(t_i) x_k = \bar{u}\},$$

where $p_k(t_i) = (U_i^k(t_i) - U_i^K(t_i))$, for $k = 1, \ldots K-1$. Consider the indifference hyperplane through $\bar{x}$ of agent i of type t_i' where $t_i' \neq t_i$:

$$H' = \{(x_1, \ldots x_{K-1}) \in R_+^{K-1} \mid \sum_{k=1}^{K-1} p_k(t_i') x_k = \bar{u}'\}.$$

Given no total indifference (condition (ii) of type diversity), we must have $p(t_i) \neq 0$ and $p(t_i') \neq 0$. Moreover, $p(t_i) \neq cp(t_i')$ for a positive number c as that would mean that $U_i(t_i) = cU_i(t_i') + \beta$, violating condition (i) of type diversity. Thus, either $p(t_i) = cp(t_i')$ where $c < 0$ or there does not exist $c \neq 0$ such that $p(t_i) = cp(t_i')$. In the former case, it is easy to see (using (ii)) that any point which lies above H must be below H' and by choosing two points (one above H and one below it) close to $\bar{x}$ one finds constant SCFs which satisfy (1). In the latter case, it is clear that we can choose two constant SCFs x and x' close to $\bar{x}$ satisfying (1).

Given (1) we can complete the proof by the same argument as in the Lemma in Abreu and Matsushima (1992a) or Lemma 1 in Abreu and Matsushima (1992b). ■

To illustrate the Lemma, see Figure 2, again drawn for the case of three pure alternatives, with alternative a_2 ranked above a_1, which in turn is ranked above a_3 (for all three types). Note how the picture is identical to Figure 1. However, its meaning is quite different. In particular, we are able to draw Figure 2 only because the condition is stated concerning preferences over constant SCFs. If an SCF is not constant, in principle the final outcome it prescribes is subject to deceptions, and an agent will find difficulties evaluating such SCFs because his Bernoulli utility or the final lottery prescribed change from state to state. Preferences over constant SCFs do not encounter this difficulty, and the surprising fact is that imposing a condition on preferences over constant SCFs alone turns out to be so powerful.

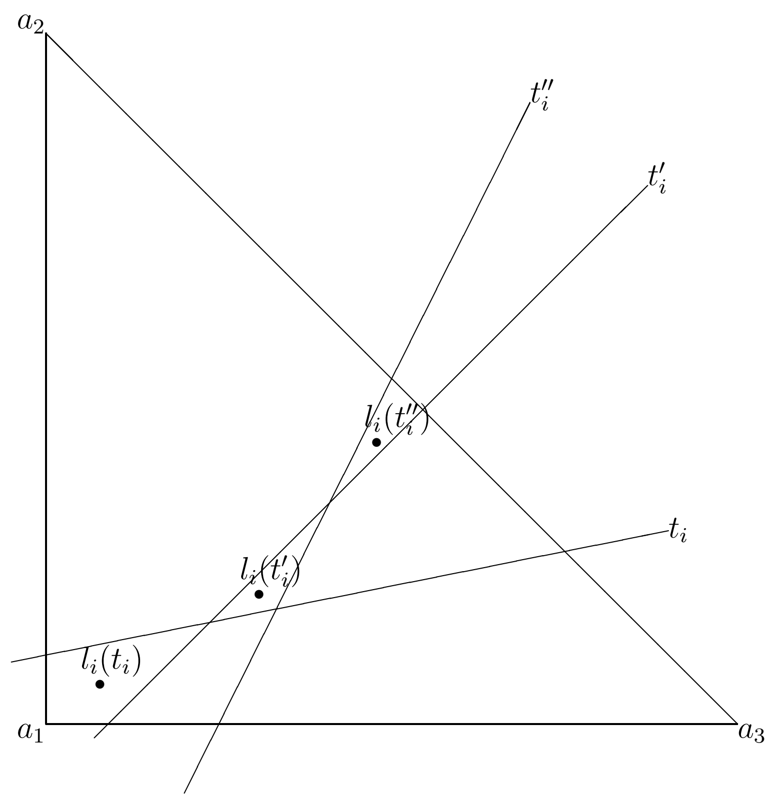

Figure 2: Type Diversity

Type diversity implies that every SCF is A-M measurable; see Abreu and Matsushima (1992b, Section 4.2), and also the related condition of interim value distinguished types in Palfrey and Srivastava (1993, definition 6.3).

It is also easy to see that if type diversity is satisfied, the SCF $f^* = \frac{\sum_{i \in N} l_i(t_i)}{n}$, where $l_i(t_i)$ satisfy the inequalities in the statement of Lemma 1, has the property that truth-telling is the only Bayesian equilibrium of the direct mechanism for f^*. This implies that, in the presence of type diversity, every SCF is incentive consistent, a condition which plays a crucial role in Duggan's (1997) sufficiency result.

While type diversity is not weaker than either AM-measurability or incentive consistency for every incentive compatible SCF, it is easier to interpret and is satisfied generically, as long as $|A| \geq 3$. Moreover, as we shall show below, our first result is not implied by the results of either Abreu and Matsushima (1992b) or Duggan (1997) because of the extra assumptions used by these authors. Note also that our result applies to environments with $n \geq 2$.

4 A Simple Positive Result

Theorem 1 *Suppose an environment $\mathcal{E}$ satisfies type diversity. Then, any incentive compatible social choice function f is virtually Bayesian implementable.*

Before proving this theorem, we consider an example due to Palfrey and Srivastava (Example 3 in Palfrey and Srivastava (1989b)) to clarify the comparison between our result and other related results in the literature. Palfrey and Srivastava use this example to show the difficulty that may arise in an environment with common values. There are two alternatives, $A = \{a, b\}$ and three agents. Each agent has two possible types, $T_i = \{t_a, t_b\}$ and each type is drawn independently with $q_i(t_b) = q$ for all i and $q^2 > 0.5$. Agents have identical preferences, given by

$$u_i(a,t) = \begin{cases} 1 & \text{if at least two agents are of type } t_a \\ 0 & \text{otherwise} \end{cases}$$

$$u_i(b,t) = \begin{cases} 1 & \text{if at least two agents are of type } t_b \\ 0 & \text{otherwise} \end{cases}$$

For each agent, the corresponding interim utilities for the constant SCFs assigning alternatives a and b are:

$$\begin{array}{lllllll} U_i^a(t_a) & = & 1 - q^2, & \qquad & U_i^b(t_a) & = & q^2, \\ U_i^a(t_b) & = & (1-q)^2, & \qquad & U_i^b(t_b) & = & 1 - (1-q)^2. \end{array}$$

Since $q^2 > 0.5$, this implies that the $U_i^b(t_i) > U_i^a(t_i)$ for all i and $t_i \in T_i$; ordinal preferences do not vary across types. Clearly, then only constant SCFs satisfy Bayesian monotonicity in this example. For example, the "majoritarian" SCF which chooses a when at least two agents are of type t_a and

b when at least two agents are of type t_b is not Bayesian implementable. As Palfrey and Srivastava (1989b) show, this SCF is not implementable in undominated Bayesian Nash equilibrium either. It can also be checked that in this environment, only constant SCFs satisfy A-M measurability or incentive consistency. This environment violates type diversity as well.

However, type diversity is easily satisfied if we modify this example to have a third alternative c. For example, suppose $u_i(c,t) = 0$ for all i and all $t \in T$ and the preferences over a and b are the same as before. Now, for each i, Clearly,

$$\frac{U_i^b(t_b)}{U_i^a(t_b)} > \frac{U_i^b(t_a)}{U_i^a(t_a)}$$

which implies that type diversity holds; and therefore, so do A-M measurability and incentive consistency. Thus, Theorem 1 applies to this modified example; any incentive compatible SCF is virtually implementable. In contrast, the results of Abreu and Matsushima (1992b) and Duggan (1997) cannot be applied to any non-constant SCF even in this example. Abreu and Matsushima (1992b) use an assumption (their assumption 2) which requires that in each state the *ex-post* preferences (over lotteries) of the agents are different, which is clearly not the case in the present example. Duggan's (1997) sufficiency theorem uses a weaker version of best element private values. This too fails in the present example. In addition, Duggan (1997, Theorem 2) assumes that $n \geq 3$, while our result also applies to the case where $n = 2$.

It is of interest to note that even in this modified example, the majoritarian SCF cannot be implemented in undominated Nash equilibrium; it can be checked that it does not satisfy the necessary condition identified by Palfrey and Srivastava (1989b). Of course, exact Bayesian implementation of a non-constant SCF remains impossible; ordinal preferences over the alternatives remain identical for all types of all agents even after the new alternative is added, and only constant SCFs satisfy Bayesian monotonicity.

Proof of Theorem 1: The proof constructs a canonical mechanism that virtually Bayesian implements f. Consider the following mechanism $G = ((M_i)_{i\in N}, g)$, where M_i is agent i's message space and g is the outcome function.

Let agent i's message space be $M_i = T_i \times T_i \times I^*$, where I^* is the set of positive integers. We shall denote a typical message sent by agent i by $m_i = (t_i^1, t_i^2, n_i)$.

Let $t^1 = (t_1^1, t_2^1, \ldots, t_n^1)$ be the profile of first type reports, and $t^2 = (t_1^2, t_2^2, \ldots, t_n^2)$ be the profile of types reported in second place by each agent.

Let $L = \{((l_i(t_i))_{t_i \in T_i})_{i \in N})\}$ be the collection of all (constant) SCFs satisfying the condition of Lemma 1. Let $m = |L|$ and $L(t) = \{(l_i(t_i))_{i \in N}\}$. For each $i \in N$ and $t_i \in T_i$, fix two SCFs $l_i^+(t_i)$ and $l_i^-(t_i)$ such that

$$l_i^+(t_i) \in \arg\max_{l \in L} U_i(l \mid t_i), \qquad l_i^-(t_i) \in \arg\min_{l \in L} U_i(l \mid t_i).$$

Note that

$$U_i(l_i^+(t_i) \mid t_i) > U_i(l_i^-(t_i) \mid t_i) \text{ for all } i \in N, \text{ for all } t_i \in T_i.$$

For each $i \in N$, $t_i \in T_i$ and $\lambda \in (0, 1)$ define the following weighted average of SCFs in L:

$$\tilde{l}_i(t_i, \lambda) = \frac{1}{m}[2\lambda l_i^+(t_i) + 2(1-\lambda) l_i^-(t_i) + \sum_{l \in L \setminus \{l_i^+(t_i), l_i^-(t_i)\}} l].$$

Clearly, for all $i \in N$, $t_i \in T_i$, and $\lambda, \lambda' \in (0, 1)$ such that $\lambda' > \lambda > 1/2$,

$$U_i(\tilde{l}_i(t_i, \lambda') \mid t_i) > U_i(\tilde{l}_i(t_i, \lambda) \mid t_i) > U_i(\frac{1}{m} \sum_{l \in L} l \mid t_i). \qquad (*)$$

With these pieces of notation in place, we are ready to write down the outcome function g of the mechanism. Let $\epsilon \in (0, 1)$.

(i) If there exists $j \in N$ such that for all $i \neq j$, $m_i = (t_i, t_i, 1)$, and $n_j = 1$:

$$g(m) = (1-\epsilon) f(t^1) + \frac{\epsilon}{2n} \sum_{i \in N} l_i(t_i^1) + \frac{\epsilon}{2m} \sum_{l \in L} l.$$

(ii) If there exists $j \in N$ such that for all $i \neq j$, $m_i = (t_i, t_i, 1)$, and $n_j > 1$:

$$g(m) = (1-\epsilon) f(t^1) + \frac{\epsilon}{2n} \sum_{i \in N} l_i(t_i^2) + \frac{\epsilon}{2m} \sum_{l \in L} l.$$

(iii) Otherwise, denoting by k the agent with the lowest index among those who announce the highest integer:

$$g(m) = (1-\epsilon) f(t^1) + \frac{\epsilon}{2n} \sum_{i \in N} l_i(t_i^2) + \frac{\epsilon}{2} \tilde{l}_k(t_k^2, \frac{n_k}{\sum_i n_i}).$$

The proof will be explained in several steps.

STEP 1: The strategy profile where for each $i \in N$ and each $t_i \in T_i$, $m_i(t_i) = (t_i, t_i, 1)$ is a Bayesian equilibrium of the mechanism G. To see this, note that changing the first type report is not a profitable deviation because f is incentive compatible and because $U_i(l_i(t_i) \mid t_i) > U_i(l_i(t_i') \mid t_i)$ for any $t_i' \neq t_i$. The latter condition also implies that it is not profitable to change the announcement of the integer and the second type report. Therefore, the proposed strategy profile is a Bayesian equilibrium of G. Note that the outcome produced converges to f as $\epsilon \to 0$.

STEP 2. There cannot be an equilibrium in case (iii) because the winner of the integer game, k, can do better by announcing an even higher integer, which shifts some probability from a least preferred lottery to a most preferred lottery (among which there is a strict preference); see (*).

STEP 3: There cannot be an equilibrium of G under rule (ii) of the outcome function g. Then, some agent j chooses $n_j > 1$ while $n_i = 1$ for all $i \neq j$. In such a case, any player i, $i \neq j$, can announce an integer higher than n_j, trigger the integer game and obtain a higher payoff.

STEP 4: There cannot be an equilibrium of G under rule (i) of the outcome function g in which $n-1$ agents report the same type twice and the integer 1, while the other agent reports different types. If this happens, one of the first $n-1$ agents can trigger the integer game and be its winner. Choosing a sufficiently large integer, by condition (ii) of type diversity, guarantees a profitable deviation.

STEP 5: There cannot be an equilibrium of G under rule (i) where all agents report the same type twice and the integer 1, but where these type reports are not truthful. The reason is that, by condition (i) of type diversity, any agent that is not reporting his true type would increase his expected payoff by reporting his true type in his second report and choosing an integer greater than 1. ■

Remark 1. In environments violating type diversity, virtual implementation may be impossible. For instance, this is the case in Example 1 of Serrano and Vohra (2001), where only constant SCFs are virtually implementable in Bayesian equilibrium, even though the set of incentive compatible SCFs contains many non-constant ones. In fact, in the environment described in that example, implementation is also impossible in other solution concepts: only constant SCFs satisfy the necessary condition for undominated Bayesian implementation identified by Palfrey and Srivastava (1989b), and the necessary condition for perfect Bayesian implementation identified by Brusco (1995). Non-constant SCFs in that example also escape the sufficient conditions for implementation in sequential equilibrium used in Baliga (1999) and in Bergin and Sen (1998).

Remark 2. Note that the integer game used in our mechanism can be replaced by a modulo game. Thus virtual implementation, under type diversity, does not require the use of infinite mechanisms for finite environments. This is to be contrasted with the result of Dutta and Sen (1994) showing that infinite mechanisms may be unavoidable for exact Bayesian implementation. While the example used by Dutta and Sen (1994) does not satisfy type diversity, it can be modified by adding a third alternative which yields 0 utility to each agent in each state to satisfy type diversity. It is easy to check that their result continues to apply to this modified example but the mechanism constructed above, with a modulo game instead of an integer game, is a finite mechanism that yields virtual Bayesian implementation. Indeed, according to the Abreu-Matsushima (1992b) result, under their conditions, virtual implementation can be accomplished through a regular mechanism.

5 A Necessary and Sufficient Condition for Economic Environments

While we have shown that type diversity is sufficient for virtual implementation, it is not a necessary condition. This follows from Example 2 in Serrano and Vohra (2001) which exhibits an economic environment violating type diversity in which *every* SCF is virtually Bayesian implementable. In that example only constant SCFs satisfy Bayesian incentive consistency or AM-measurability.[11] In this section we identify a condition which is necessary and sufficient for virtual Bayesian implementation in economic environments. The results in this section can, therefore, be applied even to 'non-generic' cases such as Example 2 of Serrano and Vohra (2001).

[11] To the extent that A-M measurability is a necessary condition for virtual implementation in iteratively undominated strategies, it follows that, unlike in complete information environments, virtual implementation in iteratively undominated strategies is less permissive than virtual Nash implementation. Also, since A-M measurability is necessary for virtual implementation using regular mechanisms, integer (or modulo) games seem essential to implementation in environments like the ones in that example.

There are two reasons why type diversity, or more precisely its implication drawn out in Lemma 1, is stronger than necessary for virtual implementation: (1) it requires a preference reversal for *all* types of *all* agents, (2) it is based exclusively on a comparison of constant SCFs.[12] As we shall see, it is possible to do with a weaker condition which uses a preference reversal only for some type of some agent; and, by considering general (non-constant) SCFs, it necessitates the reliance on the notion of deceptions.

An SCF x is said to be *interim top ranked* for agent i if there does not exist another SCF y and a type $s_i \in T_i$ such that $U_i(y|s_i) > U_i(x|s_i)$. Of course, given our assumption of a finite number of alternatives, there exists a top ranked SCF for each $i \in N$. Denote by $\hat{f}$ the constant SCF that prescribes the uniform probability distribution over alternatives in each state.

An environment $\mathcal{E}$ is said to be *economic* if:

(i) there exist at least two agents i and j for whom $\hat{f}$ is not interim top ranked;

(ii) for each $i \in N$ there exists an interim top ranked SCF ω^i with the property that there exists another agent j for whom ω^i is not a top ranked SCF.

In an exchange economy, under the standard assumptions, ω^i can be taken to be the SCF which allocates the aggregate endowment to agent i in each state. However, an economic environment is not restricted to be an exchange economy. Our notion of an economic environment is similar, but slightly weaker, than the one used in Jackson (1991). In particular, it does not rule out the possibility that there is some SCF that is interim top ranked by all agents.

For smoothness in the presentation, we first present a new condition on environments, and then we shall turn it into a condition on the SCF.

An environment $\mathcal{E}$ satisfies *type diversity with respect to deceptions* if for every deception α, there exists $i \in N$, $s_i \in T_i$, and SCFs x and y such that

$$U_i(y \circ \alpha \mid s_i) > U_i(x \circ \alpha \mid s_i) \text{ while } U_i(x \mid s_i') \geq U_i(y_{\alpha_i(s_i)} \mid s_i'), \forall s_i' \in T_i. \quad (**)$$

An SCF f satisfies *type diversity with respect to deceptions* if for every deception α satisfying that $f \neq f \circ \alpha$, there exists $i \in N$, $s_i \in T_i$, and SCFs x and y such that equation (**) holds.

[12] These are also the reasons why it is a simple condition.

To get a feel for this condition, the reader will notice how it is inspired by the statement of Bayesian monotonicity, except that the preference reversal does not necessarily involve f, our SCF of interest. Therefore, as a condition on the environment, for each deception α we have a set of test-agents: those agents for whom some of their types exhibit a preference reversal between two SCFs as specified in (**). As a condition on the SCF, the same is true only for those deceptions that turn f into a non-equivalent SCF. Note that, when written as a condition on the SCF, type diversity with respect to deceptions is clearly weaker than Bayesian monotonicity. We obtain the following characterization, to be compared to Jackson's (1991) Theorem 1.

Theorem 2 *In an economic environment, a social choice function f is virtually Bayesian implementable if and only if it satisfies incentive compatibility and type diversity with respect to deceptions.*

The proof of this result is identical to that of Theorem 3. Necessity follows from a minor modification of the proof of part (1) of Theorem 3, and sufficiency from the proof of part (2). Although Theorem 2 conforms to the standard way of stating characterization results, it is Theorem 3 which relates most directly with Theorem 1. Drawing on Theorem 1 and the weak condition on environments used there (type diversity), one gets a good feel for the meaning of the more involved condition using deceptions.

Theorem 3 (1) *Suppose f is a social choice function with the property that $f(t) \neq f(t')$ for all $t, t' \in T^*$. If f is virtually implementable in Bayesian equilibrium, then the environment $\mathcal{E}$ satisfies type diversity with respect to deceptions.*
(2) *Suppose an economic environment $\mathcal{E}$ satisfies type diversity with respect to deceptions. If f is incentive compatible, it is virtually Bayesian implementable.*

Part (1) of the theorem says that type diversity with respect to deceptions (as a condition on the environment) is essentially a necessary condition for the success of virtual Bayesian implementation, in the sense that "most" SCFs satisfy the hypothesis stated: recall that an SCF is a mapping from T to $\triangle$, and thus, the hypothesis of part (1) is satisfied by an open and dense set of SCFs. Conceptually, the assumption is also pretty minor: one should be indifferent between virtually implementing f or an arbitrarily small perturbation thereof.

In contrast, part (2) states that type diversity with respect to deceptions (on the environment) is sufficient for any incentive compatible SCF to be virtually Bayesian implementable over economic environments. Thus, in these environments, the multiplicity of equilibrium problem disappears and virtual implementation is as successful as it can be: its limits are given only by incentive compatibility. Type diversity with respect to deceptions is a weaker condition than type diversity, identified in Section 3. However, it is not a trivial condition either: for instance, as implied by the remark at the end of last section, Example 1 in Serrano and Vohra (2001) describes an environment that violates it, where virtual implementation is impossible.

Proof of Theorem 3, part (1). The argument follows closely from the Bayesian monotonicity condition (see the definition given in Section 2). If the environment does not satisfy type diversity with respect to deceptions, there exists a deception α for which condition (**) is violated. Therefore, f, which is non-constant everywhere, is not exactly implementable. But it is not virtually implementable either: in order to virtually implement f, one would need to find a sequence of exactly implementable SCFs h_ϵ, which are also non-constant everywhere. But by the same argument as above, these h_ϵ cannot be exactly implementable either (otherwise, (**) would be satisfied, which is a contradiction). ■

Proof of Theorem 3, part (2). We shall construct a canonical mechanism, $G = ((M_i)_{i \in N}, g)$. Before we describe strategy sets and outcome function, we introduce some useful notation.

Denote by D_i the set of deceptions for which i is a test-agent. For each test-agent i and each deception $\alpha \in D_i$, fix two SCFs $x_i(\alpha)$ and $y_i(\alpha)$ satisfying (**). Let

$$C_i = \{(z_i(\alpha))_{\alpha \in D_i}\},$$

where for each $\alpha \in D_i$, $z_i(\alpha)$ can take one of two possible values, either $x_i(\alpha)$ or $y_i(\alpha)$. Thus, a typical element of the set C_i is a list of $|D_i|$ components. Each component is one of the two SCFs in (**) associated with a deception α for which agent i is a test-agent. Let $x_i \in C_i$ be such that for all $\alpha \in D_i$, $z_i(\alpha) = x_i(\alpha)$. Also, if agent i is not a test-agent for any deception α, let $C_i = \{x_i\}$ for some arbitrary SCF x_i.

Let (ω^i) denote a collection of top ranked SCFs satisfying the definition of an economic environment.

Next, we define the message set M_i in the mechanism G: let $M_i = T_i \times C_i \times I^*$, where I^* is the set of positive integers. Denote by (m_i^1, m_i^2, m_i^3) a typical message sent by agent i.

For $\epsilon \in (0,1)$, the outcome function is defined in the following rules:

(i) If at least $n-1$ agents announce $m_i^2 = (z_i(\alpha))_{\alpha \in D_i} = x_i$ and $m_i^3 = 1$:

$$g(m) = (1-\epsilon)f(m^1) + \frac{\epsilon}{2}\hat{f} + \frac{\epsilon}{2n}\sum_{i \in N}[\frac{1}{\max\{1, |D_i|\}}\sum_{\alpha \in D_i} z_i(\alpha)].$$

(ii) Otherwise, denoting by k the agent with the lowest index among those who announce the highest integer, we have:

$$g(m) = (1-\epsilon)f(m^1) + \frac{\epsilon}{2}\omega^k + \frac{\epsilon}{2n}\sum_{i \in N}[\frac{1}{\max\{1, |D_i|\}}\sum_{\alpha \in D_i} z_i(\alpha)].$$

To prove the theorem, we take the following steps:

STEP 1: The strategy profile where for each $i \in N$ and each $t_i \in T_i$, $m_i(t_i) = (t_i, x_i, 1)$ is a Bayesian equilibrium of G. To see this, note that this strategy profile corresponds to the outcome of rule (i). Moreover, no unilateral deviation from it can trigger rule (ii). Since f is incentive compatible, reporting a false type is not a profitable deviation for any agent. By (**), changing the second component (or both the first and the second components at the same time) of the message is not profitable either. Changing the reported integer does not change the outcome in this case. Thus, as claimed, this profile is a Bayesian equilibrium of G. Note that as $\epsilon \to 0$, the equilibrium outcome converges to f.

STEP 2: An equilibrium under rule (ii) of the outcome function is impossible. Suppose not, i.e., suppose there is an equilibrium in which agent k wins the integer game. Because the environment is economic, there is some agent $j \neq k$ for whom ω^j interim dominates ω^k. Thus agent j has an incentive to deviate and announce an integer higher than n_k, thereby becoming the winner of the integer game.

STEP 3: An equilibrium under rule (i) of the outcome function g where exactly $n-1$ agents j are announcing $m_j^2 = x_j$ and $m_j^3 = 1$, while agent i is announcing something else, is also impossible. For a strategy profile of this kind, any agent $j \neq i$ can trigger the integer game and become its winner by announcing a high enough integer. By the definition of an economic environment, there exists $j \neq i$ for whom ω^i interim dominates $\hat{f}$. But then j has an incentive to deviate and trigger the integer game.

STEP 4: Finally, it is also impossible to have an equilibrium of G under rule (i) where each agent i announces $m_i^2 = x_i$ and $m_i^3 = 1$, but where a deception α is being used to misreport the types. In this case, for this α and by type diversity with respect to deceptions, there exists an agent i and two SCFs $x_i(\alpha)$ and $y_i(\alpha)$ satisfying (**). Therefore, type s_i of agent i has an incentive to deviate and change the second component of his announcement to $(y_i(\alpha), x_i(D_i \setminus \{\alpha\}))$. ■

6 Concluding Remarks

We have shown that the difference between truthful implementation (the approach based on the revelation principle) and full implementation in incomplete information environments is perhaps smaller than at first thought. Indeed, thanks to the type diversity condition, we have established that in almost every environment any incentive compatible SCF is virtually Bayesian implementable. Furthermore, we have provided a characterization of virtual Bayesian implementation in economic environments using a weakening of type diversity (type diversity with respect to deceptions).

A final observation is in order, given the extremely positive results reported in this paper. One may wonder how much the results depend on the expected utility assumption. To the extent that this is just an approximation of the preferences that agents may have in the "real world," it would be desirable that the assumption of expected utility be not a crucial one for the theory. Indeed, one can easily see that all our conclusions extend to preferences over lotteries that are monotonic in the sense of first-order stochastic dominance. Reflection on Figure 2 should suffice to convince the reader of this assertion: note that what is needed is that the relevant indifference surfaces yield non-nested lower contour sets in the interior of the probability simplex. This is completely independent from having a map of parallel straight lines (see Abreu and Sen (1991) for a similar observation in the context of virtual Nash implementation).

References

Abreu, D. and H. Matsushima (1992a), "Virtual Implementation in Iteratively Undominated Strategies: Complete Information," *Econometrica*, 60, 993-1008.

Abreu, D. and H. Matsushima (1992b), "Virtual Implementation in Iteratively Undominated Strategies: Incomplete Information," mimeo, Princeton University.

Abreu, D. and A. Sen (1991), "Virtual Implementation in Nash Equilibrium," *Econometrica*, 59, 997-1021.

Baliga, S. (1999), "Implementation in Economic Environments with Incomplete Information: the Use of Multi-Stage Games," *Games and Economic Behavior*, 27, 173-183.

Bergin, J. and A. Sen (1998), "Extensive Form Implementation in Incomplete Information Environments," *Journal of Economic Theory*, 80, 222-256.

Brusco, S. (1995), "Perfect Bayesian Implementation," *Economic Theory*, 5, 419-444.

Chakravorti, B. (1992), "Efficiency and Mechanisms with no Regret," *International Economic Review*, 33, 45-59.

Duggan, J. (1997), "Virtual Bayesian Implementation," *Econometrica*, 65, 1175-1199.

Dutta, B. and A. Sen (1994), "Bayesian Implementation: the Necessity of Infinite Mechanisms," *Journal of Economic Theory*, 64, 130-141.

Jackson, M. (1991), "Bayesian Implementation," *Econometrica*, 59, 461-477.

Maskin, E. (1977), "Nash Equilibrium and Welfare Optimality," mimeo, M.I.T, published (1999): *Review of Economic Studies*, 66, 23-38.

Matsushima, H. (1988), "A New Approach to the Implementation Problem," *Journal of Economic Theory*, 45, 128-144.

Matsushima, H. (1993), "Bayesian Monotonicity with Side Payments," *Journal of Economic Theory*, 59, 107-121.

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

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

Myerson, R. (1987), "Mechanism Design," in J. Eatwell, M. Milgate and P. Newman (eds.) *The New Palgrave Dictionary of Economics*, Norton, New York.

Palfrey, T. and S. Srivastava (1987), "On Bayesian Implementable Allocations," *Review of Economic Studies*, 54, 193-208.

Palfrey, T. and S. Srivastava (1989a), "Implementation with Incomplete Information in Exchange Economies," *Econometrica*, 57, 115-134.

Palfrey, T. and S. Srivastava (1989b), "Mechanism Design with Incomplete Information: a Solution to the Implementation Problem," *Journal of Political Economy*, 97, 668-691.

Palfrey, T. and S. Srivastava (1993), *Bayesian Implementation*, Harwood Academic Publishers, New York.

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

Saijo, T. (1987), "On Constant Maskin Monotonic Social Choice Functions," *Journal of Economic Theory*, 42, 382-386.

Serrano, R. and R. Vohra (2001), "Some Limitations of Virtual Bayesian Implementation," *Econometrica*, 69, 785-792.