



Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Games and Economic Behavior 50 (2005) 312–331

GAMES and
Economic
Behavior

www.elsevier.com/locate/geb

A characterization of virtual Bayesian implementation[☆]

Roberto Serrano and Rajiv Vohra^{*}

Department of Economics, Brown University, Providence, RI 02912, USA

Received 20 June 2003

Available online 27 April 2004

Abstract

We provide a characterization of virtual Bayesian implementation in pure strategies for environments satisfying no-total-indifference. A social choice function in such environments is virtually Bayesian implementable if and only if it satisfies incentive compatibility and a condition we term virtual monotonicity. The latter is weaker than Bayesian monotonicity—known to be necessary for Bayesian implementation. Virtual monotonicity is weak in the sense that it is generically satisfied in environments with at least three alternatives. This implies that in most environments virtual Bayesian implementation is as successful as it can be (incentive compatibility is the only condition needed).
© 2004 Elsevier Inc. All rights reserved.

JEL classification: C72; D78; D82

Keywords: Virtual monotonicity; Bayesian monotonicity; Maskin monotonicity; Virtual Bayesian implementation; Incentive compatibility; Type diversity

1. Introduction

Consider the problem of implementing a social choice function (SCF) in an environment with incomplete information among the agents. Since the socially desirable outcome depends on agents' private information, it is possible that agents do not have the incentive to correctly reveal their private information. Bayesian incentive compatibility of an SCF is

[☆] Part of this paper circulated as "Type Diversity and Virtual Bayesian Implementation," Working Paper No. 2000-16, Department of Economics, Brown University.

^{*} Corresponding author.

E-mail addresses: roberto_serrano@brown.edu (R. Serrano), Rajiv_Vohra@brown.edu (R. Vohra).

URLs: <http://www.econ.brown.edu/faculty/serrano>, <http://www.econ.brown.edu/~rvohra>.

simply the requirement that each agent has the incentive to truthfully reveal her information when all other agents report their information truthfully. More precisely, an SCF is said to be incentive compatible if truth-telling is a Bayesian Nash equilibrium of the direct revelation game (in which agents report their private information and the outcome is the social choice corresponding to these reports). By the revelation principle (see Myerson, 1989 and the references therein), incentive compatibility of an SCF is necessary for it to be implemented through a Bayesian Nash equilibrium of *any* mechanism. In general, however, a mechanism applied to an incentive compatible SCF may possess other equilibria (which do not correspond to the socially desired outcome). Full implementation refers to the design of a mechanism that resolves this multiplicity problem by ensuring that *all* equilibria correspond to the socially desired outcome in each information state, and requires some condition in addition to incentive compatibility.

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.¹ 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. Within economic environments, the gap after Palfrey and Srivastava's work between necessary and sufficient conditions was closed by Jackson (1991) with a strengthening of Bayesian monotonicity.² Unfortunately, Bayesian monotonicity is not satisfied by many well-known social choice functions (SCFs) for exchange economies with incomplete information; see Palfrey and Srivastava (1987), Chakravorti (1992) and Serrano and Vohra (2001). In this sense, Bayesian monotonicity is a demanding condition. In light of Jackson's characterization result this means that (full) Bayesian implementation is a demanding requirement.

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 a correspondence that is arbitrarily close to the given correspondence.³ 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

¹ 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).

² 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.

³ This can be explained, in part, by the fact that Maskin monotonicity defined over lotteries is a very weak condition.

in the more attractive notion of iteratively undominated strategies, and that 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 settle two open issues in this literature:

- Characterize the SCFs that are virtually Bayesian implementable;
- Provide a simple, weak and readily interpretable condition that is sufficient for virtual Bayesian implementation of an incentive compatible SCF.

It should be emphasized that we are concerned with implementation in pure Bayesian Nash equilibrium. In this sense our characterization of virtual Bayesian implementation is comparable to Jackson's (1991) characterization of Bayesian implementation.

Two sufficient conditions for virtual Bayesian implementation (in addition to incentive compatibility) are available in the existing literature. Abreu and Matsushima (1992b) dispense with Bayesian monotonicity and introduce a new condition termed *measurability* (henceforth A–M measurability) which, under other weak assumptions, along with incentive compatibility, 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 mixed Bayesian Nash equilibrium. Duggan (1997) suggests the condition of *incentive consistency* and presents a sufficiency result for environments with “best-element private values.”⁴ Serrano and Vohra (2001) criticize A–M measurability and incentive consistency, by showing them to be sometimes even 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. Thus, A–M measurability or incentive consistency are far from necessary for virtual Bayesian implementation. For a more detailed comparison of these results with ours, see Section 1.1 below.

Our main result shows that in environments satisfying no-total-indifference (NTI), an incentive compatible SCF is virtually implementable if and only if it satisfies a condition we term *virtual monotonicity*. This result includes two-agent and ‘non-economic’ environments. Virtual monotonicity is inspired by the Bayesian monotonicity condition stated in Jackson (1991), and by the arguments used by Matsushima (1988) and Abreu and Sen (1991) for virtual Nash implementation. Like Maskin monotonicity (Maskin, 1977) and the conditions and insights used by those authors, virtual monotonicity is a condition

⁴ In the same endeavor of attempting to dispense with Bayesian monotonicity, Matsushima (1993) shows that this can be done if side payments are allowed. In contrast, our results do not rely on transfers.

on lower contour sets of preferences. Those papers and our characterization result provide a unified theory of implementation using pure Nash equilibrium as a solution concept. Our new monotonicity condition requires, for each deception that undermines the given SCF, the existence of a certain preference reversal for at least one type of one agent. Such a preference reversal is not required to hold for the given SCF but for *some* incentive compatible SCF, thereby making it a weaker condition than Bayesian monotonicity. In Section 5, we shall consider an important example, due to Palfrey and Srivastava (1989b), of a common-values environment. In this example, the majoritarian SCF does not satisfy Bayesian monotonicity (or indeed, any of the other available sufficient conditions for implementation in other solution concepts). Nor does it satisfy A–M measurability or incentive consistency. But it does satisfy virtual monotonicity, and is therefore virtually Bayesian implementable. In fact, in this example, *all* SCFs satisfy virtual monotonicity.

Our characterization result demonstrates that virtual Bayesian implementation is far more permissive than Bayesian implementation. The argument goes beyond the fact that virtual monotonicity is weaker than Bayesian monotonicity. As we will show, there is a related condition on environments, which we term type diversity, such that *any* incentive compatible SCF is virtually implementable in such environments. This condition is stronger than the necessary condition (it implies that every SCF satisfies virtual monotonicity), but it is much easier to state and interpret. It requires that the interim (cardinal) preferences over pure alternatives of different types of an agent be different.⁵ It is considerably easier to verify than the other conditions mentioned above. More importantly, type diversity turns out to be generic in the set of all environments with at least three alternatives. This implies that in most environments, since every SCF satisfies virtual monotonicity, every SCF is virtually implementable in Bayesian equilibrium provided it is incentive compatible. In other words, 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.

1.1. A comparison with the related literature

As indicated above, our characterization of virtual Bayesian implementation in pure strategies is most related to Jackson's (1991) Theorem 1—both are concerned with implementation in pure strategies. The difference is that we also include non-economic environments and the case of two agents; and in weakening the implementation

⁵ This condition appears in Abreu and Matsushima (1992b) as a simple way of ensuring A–M measurability. In environments satisfying type diversity, every SCF also satisfies incentive consistency. In a private values model, it reduces to the condition of value-distinguished types introduced in Palfrey and Srivastava (1989b).

requirement from exact to virtual Bayesian implementation we can weaken his Bayesian monotonicity condition to virtual monotonicity.

Abreu and Matsushima (1992b) show that if mixed strategies are included and mechanisms are restricted to be regular, then virtual Bayesian Nash implementation is characterized by incentive compatibility and A–M measurability. A characterization for the case in which mixed strategies are included, but mechanisms are not required to be regular, remains open.

Duggan (1997) covers mixed equilibria in his sufficiency result. He also uses general type spaces, whereas we assume a finite set of types. His main result assumes an “ex post no-total-indifference” condition, instead of our somewhat stronger NTI. However, he also assumes “best element private values,” which in conjunction with his version of NTI implies our NTI condition. An important issue for future work is a characterization that allows mixed strategies and general type spaces.

2. The model and definitions

We shall consider implementation in the context of a general environment with asymmetric information. Let $N = \{1, \dots, n\}$ be a finite set of agents. Let T_i denote the (finite) set of agent i 's types. The interpretation is that $t_i \in T_i$ describes the *private information* possessed by agent i . We refer to a profile of types $t = (t_1, \dots, t_n)$ 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 $q_i(t_{-i}|\bar{t}_i)$. Let $T^* \subseteq T$ be the set of states with positive probability. We assume that agents agree on the states in T^* , i.e., for all $i \in N$, $q_i(t) = 0$ if and only if $t \notin T^*$.

Let A denote the set of social alternatives, which are assumed to be independent of the information state. Let \mathcal{A} be a σ -algebra on A and Δ denote the set of probability measures on (A, \mathcal{A}) . We shall assume that \mathcal{A} contains all singleton sets. The Bernoulli utility of agent i for alternative a in state t is $u_i(a, t)$.

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

A *social choice function* (SCF) is a function $f: T \mapsto \Delta$. Two SCFs, f and h are equivalent ($f \approx h$) if $f(t) = h(t)$ for every $t \in T^*$ (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.

Abusing notation slightly, given an SCF f , $u_i(f, t)$ will refer to agent i 's expected utility evaluation of lottery $f(t)$ in state t . The (interim/conditional) expected utility of agent i of type t_i corresponding to an SCF f is defined as:

$$U_i(f|t_i) \equiv \sum_{t'_{-i} \in T_{-i}} q_i(t'_{-i}|t_i) u_i(f, (t'_{-i}, t_i)).$$

We shall make the (weak) regularity assumption that there is no-total-indifference. This assumption will be in force throughout the paper.

An environment \mathcal{E} satisfies *no-total-indifference* (NTI) if for every $j \in N$, $t_j \in T_j$ and $T'_{-j} \subseteq T_{-j}$ such that $\{t_j\} \times T'_{-j} \subseteq T^*$, there exist $a, a' \in A$ such that

$$\sum_{t_{-j} \in T'_{-j}} q_j(t_{-j}|t_j)u_j(a, t) \neq \sum_{t_{-j} \in T'_{-j}} q_j(t_{-j}|t_j)u_j(a', t).$$

Since N and T are finite, it follows that there is a finite set $\bar{A} \subseteq A$ such that a and a' in the above condition belong to \bar{A} .

This assumption amounts to the statement that there is no total-indifference for each agent of each type whatever the updated beliefs about the other agents' types, provided that the updating is consistent with Bayes' rule. Hence, NTI includes the assumption of no-total-indifference ex post (made in Duggan, 1997) as well as interim (Abreu and Matsushima, 1992b). While our assumption is stronger than the corresponding assumption in Duggan (1997), he also makes another assumption, best element private values, and it can be shown that the conjunction of that and his ex post NTI condition is stronger than our version of NTI.

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

A (pure strategy) *Bayesian equilibrium* of G is a profile of strategies, $\sigma = (\sigma_i)_{i \in N}$ where $\sigma_i : T_i \mapsto M_i$ such that $\forall i \in N, \forall t_i \in T_i$,

$$U_i(g(\sigma)|t_i) \geq U_i(g(\sigma_{-i}, \sigma'_i)|t_i) \quad \forall \sigma'_i : T_i \mapsto M_i.$$

Denote by $B(G)$ the set of Bayesian equilibria of the mechanism G . Let $g(B(G))$ be the corresponding set of equilibrium outcomes.

An SCF f is *exactly Bayesian implementable* if there exists a mechanism G such that every $h \in g(B(G))$ is equivalent to f .⁶

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

Consider the following metric on SCFs:

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

An SCF f is *virtually Bayesian implementable* if $\forall \epsilon > 0$ there exists an SCF f^ϵ such that $d(f, f^\epsilon) < \epsilon$ and f^ϵ is exactly Bayesian implementable.

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

⁶ 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.

The next condition is necessary for exact Bayesian implementation (see Jackson, 1991).

An SCF f satisfies *Bayesian monotonicity* if for any deception α , whenever $f \circ \alpha \not\approx f$, there exist $i \in N$, $t_i \in T_i$ and an SCF y such that

$$U_i(y \circ \alpha | t_i) > U_i(f \circ \alpha | t_i) \quad \text{while} \quad U_i(f | t'_i) \geq U_i(y_{\alpha_i(t_i)} | t'_i), \quad \forall t'_i \in T_i.$$

An SCF f satisfies *incentive compatibility* if for all $i \in N$, $t_i \in T_i$ and all deceptions α ,

$$U_i(f | t_i) \geq U_i(f_{\alpha_i(t_i)} | t_i).$$

Jackson (1991) provides a characterization result for economic environments with at least three agents: a social choice function is Bayesian implementable if and only if it satisfies incentive compatibility and Bayesian monotonicity.⁷ This result can be readily compared to our characterization theorem for virtual Bayesian implementation, found in the next section.

3. A characterization result

In this section we show that a substantial weakening of Bayesian monotonicity yields a necessary and sufficient condition, together with incentive compatibility, for virtual Bayesian implementation.

An SCF f satisfies *virtual monotonicity* if for every deception α , whenever $f \circ \alpha \not\approx f$, there exists $i \in N$, $t_i \in T_i$, an incentive compatible SCF x and an SCF y such that

$$U_i(y \circ \alpha | t_i) > U_i(x \circ \alpha | t_i) \quad \text{while} \quad U_i(x | t'_i) \geq U_i(y_{\alpha_i(t_i)} | t'_i), \quad \forall t'_i \in T_i. \quad (*)$$

The difference with Bayesian monotonicity is that the preference reversal in the new condition does not necessarily involve the SCF f . For each deception α such that $f \circ \alpha \not\approx f$, we have an agent for whom some of their types exhibit a preference reversal between two SCFs as specified in (*). Clearly, under incentive compatibility, virtual monotonicity is weaker than Bayesian monotonicity. A more detailed comparison of the two is provided in Section 3.1, where we show that the necessary preference reversal can be shown to hold for SCFs that are arbitrarily close to f . This is our rationale for the term “virtual monotonicity.” We postpone to Section 4 a discussion of how weak virtual monotonicity really is, and how it follows from a much simpler condition that does not involve any reference to deceptions.

Our main result is the following characterization theorem.

Theorem 1. *Suppose an environment \mathcal{E} satisfies NTI. Then, a social choice function f is virtually Bayesian implementable if and only if it satisfies incentive compatibility and virtual monotonicity.*

In comparing this result to Jackson’s (1991) Theorem 1, note that our characterization of virtual Bayesian implementation does not assume the environment to be economic, nor

⁷ In the more general case of a social choice set, an added condition, closure, is also needed. This condition requires that the social choice set be closed under concatenation of common knowledge events.

does it require the number of agents to be at least three. In terms of the conditions on the SCF, closure being a trivial requirement then, the only difference reduces to requiring Bayesian monotonicity instead of its virtual counterpart.

Proof of Theorem 1. Necessity. Since the necessity of incentive compatibility is well known,⁸ we shall show that virtual monotonicity is necessary for virtually implementing an incentive compatible SCF.

Consider a deception α such that $f \circ \alpha \not\approx f$, and suppose f is virtually implementable, i.e., for every $\epsilon > 0$ there exists and SCF f^ϵ which is implementable and is ϵ -close to f . Thus f^ϵ satisfies incentive compatibility and Bayesian monotonicity. Choosing ϵ small enough, it follows that for the given deception α , $f^\epsilon \circ \alpha \not\approx f^\epsilon$, and by Bayesian monotonicity of f^ϵ , there exists $i \in N$, $t_i \in T_i$ and an SCF y such that

$$U_i(y \circ \alpha | t_i) > U_i(f^\epsilon \circ \alpha | t_i) \quad \text{while} \quad U_i(f^\epsilon | t'_i) \geq U_i(y_{\alpha_i(t_i)} | t'_i), \quad \forall t'_i \in T_i. \quad (1)$$

But this means that whenever $f \circ \alpha \not\approx f$, there exist $i \in N$, $t_i \in T_i$ and a pair of SCFs, x (incentive compatible) and y satisfying (*); simply choose $x = f^\epsilon$. Thus f satisfies virtual monotonicity.

Sufficiency. Suppose f satisfies incentive compatibility and virtual monotonicity. We shall construct a canonical mechanism, $G = ((M_i)_{i \in N}, g)$ to virtually implement f in Bayesian equilibrium. Before we describe the strategy sets and the outcome function, we introduce some additional notation.

For a deception α such that $f \circ \alpha \not\approx f$, a test-agent is any agent for whom condition (*) holds. 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^α and y_i^α satisfying (*) for i of type t_i , where x_i^α is incentive compatible. Notice that condition (*) concerns the SCF y only in those states in which agent i is of type $\alpha_i(t_i)$. There is, therefore, no loss of generality in assuming that y_i^α is of the form:

$$y_i^\alpha(t_{-i}, t'_i) = y_i^\alpha(t_{-i}, t_i) \quad \text{for all } t_{-i} \in T_{-i} \quad \text{and} \quad t'_i \in T_i.$$

Thus y_i^α is constant over T_i .⁹

If agent i is a test-agent for some α , let

$$C_i = \{(z_i^\alpha)_{\alpha \in D_i} \mid \forall \alpha \in D_i, z_i^\alpha \in \{x_i^\alpha, y_i^\alpha\}\}.$$

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. For notational reasons, we will also find it convenient to construct these sets for agents who are not test-agents. For each i who is not a test-agent for any deception fix an arbitrary deception $\bar{\alpha}$ and an arbitrary incentive compatible SCF, $x_i^{\bar{\alpha}}$, and let $D_i = \{\bar{\alpha}\}$ and let $C_i = \{x_i^{\bar{\alpha}}\}$.

⁸ See Duggan (1997). Strictly speaking, what is necessary is the existence of an equivalent SCF that is incentive compatible, but given the definition of implementation, there is no loss of generality in taking f itself to be incentive compatible. This should also be understood in the way Theorem 1 is stated.

⁹ This observation also applies to the definition of Bayesian monotonicity.

Denote by $\bar{a} \in \Delta$ the uniform probability distribution over \bar{A} . For $\epsilon \in (0, 1)$, define the SCF f^ϵ as

$$f^\epsilon(t) = (1 - \epsilon)f(t) + \frac{\epsilon}{2}\bar{a} + \frac{\epsilon}{2n} \sum_{i \in N} \left[\frac{1}{|D_i|} \sum_{\alpha \in D_i} x_i^\alpha(t) \right].$$

We will now prove that f^ϵ is Bayesian implementable with the following mechanism:

The message set of agent i is defined as $M_i = T_i \times C_i \times \bar{A} \times I$, where I is the set of non-negative integers. Denote a typical message of agent i as $m_i = (\hat{t}_i, c_i, a_i, k_i)$, where $c_i = (z_i^\alpha)_{\alpha \in D_i}$, and let m denote a profile of messages. The outcome function $g: \prod_{i \in N} M_i \mapsto \Delta$ is defined by the following rules:

- (i) If m is such that at least $n - 1$ agents announce $c_i = (x_i^\alpha)_{\alpha \in D_i}$ and $k_i = 0$,

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

- (ii) Otherwise, denoting by h the agent with the lowest index among those who announce the highest integer,

$$g(m) = (1 - \epsilon)f(\hat{t}) + \frac{\epsilon}{2} \left[\frac{1}{k_h + 1} \bar{a} + \frac{k_h}{k_h + 1} a_h \right] + \frac{\epsilon}{2n} \sum_{i \in N} \left[\frac{1}{|D_i|} \sum_{\alpha \in D_i} z_i^\alpha(\hat{t}) \right].$$

To prove the theorem, we take the following steps:

Step 1. A strategy profile where for each $i \in N$ and each $t_i \in T_i$, $\hat{t}_i = t_i$, $c_i = (x_i^\alpha)_{\alpha \in D_i}$ and $k_i = 0$ 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), and therefore a_i and k_i have no effect on the outcome. The only way an agent can change the outcome is by changing his announcement of \hat{t}_i or c_i . Since f is incentive compatible, and so is x_i^α for all $i \in N$ and $\alpha \in D_i$, reporting a false type is not a profitable deviation for any agent. By condition (*) it is not profitable to report a change in c_i . Nor is it possible to profit by changing both \hat{t}_i and c_i because each y_i^α is constant with respect to i 's type. Thus, as claimed, any such profile is a Bayesian equilibrium of G . Note that as $\epsilon \rightarrow 0$, the equilibrium outcome converges to f .

Step 2. There cannot be an equilibrium σ that induces case (ii) in any state in T^* . Suppose not. Let $T^l \subseteq T^*$, $T^l \neq \emptyset$ be the set of states in T^* in which σ induces the integer game (case (ii)). Let \hat{k} be the highest integer announced in any state in T^l , and let h be the lowest indexed agent who announces \hat{k} in some state t in T^l . Let

$$T'_{-h} = \{t'_{-h} \in T_{-h} \mid (t_h, t'_{-h}) \in T^l\}.$$

By hypothesis, this set is non-empty. By construction, agent h of type t_h , by announcing $k_h = \hat{k}$, wins the integer game in all states in $\{t_h\} \times T'_{-h}$. Since h is the lowest indexed agent who announces \hat{k} in T^l , if agent h of type t_h changes her announcement of the integer to $k'_h > \hat{k}$, everything else being the same, she continues to be the winner in precisely the

same states as before, namely $\{t_h\} \times T'_{-h}$. Let $\sigma_h(t_h) = (\hat{t}_h, c_h, a_h, \hat{k})$. Consider a strategy σ'_h such that $\sigma'_h(t_h) = (\hat{t}_h, c_h, a'_h, k'_h)$, where

$$k'_h > \hat{k}, \tag{2}$$

$$\sum_{t_{-h} \in T'_{-h}} q_h(t'_{-h} | t_h) u_h(a'_h, (t'_{-h}, t_h)) \geq \sum_{t_{-h} \in T'_{-h}} q_h(t'_{-h} | t_h) u_h(a_h, (t'_{-h}, t_h)) \tag{3}$$

and

$$\sum_{t_{-h} \in T'_{-h}} q_h(t'_{-h} | t_h) u_h(a'_h, (t'_{-h}, t_h)) > \sum_{t_{-h} \in T'_{-h}} q_h(t'_{-h} | t_h) u_h(\bar{a}, (t'_{-h}, t_h)). \tag{4}$$

The last inequality is possible because of NTI and the fact that \bar{a} assigns uniform probability to all outcomes in \bar{A} .

Observe that

$$\begin{aligned} & U_h((\sigma_{-h}, \sigma'_h) | t_h) - U_h(\sigma | t_h) \\ &= \frac{\epsilon}{2} \sum_{t_{-h} \in T'_{-h}} q_h(t'_{-h} | t_h) [u_h(b', (t'_{-h}, t_h)) - u_h(b, (t'_{-h}, t_h))], \end{aligned}$$

$$\text{where } b' = \frac{1}{k'_h + 1} \bar{a} + \frac{k'_h}{k'_h + 1} a'_h \quad \text{and} \quad b = \frac{1}{\hat{k} + 1} \bar{a} + \frac{\hat{k}}{\hat{k} + 1} a_h.$$

From (2), (3) and (4), it follows that this expression is positive. This contradicts the hypothesis that the strategy profile σ is a Bayesian equilibrium.

Step 3. There cannot be an equilibrium σ that, in a state in T^* , induces rule (i) of the outcome function g where exactly $n - 1$ agents announce $c_i = (x_i^\alpha)_{\alpha \in D_i}$ and $k_i = 0$, while agent j announces something else. Suppose this happened under σ in state $t \in T^*$. Then, any $h \neq j$, of type t_h , can announce k'_h sufficiently high so that he becomes the winner in all states $\{t_h\} \times T'_{-h}$ where

$$T'_{-h} = \{t'_{-h} \in T_{-h} \mid \text{there is exactly one } i \neq h \text{ with } k_i > 0 \text{ or } c_i \neq (x_i^\alpha)_{\alpha \in D_i}\}.$$

These are precisely the states involving t_h where σ induces a non-unanimous report within case (i). Let $\sigma_h(t_h) = (\hat{t}_h, c_h, a_h, 0)$ and consider a strategy σ'_h , such that $\sigma'_h(t_h) = (\hat{t}_h, c_h, a'_h, k'_h)$ where k'_h is chosen to ensure that h wins the integer game in all states in $\{t_h\} \times T'_{-h}$, and a'_h is chosen to satisfy (4). By changing $\sigma_h(t_h)$ to $\sigma'_h(t_h)$, agent h of type t_h can shift some of the probability weight from \bar{a} to a'_h (in states in $\{t_h\} \times T'_{-h}$) and gain in terms of interim utility. This contradicts the hypothesis that σ is a Bayesian equilibrium.

Step 4. Finally, we claim that in any equilibrium of G under rule (i) where each agent i announces $c_i = (x_i^\alpha)_{\alpha \in D_i}$ and $k_i = 0$, agents do not use a deception α (in the sense that type t_i announces $\hat{t}_i = \alpha_i(t_i)$) where $f \circ \alpha \not\approx f$. Suppose not, i.e., there is an equilibrium under rule (i) in which a deception α is used where $f \circ \alpha \not\approx f$. Since f satisfies virtual monotonicity, there exists a test-agent i and two SCFs x_i^α and y_i^α satisfying (*). Therefore, type t_i of agent i has an incentive to deviate and change the second component of his

announcement to $(y_i^\alpha, (x_i^{\alpha'})_{\alpha' \in D_i \setminus \{\alpha\}})$, which is a contradiction. Thus, either no deception is used in equilibrium or the deception being used is such that $f \circ \alpha \approx f$. In either case, the equilibrium outcome is f^ϵ . This proves that our mechanism virtually implements f . \square

3.1. ϵ -Bayesian monotonicity and virtual monotonicity

If an SCF f is virtually implementable, then for every $\epsilon > 0$ there exists an SCF f^ϵ such that $d(f, f^\epsilon) < \epsilon$, and f^ϵ is Bayesian implementable. Thus, f^ϵ must satisfy the necessary conditions for Bayesian implementation, namely incentive compatibility and Bayesian monotonicity. By the same argument we used in the necessity part of the proof of Theorem 1, this yields the following necessary condition for virtual Bayesian implementation:

An SCF f satisfies ϵ -Bayesian monotonicity if for every deception α satisfying that $f \circ \alpha \not\approx f$, and every $\epsilon > 0$, there exists $i \in N$, $t_i \in T_i$, an incentive compatible SCF f^ϵ , with $d(f, f^\epsilon) < \epsilon$, and an SCF y such that

$$U_i(y \circ \alpha | t_i) > U_i(f^\epsilon \circ \alpha | t_i) \quad \text{while} \quad U_i(f^\epsilon | t'_i) \geq U_i(y_{\alpha_i(t_i)} | t'_i), \quad \forall t'_i \in T_i.$$

Evidently, this condition implies virtual monotonicity, since it concerns an SCF, f^ϵ , close to f , rather than some arbitrary SCF. However, the sufficiency part of Theorem 1 implies that in fact the two conditions are equivalent. A direct proof that (for an incentive compatible SCF) virtual monotonicity implies ϵ -Bayesian monotonicity is as follows.

Suppose f satisfies virtual monotonicity and is incentive compatible. Let α be such that $f \circ \alpha \not\approx f$. Let i, t_i, x and y satisfy (*). Define $f^\epsilon = (1 - \epsilon)f + \epsilon x$ and $y' = (1 - \epsilon)f + \epsilon y$. We claim that f^ϵ and y' satisfy the ϵ -Bayesian monotonicity condition for agent i of type t_i . Since

$$U_i(y' \circ \alpha | t_i) - U_i(f^\epsilon \circ \alpha | t_i) = \epsilon[U_i(y \circ \alpha | t_i) - U_i(x \circ \alpha | t_i)],$$

it follows from (*) that

$$U_i(y' \circ \alpha | t_i) > U_i(f^\epsilon \circ \alpha | t_i). \tag{5}$$

From (*) we also know that

$$U_i(x | t'_i) \geq U_i(y_{\alpha_i(t_i)} | t'_i), \quad \forall t'_i \in T_i.$$

Thus

$$(1 - \epsilon)U_i(f | t'_i) + \epsilon U_i(x | t'_i) \geq (1 - \epsilon)U_i(f | t'_i) + \epsilon U_i(y_{\alpha_i(t_i)} | t'_i), \quad \forall t'_i \in T_i.$$

Since, f is incentive compatible,

$$U_i(f | t'_i) \geq U_i(f_{\alpha_i(t_i)} | t'_i), \quad \forall t'_i \in T_i.$$

The last two inequalities imply that

$$U_i(f^\epsilon | t'_i) \geq U_i(y'_{\alpha_i(t_i)} | t'_i), \quad \forall t'_i \in T_i. \tag{6}$$

Since f and x are incentive compatible, so is f^ϵ . From (5) and (6) it now follows that f^ϵ and y' satisfy the ϵ -Bayesian monotonicity conditions for agent i of type t_i . Thus, virtual monotonicity implies ϵ -Bayesian monotonicity.

4. Type diversity

According to Theorem 1, virtual monotonicity is a necessary condition for an incentive compatible SCF to be virtually implemented. It is not possible, therefore, to achieve virtual Bayesian implementation through a weaker condition. However, this condition, like Bayesian monotonicity, is quite involved, and it is difficult to check whether a given SCF satisfies it or not. Consequently, it is not easy to see (apart from the necessity result) how weak the condition is. The aim of this section is to identify a simple condition that is easy to check, readily interpreted, and that implies that every SCF satisfies virtual monotonicity. In addition, we show that this condition holds generically in environments. In most environments, therefore, virtual monotonicity is vacuously satisfied by any incentive compatible SCF, and virtual implementation is as permissive as it can possibly be.

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 $A = \{a_1, \dots, a_K\}$ be the finite set of alternatives. Henceforth, we will find it convenient to identify a lottery, $x \in \Delta$, 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.,

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

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

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* (TD) if there do not exist $i \in N$, $t_i, t'_i \in T_i$, $t_i \neq t'_i$, $\beta \in \mathbb{R}_{++}$ and $\gamma \in \mathbb{R}$ such that

$$U_i(t_i) = \beta U_i(t'_i) + \gamma e,$$

where e is the unit vector in R^K .

Condition TD has a simple interpretation: it requires that the interim (cardinal) preferences over pure alternatives of different types of an agent be different. Note that this does not require ordinal preferences over pure alternatives to differ across types unless $|A| = 2$. Moreover, the condition only concerns constant SCFs. This condition appears in Section 4.2 of Abreu and Matsushima (1992b) as a simple way of ensuring A–M measurability. In a private values model, TD 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.¹⁰

Type diversity (TD) has the obvious virtue of being simple and easy to check, especially compared to virtual monotonicity or to other conditions in the literature, such as Bayesian

¹⁰ For private values environments, TD rules out the possibility that two types differ only in their beliefs.

monotonicity, A–M measurability or incentive consistency. Importantly, it is easy to see that in the space of preferences over pure alternatives, TD is satisfied *generically* if $|A| \geq 3$. In this sense, TD is indeed a very weak condition if $|A| \geq 3$.¹¹ It is, therefore, remarkable that TD is sufficient for virtual implementation of an incentive compatible SCF, as will be shown.

The following lemma provides a useful implication of TD from the point of view of implementation.

Lemma 1. *Suppose an environment \mathcal{E} satisfies TD and NTI. Then there exist constant SCFs $(l_i(t_i))_{t_i \in T_i}$ such that for every $i \in N$, $t_i, t'_i \in T_i$, $t_i \neq t'_i$,*

$$U_i(l_i(t_i)|t_i) > U_i(l_i(t'_i)|t_i).$$

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., $\bar{x}(t) = (1/K, \dots, 1/K)$ 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

$$U_i(x|t_i) > U_i(x'|t_i) \quad \text{and} \quad U_i(x'|t'_i) > U_i(x|t'_i). \quad (7)$$

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

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

where $p_k(t_i) = (U_i^k(t_i) - U_i^K(t_i))$, for $k = 1, \dots, K - 1$. Consider the indifference hyperplane through \bar{x} of agent i of type t'_i where $t'_i \neq t_i$:

$$H' = \left\{ (x_1, \dots, x_{K-1}) \in R_+^{K-1} \mid \sum_{k=1}^{K-1} p_k(t'_i)x_k = \bar{u}' \right\}.$$

Given NTI, 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 $U_i(t_i) = cU_i(t'_i) + \gamma e$, violating condition TD. 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 NTI) 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 (7). In the latter case, it is clear that we can choose two constant SCFs x and x' close to \bar{x} satisfying (7).

Given (7) 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). \square

It is now easy to show that every SCF satisfies virtual monotonicity.

¹¹ There is another reason why the weakness of condition TD 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.

Lemma 2. *Suppose an environment \mathcal{E} satisfies TD and NTI. Then every SCF satisfies virtual monotonicity.*

Proof. Suppose the environment satisfies NTI and TD. For any deception α , there exists $i \in N$ and $t_i \in T_i$ such that $\alpha_i(t_i) \neq t_i$. Given the SCFs, $(l_i(t_i))$ described in Lemma 1, define a pair of SCFs, y and x , where

$$y(t') = l_i(t_i) \quad \text{for all } t' \in T \quad \text{and} \quad x(t') = l_i(t'_i) \quad \text{for all } t' \in T.$$

By construction, x is incentive compatible. Note also that y is a constant SCF. Since, $\alpha_i(t_i) \neq t_i$,

$$U_i(y \circ \alpha | t_i) = U_i(l_i(t_i) | t_i) > U_i(l_i(\alpha_i(t_i)) | t_i) = U_i(x \circ \alpha | t_i).$$

Moreover,

$$U_i(x | t'_i) = U_i(l_i(t'_i) | t'_i) \geq U_i(l_i(t_i) | t'_i) = U_i(y_{\alpha_i(t_i)} | t'_i) \quad \text{for all } t'_i \in T_i.$$

Thus, for any deception α , and any SCF, condition (*) is satisfied with $i \in N$, $t_i \in T_i$, and SCFs y and x chosen as above. \square

Applying Theorem 1, we have the following corollary.

Corollary 1. *In an environment satisfying NTI and TD, every incentive compatible SCF is virtually Bayesian implementable.*

This result can also be proved directly by constructing a mechanism based on the constant SCFs, $(l_i(t_i))$. We have constructed such a mechanism without relying on the notion of deceptions, and it is available upon request.

To illustrate TD, or its implications drawn out in Lemma 1, see Fig. 1, 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, the direction of preference increases towards a_2). This figure is very similar to one that could be drawn to illustrate the power of virtual Nash implementation in the complete information case. We can illustrate TD in this figure only because condition TD concerns 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, as shown in Corollary 1.

Conditions TD and NTI imply 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 TD and NTI are satisfied, the SCF $f^* = (1/n) \sum_{i \in N} l_i(t_i)$, 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, under TD and NTI, every SCF is incentive consistent, a condition which plays a crucial role in Duggan’s (1997) sufficiency result.

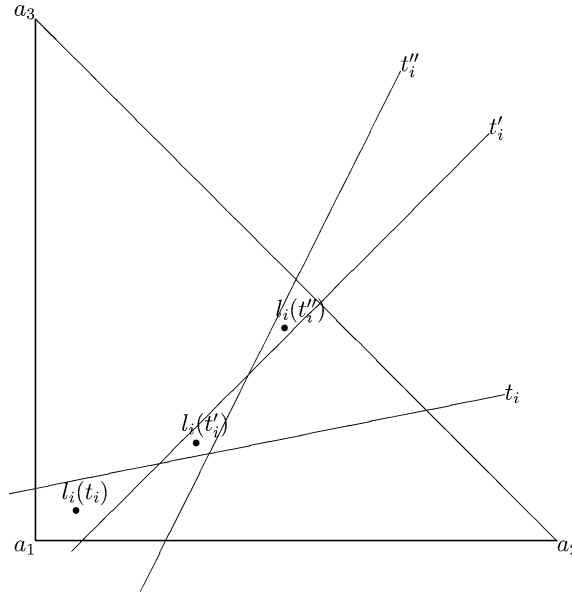


Fig. 1. Type diversity.

5. An example

We now consider an important example (Example 3 in Palfrey and Srivastava, 1989b) to show the permissiveness of our conditions and to clarify the comparison between our results and related results in the literature. Palfrey and Srivastava use this example to show the difficulties 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:

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

Since $q^2 > 0.5$, this implies that $U_i^b(t_i) > U_i^a(t_i)$ for all i and $t_i \in T_i$; ordinal preferences do not vary across types.

Consider the “majoritarian” SCF, x^* , 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 . This SCF does not satisfy Bayesian

monotonicity. To see this consider the deception, $\alpha_i(t_i) = t_b$ for all i and t_i . Of course, $x^* \circ \alpha \not\approx x^*$. Since $x^* \circ \alpha(t) = b$ for all t and $U_i^b(t_i) > U_i^a(t_i)$ for all i and $t_i \in T_i$, there does not exist y such that $U_i(y \circ \alpha | t_i) > U_i(x^* \circ \alpha | t_i)$ for any i and t_i . 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.

We show now that every SCF satisfies virtual monotonicity in this example. To begin with, consider the majoritarian SCF x^* . Let

$$\lambda = (1 - q)^2 < 1/2$$

and define x as

$$x(t) = (1 - \lambda)x^* + \lambda z(t)$$

where z makes the choice least preferred by the majority, i.e., $z(t) = a$ if $x^*(t) = b$ and $z(t) = b$ if $x^*(t) = a$.

The SCF x^* , choosing always the best alternative, yields a utility of 1 in each state, while z , choosing the worst alternative, yields 0 in each state. Thus, x yields a utility of $(1 - \lambda)$ in each state, and the interim utility of each agent of each type is therefore $(1 - \lambda)$, i.e.,

$$U_i(x|t_a) = U_i(x|t_b) = (1 - \lambda) \quad \text{for all } i.$$

Notice that x chooses the best alternative with probability $1 - \lambda$ and the worst one with probability λ . Consider a unilateral deception from truth-telling by agent i . This does not change the outcome if i is not pivotal. But in each state where i is pivotal, this causes the outcome to be the best one with probability λ and the worst one with probability $(1 - \lambda)$. Since $\lambda < 1/2$, this results in an interim utility less than $(1 - \lambda)$. Thus, x is incentive compatible.

In fact, the argument of the previous paragraph can be extended to show that any deception α applied to x results in some agent receiving an interim utility strictly less than $(1 - \lambda)$:

$$\text{If } \alpha(t) \neq t \text{ for some } t, \text{ then there exists } i \text{ such that } U_i(x \circ \alpha | t_b) < (1 - \lambda). \quad (8)$$

Of course, in each state t , $u_i(x \circ \alpha, t) \leq 1 - \lambda$. To prove (8), it suffices to show that there exists i and t , with $t_i = t_b$ such that $x \circ \alpha(t) \neq x(t)$. There are two cases to consider.

Case 1. Suppose there exists i such that $\alpha_i(t_b) = t_a$. If for some $j \neq i$, $\alpha_j(t_a) = t_a$, this implies that for t such that $t_i = t_b$ and $t_j = t_a$, $t_k = t_b$,

$$x \circ \alpha(t) = (1 - \lambda)a + \lambda b \quad \text{while} \quad x(t) = (1 - \lambda)b + \lambda a.$$

If $\alpha_j(t_a) = t_b$ for both $j \neq i$, then for t such that $t_i = t_b$ and $t_j = t_a$ for $j \neq i$, we have

$$x \circ \alpha(t) = (1 - \lambda)b + \lambda a \quad \text{while} \quad x(t) = (1 - \lambda)a + \lambda b.$$

Thus, in either case, $U_i(x \circ \alpha | t_b) < 1 - \lambda$.

Case 2. Suppose α is such that all agents report truthfully when they are of type t_b . Then, there exists j such that $\alpha_j(t_a) = t_b$. Consider $i \neq j$ and the state t where $t_i = t_b$, $t_j = t_a$, $t_k = t_a$. Since $\alpha_i(t_b) = t_b$,

$$x \circ \alpha(t) = (1 - \lambda)b + \lambda a \quad \text{while} \quad x(t) = (1 - \lambda)a + \lambda b,$$

and $U_i(x \circ \alpha | t_b) < 1 - \lambda$.

Define y to be the constant SCF which prescribes b in each state. Clearly, for each i ,

$$U_i(y|t_a) = q^2, \quad U_i(y|t_b) = 1 - (1 - q)^2.$$

Since y is constant, this means that for any deception α and any type t_i , $y_{\alpha_i(t_i)}(t) = y(t)$ for all t . In particular,

$$U_i(y \circ \alpha | t_b) = 1 - (1 - q)^2 = (1 - \lambda).$$

This, along with (8), implies that there exists i such that

$$U_i(y \circ \alpha | t_b) > U_i(x \circ \alpha | t_b). \quad (9)$$

Recall that $U_i(x|t_a) = U_i(x|t_b) = (1 - \lambda) = 1 - (1 - q)^2 > q^2$. Thus,

$$U_i(x|t_a) \geq U_i(y_{\alpha_i(t_i)} | t_a) \quad \text{and} \quad U_i(x|t_b) \geq U_i(y_{\alpha_i(t_i)} | t_b).$$

Given (9), this implies that x and y satisfy our condition, for any deception α .

In fact, a slight modification of the arguments above shows that for any deception α , one can choose x to be $x = (1 - \beta)x^* + \beta z$ for $\beta > 0$ arbitrarily small (together with the same y , where the test-agent is always a type t_b , chosen as above).

Since for every deception α , $x^* \not\approx x^* \circ \alpha$, the desired preference reversal has been found and x^* satisfies virtual monotonicity. Moreover, since we have found a preference reversal for every deception, it follows that every SCF in this example satisfies virtual monotonicity. Since the environment clearly satisfies NTI, it follows from Theorem 1 that every incentive compatible SCF is virtually Bayesian implementable.

Checking for virtual monotonicity may sometimes be cumbersome. It may then be easier to check that the environment satisfies TD. Actually, this environment does not. However, TD is easily satisfied if we modify this example to have a third alternative c . For instance, 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. Note that

$$\begin{aligned} U_i^a(t_a) &= 1 - q^2, & U_i^b(t_a) &= q^2, & U_i^c(t_a) &= 0, \\ U_i^a(t_b) &= (1 - q)^2, & U_i^b(t_b) &= 1 - (1 - q)^2, & U_i^c(t_b) &= 0. \end{aligned}$$

Clearly, TD is now satisfied. Thus, Corollary 1 applies to this modified example; any incentive compatible SCF is virtually implementable. Since TD holds, every SCF satisfies A–M measurability and incentive consistency in this environment with three alternatives. However, the results of Abreu and Matsushima (1992b) and Duggan (1997) cannot be applied to any non-constant SCF even then. 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.

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.

6. Concluding remarks

We conclude with a few remarks that apply to both Theorem 1 and Corollary 1.

Remark 1. Theorem 1 and the discussion in Section 5 show that, in some environments, virtual implementation in Bayesian equilibrium is more permissive than virtual implementation in iterative undominated strategies or exact implementation in undominated Bayesian equilibrium.

Remark 2. In environments violating TD, virtual implementation may be impossible, through a general violation of virtual monotonicity. 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 3. In this paper we have used the traditional notion of implementability which ignores mixed strategies and imposes no restrictions on the nature of the mechanism. If one were allowed to use small transfers, and one were to insist on ‘regular mechanisms’ and virtual Bayesian implementation using mixed strategies, virtual monotonicity is not sufficient; A–M measurability then becomes necessary, as shown by Abreu and Matsushima (1992b).

Remark 4. Our mechanism makes a use of the integer game that is a bit different from the usual. In the traditional mechanisms, the integer game cannot be triggered in equilibrium because (given some version of the no-veto-power assumption) different agents would attempt to get different alternatives. Our mechanism works even if all agents have identical preferences (as in the example in Section 5) because the set of SCFs achievable by a winner of the integer game is ‘open’: a higher integer shifts more of the probability weight from \bar{a} to the chosen outcome in \bar{A} . Of course, this makes essential use of NTI and the fact that virtual rather than exact implementation is being sought.

The integer game used in our mechanism can be replaced by a modulo game, thereby making the mechanism compact. However, it is not clear to us whether it can be replaced by a *finite* modulo game. Dutta and Sen (1994) showed the necessity of infinite mechanisms for *exact* Bayesian implementation even in finite environments. (While the example used by Dutta and Sen, 1994 does not satisfy TD, it can be modified by adding a third alternative which yields 0 utility to each agent in each state to satisfy TD. It is easy to check that their result continues to apply to this modified example.) Yet TD implies A–M measurability, and under A–M’s conditions, virtual implementation can be accomplished through a regular mechanism. Clarifying this aspect of the theory should be the object of further research.

Remark 5. Given the very positive results reported in this paper, one may wonder how much they depend on the expected utility assumption. To the extent that this is just an approximation of more realistic preferences, it would be desirable that the assumption of expected utility be not a crucial one for the theory. Indeed, it is not. One can easily see that all our conclusions extend to preferences over lotteries that have lower contour sets that are not nested in two different environments (many monotonic preferences in the sense of first-order stochastic dominance will satisfy this). Reflection on Fig. 1 should suffice to convince the reader of this assertion: the relevant indifference surfaces yielding non-nested lower contour sets in the interior of the probability simplex 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).

Remark 6. A connection of our analysis with subgame perfect implementation is of interest. Note that virtual monotonicity differs from Bayesian monotonicity much like the necessary condition for subgame perfect implementation (Moore and Repullo, 1988; Abreu and Sen, 1990) differs from Maskin monotonicity: while the latter requires preference reversals around the SCF, the former can be satisfied by reversals over arbitrary pairs (provided they are connected to the SCF by a suitable sequence). It remains as an open question whether a similar condition can be found for implementation with dynamic mechanisms in incomplete information environments.

Acknowledgments

We thank several anonymous referees, whose careful reading improved the presentation substantially. We acknowledge numerous conversations with O. Volij on the subject. We also thank S. Baliga, J. Bergin, L. Corchon, M. Jackson, G. Mailath, E. Maskin, T. Palfrey and P. Reny for helpful comments. Conference and seminar audiences at Venice, Pompeu Fabra, Stanford, Arizona, Carlos III, Brown, Jerusalem, Minnesota, Caltech, Maryland, Chicago, Pittsburgh, Institute for Advanced Study, Sao Paulo, Princeton, Rutgers, Purdue Northwestern and NYU also had useful suggestions. Both authors acknowledge support from NSF grant SES-0133113 and Salomon research awards from Brown University. Serrano also thanks the Institute for Advanced Study for its warm hospitality and acknowledges research support from Deutsche Bank and the Alfred P. Sloan Foundation.

References

- Abreu, D., Matsushima, H., 1992a. Virtual implementation in iteratively undominated strategies: complete information. *Econometrica* 60, 993–1008.
- Abreu, D., Matsushima, H., 1992b. Virtual implementation in iteratively undominated strategies: incomplete information. Mimeo. Princeton University.
- Abreu, D., Sen, A., 1990. Subgame perfect implementation: a necessary and almost sufficient condition. *J. Econ. Theory* 50, 285–299.
- Abreu, D., Sen, A., 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 Econ. Behav.* 27, 173–183.
- Bergin, J., Sen, A., 1998. Extensive form implementation in incomplete information environments. *J. Econ. Theory* 80, 222–256.
- Brusco, S., 1995. Perfect Bayesian implementation. *Econ. Theory* 5, 419–444.
- Chakravorti, B., 1992. Efficiency and mechanisms with no regret. *Int. Econ. Rev.* 33, 45–59.
- Duggan, J., 1997. Virtual Bayesian implementation. *Econometrica* 65, 1175–1199.
- Dutta, B., Sen, A., 1994. Bayesian implementation: the necessity of infinite mechanisms. *J. Econ. Theory* 64, 130–141.
- Jackson, M., 1991. Bayesian implementation. *Econometrica* 59, 461–477.
- Maskin, E., 1977. Nash equilibrium and welfare optimality. Mimeo. MIT. Published in 1999 in: *Rev. Econ. Stud.* 66, 23–38.
- Matsushima, H., 1988. A new approach to the implementation problem. *J. Econ. Theory* 45, 128–144.
- Matsushima, H., 1993. Bayesian monotonicity with side payments. *J. Econ. Theory* 59, 107–121.
- Moore, J., Repullo, R., 1988. Subgame perfect implementation. *Econometrica* 56, 1191–1220.
- Mueller, E., Satterthwaite, M., 1977. The equivalence of strong positive association and strategy-proofness. *J. Econ. Theory* 14, 412–418.
- Myerson, R., 1989. Mechanism design. In: Eatwell, J., Milgate, M., Newman, P. (Eds.), *The New Palgrave: Allocation, Information, and Markets*. Norton, New York.
- Palfrey, T., Srivastava, S., 1987. On Bayesian implementable allocations. *Rev. Econ. Stud.* 54, 193–208.
- Palfrey, T., Srivastava, S., 1989a. Implementation with incomplete information in exchange economies. *Econometrica* 57, 115–134.
- Palfrey, T., Srivastava, S., 1989b. Mechanism design with incomplete information: a solution to the implementation problem. *J. Polit. Economy* 97, 668–691.
- Palfrey, T., Srivastava, S., 1993. *Bayesian Implementation*. Harwood Academic Publishers, New York.
- Postlewaite, A., Schmeidler, D., 1986. Implementation in differential information economies. *J. Econ. Theory* 39, 14–33.
- Saijo, T., 1987. On constant Maskin monotonic social choice functions. *J. Econ. Theory* 42, 382–386.
- Serrano, R., Vohra, R., 2001. Some limitations of virtual Bayesian implementation. *Econometrica* 69, 785–792.