Econometrica, Vol. 69, No. 3 (May, 2001), 785–792

SOME LIMITATIONS OF VIRTUAL BAYESIAN IMPLEMENTATION

By Roberto Serrano and Rajiv Vohra[1]

1. INTRODUCTION

As is well known, Maskin monotonicity (Maskin (1977)) is a necessary condition for a social choice correspondence to be Nash implementable. In economic environments with at least three agents and a private good, this condition is also sufficient for Nash implementation. Remarkably, Abreu and Sen (1991) and Matsushima (1988) showed that this condition can be entirely dispensed with if the following modifications are introduced: first, random allocations are permitted; and second, the notion of implementation is weakened to virtual Nash implementation, requiring the implementation of a social choice function that is arbitrarily close to the given function.

While Nash implementation pertains to environments with complete information, there is also an extensive literature on Bayesian Nash implementation in environments with incomplete information; see, for example, Postlewaite and Schmeidler (1986), Palfrey and Srivastava (1987, 1989), Mookherjee and Reichelstein (1990), and Jackson (1991). It has been shown that, apart from incentive compatibility (which is clearly a necessary condition for implementation in any solution concept), a Bayesian implementable social choice set must also satisfy Bayesian monotonicity—a suitable analog of Maskin monotonicity. In economic environments with at least three agents, these conditions are also sufficient; see, for example, Jackson (1991). Bayesian monotonicity is an involved and sometimes quite strong condition (as will be apparent from Example 1 below). It is therefore natural to examine whether the virtual approach can help dispense with it, or replace it with weaker conditions.

Besides incentive compatibility, two sufficient conditions for virtual Bayesian implementation have been identified in the literature. Abreu and Matsushima (1992b), in analyzing virtual implementation in iteratively undominated strategies, introduce a measurability condition (referred to as A-M measurability in the sequel), which they show is necessary for implementation in their solution concept. Further, they show that under weak domain restrictions this is sufficient for virtual implementation in iteratively undominated strategies and, a fortiori, in the weaker notion of implementation in Bayesian Nash equilibrium. They also show that this condition is necessary for virtual Bayesian Nash implementation if one insists on 'regular game forms', which rule out unattractive features such as integer games, and is in this sense close to being necessary. On the other hand, Duggan (1997) introduces a condition termed incentive consistency, which he argues is weak under standard topological and informational assumptions. Our aim is to clarify the strength of these sufficient conditions. We do so with the help of two examples.

---

[1] Part of the material in this paper circulated as Serrano and Vohra (1999). We thank John Duggan, a co-editor, and an anonymous referee for their comments. Both authors acknowledge Salomon research awards from Brown University. Serrano also acknowledges an Alfred P. Sloan Foundation research fellowship.

Example 1 concerns an economy in which a social choice function can be virtually implemented if and only if it is constant (with respect to the information state). Thus, only constant social choice functions satisfy Bayesian monotonicity, A-M measurability, or incentive consistency. However, in this economy there are many interesting (incentive compatible) social choice functions that are *not* constant. Indeed, in this example, any social choice function that satisfies interim individual rationality and interim (incentive) efficiency is *not* constant. This shows that in environments with incomplete information it is not generally possible to virtually implement every incentive compatible social choice function. Virtual implementation imposes nontrivial restrictions on a social choice function beyond incentive compatibility. A sufficient condition for virtual implementation (e.g., incentive consistency, A-M measurability) cannot, therefore, by innocuous, *at least in general environments*.[2]

In our second example, *every* social choice function is virtually implementable: Bayesian monotonicity is satisfied by almost every social choice function. However, only constant functions satisfy A-M measurability or incentive consistency. Thus, neither condition is necessary for virtual implementation. And virtual Bayesian implementation may be more permissive (admitting the use of nonregular game forms) than virtual implementation in iteratively undominated strategies.[3]

We conclude that in environments with incomplete information the virtual approach has some limitations with respect to what it delivers in the complete information case. Moreover, the task of identifying weak sufficient conditions (in the sense of being close to necessary) for virtual Bayesian implementation remains.

2. NOTATION AND DEFINITIONS

We shall consider exchange economies with a finite set of agents $N = \{1, \dots, n\}$. Let T_i denote the (finite) set of agent i's types and let $T = \prod_{i \in N} T_i$. Each agent has a prior probability distribution q_i defined on T. We shall assume that all agents agree on zero probability states. Let T^* be the set of states with positive probability. The consumption set for each agent in each state is $\mathbb{R}^l_+$, and the aggregate endowment ω is constant across states. The set of feasible allocations in each state is denoted

$$A = \left\{ x = (x_i) \in \mathbb{R}^{ln}_+ \,\middle|\, \sum_i x_i \leq \omega \right\}.$$

Let $\mathscr{A}$ denote the Borel σ-algebra on A and Δ denote the set of probability measures on $(A, \mathscr{A})$.

A *social choice function* (SCF) is a function $f: T \mapsto \Delta$. Let $f_i(s)$ denote agent i's (possibly random) commodity bundle in state s according to f. We shall say that 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).

---

[2] With domain restrictions, such as the possibility of transfers, it is possible that incentive compatibility alone may suffice; see, for example, Matsushima (1993).

[3] In contrast, recall that in the complete information framework, Abreu and Matsushima (1992a) provide a significant improvement over the Abreu and Sen (1991) result by showing that nothing is lost by using the more appealing solution concept of iteratively undominated strategies.

The Bernoulli utility of agent i for allocation x in state s is $u_i(x,s)$.[4] Abusing notation slightly, $u_i(f,s)$ 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

$$U_i(f|s_i) \equiv \sum_{s'_{-i} \in T_{-i}} q_i(s'_{-i}|s_i) u_i(f,(s'_{-1},s_i)).$$

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

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

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

Consider the following metric on SCFs:

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

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 $\alpha_i: T_i \mapsto T_i$. A deception is said to be *compatible* 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)).

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

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

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

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

Along with incentive compatibility, the next two conditions have been proved to be sufficient for virtual Bayesian implementation in certain classes of environments (see Abreu and Matsushima (1992b) and Duggan (1997)).

An SCF f is said to be *incentive consistent* if there exists an incentive compatible SCF, f^*, such that for any deception α, $f \circ \alpha \neq f$ implies that α is not a Bayesian equilibrium of the direct mechanism for f^*.

For the next definition we shall need some additional notation. Let P_i be a partition of T_i, where $\phi_i(s_i)$ denotes the element of P_i that contains s_i. Let $P = \prod_i P_i$. An SCF f is said to be *measurable with respect to P* if for every $i \in N$,

$$\phi_i(s_i) = \phi_i(s'_i) \quad \text{implies that} \quad f(s) = f(s_{-i}, s'_i) \qquad \text{for all} \quad s_{-i} \in T_{-i}.$$

Given P_{-i}, define the partition

$$R_i(P_{-i}) = \{\rho_i(s_i, P_{-i})|s_i \in T_i\}$$

---

[4] Although we will consider only economies without externalities, it is notationally simpler to define agent i's utility in terms of an allocation rather than a commodity bundle.

where $\rho_i(s_i, P_{-i})$ is the set of all $s'_i \in T_i$ such that for all SCFs f and h that are measurable with respect to $\{T_i\} \times P_{-i}$,

$$U_i(f|s_i) \geq U_i(h|s_i) \quad \text{if and only if} \quad U_i(f|s'_i) \geq U_i(h|s'_i).$$

Let $P_i^0 = \{T_i\}$ for all i and define recursively, for every $i \in N$ and $k > 0$,

$$P_i^k = R_i(P_{-i}^{k-1})$$

and let $P_i^* = P_i^L$ where L is such that $P_i^k = P_i^L$ for all $k \geq L$ and all $i \in N$.

An SCF is said to be *A-M measurable* if it is measurable with respect to P^*.

3. RESULTS

EXAMPLE 1: The set of agents is $N = \{1,2,3,4\}$. There is a single commodity and all consumers have 1 unit of endowment in each state. The sets of types are $T_k = \{t_k, t'_k, t''_k\}$ for $k = 1,2$, while $T_j = \{t_j, t'_j\}$ for $j = 3,4$. There are only three states that arise with positive probability: $T^* = \{t, t', t''\}$, where $t = (t_1, t_2, t_3, t_4)$, $t' = (t'_1, t'_2, t'_3, t'_4)$, and $t'' = (t''_1, t''_2, t'_3, t'_4)$. Agents 1 and 2 are fully informed, so that for $k = 1,2$, the posterior probability distributions are:

$$q_k(t|t_k) = q_k(t'|t'_k) = q_k(t''|t''_k) = 1, \qquad k = 1,2.$$

Agents 3 and 4 are fully informed only when they are of type t_j, i.e., for $j = 3,4$, $q_j(t|t_j) = 1$, but

$$q_3(t'|t'_3) = 0.25, \qquad q_3(t''|t'_3) = 0.75,$$

$$q_4(t'|t'_4) = 0.75, \qquad q_4(t''|t'_4) = 0.25.$$

The utility functions are as follows:

$$u_i(x, s) = \sqrt{x_i} \qquad \forall s \in T, \quad \forall i \in N.$$

CLAIM 1.1: *Consider the exchange economy in Example* 1. *Let f be an SCF such that for some $s \in T^*$, $s \neq t$, $f(s) \neq f(t)$. Then f is not virtually Bayesian implementable.*

PROOF: Let f be such an SCF and let α be the deception such that for all $i \in N$ and for all types s_i, $\alpha_i(s_i) = t_i$. For an SCF h let h^t denote the constant SCF where $h^t(s) = h(t)$ for every $s \in T$. Since $\alpha(s) = t$ for all $s \in T$, $h \circ \alpha = h^t$. This means that

$$U_i(h \circ \alpha|s_i) = U_i(h^t|s_i) = U_i(h^t|t_i) = U_i(h|t_i),$$

where the second last equality uses the fact that the Bernoulli utility function of each agent is independent of the state and the last equality follows from the fact that t is a common knowledge state. Notice also that $U_i(h_{\alpha_i(s_i)}|t_i) = U_i(h|t_i)$. Thus we have shown that for all $i \in N$ and $s_i \in T_i$, and any SCF h,

$$U_i(h \circ \alpha|s_i) = U_i(h_{\alpha_i(s_i)}|t_i) = U_i(h|t_i).$$

Since $f \neq f \circ \alpha$, this implies that f is not Bayesian monotonic. And it follows (for example, from Theorem 1 in Jackson (1991)) that f is not exactly implementable.

Now suppose, contrary to our claim, that f is virtually implementable, i.e., there exists f' arbitrarily close to f which is exactly implementable. Since Bayesian monotonicity is a necessary condition for exact implementation, it follows from the previous paragraph that f' is constant. But a nonconstant function, such as f, cannot be approximated by a constant one, which contradicts our supposition. *Q.E.D.*

CLAIM 1.2: *Every selection f from the interim individually rational interim efficient correspondence in Example* 1 *is nonconstant.*

PROOF: Let f be a selection from such a correspondence. It is easy to see that $f(t) = (1,1,1,1)$. However, efficient risk sharing among agents 3 and 4 in states t' and t'' clearly implies nonzero trade, i.e., $f(t') \neq (1,1,1,1)$ and $f(t'') \neq (1,1,1,1)$. Thus, by Claim 1.1, f is not virtually Bayesian implementable. *Q.E.D.*

REMARK 1: The example requires only one informed agent. Having two of them, though, guarantees incentive compatibility because it makes the environment one of nonexclusive information. More general domains of economies in which the same problem persists have been identified in Serrano and Vohra (1999).

REMARK 2: By introducing the following simple modification, the posterior beliefs in Example 1 can be derived from a common prior: let $q_j(t'|t_j') = q_j(t''|t_j') = 0.5$ for $j = 3,4$, $u_3(x,t') = \sqrt{x_3}$, $u_3(x,t'') = 3\sqrt{x_3}$, $u_4(x,t') = 3\sqrt{x_4}$, and $u_4(x,t'') = \sqrt{x_4}$. The important feature of this example is the difference in interim preferences of the second types of agents 3 and 4. For our purpose, it does not matter whether this difference arises from different probability assessments or from different cardinal representations of the utility functions.

REMARK 3: Proposition 4 in Duggan (1997) shows that incentive consistency is a weak condition in the following sense. Under standard topological conditions and under the assumption of private values, every *value-measurable* SCF is incentive consistent. An SCF f is said to be value-measurable if, for all $s, s' \in T$, $f(s) \neq f(s')$ implies that for some i, $u_i(\cdot, s) \neq u_i(\cdot, s')$. Recall that Example 1 satisfies private values and, as pointed out in Remark 2, the utility functions in each state can be multiplied by arbitrary constants without changing the conclusions of the example. Thus, Proposition 4 in Duggan (1997) and Claims 1.1 and 1.2 imply that in the definition of value-measurability in Duggan (1997), $u_i(\cdot, s)$ must be interpreted to mean the equivalence class of positive affine transformations of a given Bernoulli utility function $u_i(\cdot, s)$. With this interpretation in mind, Example 1 demonstrates that value-measurability cannot be dropped from Proposition 4 in Duggan (1997), and that there exist well-behaved economies with interesting SCFs that do not satisfy value-measurability.

Our next example concerns an economy in which *every* SCF is virtually implementable, but only constant SCFs satisfy A-M measurability, or incentive consistency.[5] Thus, neither AM-measurability nor incentive consistency are necessary for virtual implementation. In environments like those in Example 2, both are strong conditions, certainly much stronger than Bayesian monotonicity.

---

[5] Duggan (1997) has already exhibited an example of an SCF that is implementable but fails to be A-M measurable.

Example 2: Let $N = \{1, 2, 3\}$. There is a single commodity and the aggregate endowment is 1 unit in each state. The sets of types are $T_i = \{t_i, t'_i\}$ for $i = 1, 2$, while $T_3 = \{t_3\}$. There are two states in T^*, denoted t and t', where $t = (t_1, t_2, t_3)$ and $t' = (t'_1, t'_2, t_3)$. Agents 1 and 2 are fully informed, so that for $i = 1, 2$: $q_i(t \mid t_i) = 1$, $q_i(t' \mid t'_i) = 1$. Agent 3 is uninformed and

$$q_3(t \mid t_3) = q_3(t' \mid t_3) = 0.5.$$

The utility functions are as follows:

$$u_i(x, s) = x_i, \qquad \text{for} \quad i = 1, 2, \qquad \text{for all} \quad s \in T^*.$$

However,

$$u_3(x, t) = x_3, \qquad u_3(x, t') = \sqrt{x_3}\,.$$

Thus agent 3's preferences over lotteries differ across states t and t'.

Claim 2.1: *If f is an SCF that satisfies A-M measurability in Example* 2, *f must be constant over T^*, i.e., $f(t) = f(t')$.*

Proof: This follows because the two types of agents 1 and 2 have identical preferences, while there is only one type of agent 3. Therefore, the final partition in the A-M algorithm is the coarse one, and any f measurable with respect to that must be constant. *Q.E.D.*

Claim 2.2: *If f is an SCF that satisfies incentive compatibility and incentive consistency in Example* 2, *f must be constant over T^*, i.e., $f(t) = f(t')$.*

Proof: Note that the environment in Example 2 is one of nonexclusive information with three agents, and therefore, every SCF is equivalent to an incentive compatible SCF. Suppose then that f is incentive compatible and incentive consistent. Clearly, truth-telling is not the only equilibrium in the direct mechanism corresponding to f. In particular, consider the deception α, where $\alpha_i(s_i) = t_i$ for all $s_i \in T_i$ and $i = 1, 2$. Since both agents 1 and 2 have the same preferences in each state, it is easy to see that α is a Bayesian Nash equilibrium of the direct mechanism corresponding to f. Indeed, this argument holds for *any* incentive compatible SCF. Thus there does not exist an incentive compatible f^* such that α is not an equilibrium of the direct mechanism corresponding to f^*. Therefore, incentive consistency of f implies that $f \circ \alpha = f$, i.e., f is constant over T^*. *Q.E.D.*

Claim 2.3: *Every SCF f in Example* 2 *is virtually Bayesian implementable.*

Proof: Since Example 2 concerns an economic environment, as in Jackson (1991, Theorem 1), there is no loss of generality in assuming that $f(s) = 0$ for all $s \notin T^*$. Recall that incentive compatibility is trivially satisfied because of nonexclusive information. Thus, by Jackson (1991, Theorem 1), f is Bayesian implementable if and only if it satisfies Bayesian Monotonicity. To prove that f is virtually Bayesian implementable it suffices to show that one can find a sequence of Bayesian monotonic SCFs converging to f. Observe also that any SCF is arbitrarily close to one that assigns to each consumer, in each state, both the commodity bundle 0 and the commodity bundle 1 with positive probabilities.

Accordingly, it suffices to show that Bayesian Monotonicity is satisfied by any SCF f with the following properties:

- $f(s) = 0$ for all $s \notin T^*$.
- $f_3(s)$ assigns positive probability to 0 and to 1 for every $s \in T^*$.

Of course, any constant f is implementable. We, therefore, need to consider an f that is not constant. To verify Bayesian monotonicity we will show that the necessary preference reversal is satisfied by agent 3. More precisely, we will show that for any deception α, which is not the identity function, there exists an SCF h such that

$$(*) \qquad U_3(h \circ \alpha) > U_3(f \circ \alpha) \quad \text{while} \quad U_3(f) \geq U_3(h).$$

There are four kinds of deceptions, which we consider in turn.

(i) α is an incompatible deception in the sense that $\alpha(s) \notin T^*$ for some $s \in T^*$. Suppose $\alpha(t) = (t'_1, t_2)$, while $\alpha(t') = t'$. Define h to be an SCF such that $h(s) = f(s)$ for all $s \neq (t'_1, t_2)$ and $u_3(h, (t'_1, t_2)) > 0$. It is now easy to see that $(*)$ is satisfied. Clearly, the same kind of construction is possible for any incompatible deception.

(ii) $\alpha(s) = t$ for all $s \in T$. Define c to be a deterministic allocation such that $u_3(c, t) = u_3(f, t)$, i.e., c assigns to agent 3 his certainty equivalent of lottery $f(t)$ in state t. Consider the deterministic SCF h, where $h(t) = c$ and $h(t') = 0$. Clearly, $U_3(f) > U_3(h)$, but $U_3(f \circ \alpha) < U_3(h \circ \alpha)$. The last inequality follows from the fact that f is not deterministic and agent 3's utility function in state t' is strictly concave, while it is linear in state t.

(iii) $\alpha(s) = t'$ for all $s \in T$. Define h such that $h(s) = 0$ for all $s \neq t'$ and $h_3(t')$ places a higher probability mass on 1 than does $f_3(t')$, so that $u_3(h, t') > u_3(f, t')$. (This is possible because f assigns a strictly positive probability mass to 0.) Moreover, since $u_3(f, t) > 0$, this can be done to ensure that $u_3(h, t') < u_3(f, t) + u_3(f, t')$. But then it is easy to see that $(*)$ is satisfied.

(iv) $\alpha(t) = t'$ and $\alpha(t') = t$. Consider the function h, where $h(t) = c$ and $h(s) = f(s)$ for all $s \neq t$. Then, $U_3(f) \geq U_3(h)$, but $U_3(f \circ \alpha) < U_3(h \circ \alpha)$. *Q.E.D.*

Remark 4: This example does not satisfy private values, but does satisfy the assumption of 'best element private values.'[6] However, in this example, every SCF is value-measurable. This shows that in Proposition 4 in Duggan (1997), the assumption of private values cannot be weakened to best element private values.

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

and

Department of Economics, Brown University, Providence, RI 02912, U.S.A.; Rajiv_Vohra@brown.edu; http://econ.pstc.brown.edu/~rvohra

REFERENCES

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

——— (1992b): "Virtual Implementation in Iteratively Undominated Strategies: Incomplete Information," mimeo, Princeton University.

---

[6] See Duggan (1997) for a precise definition, where this assumption is used in the main sufficiency result.

ABREU, D., AND A. SEN (1991): "Virtual Implementation in Nash Equilibrium," *Econometrica*, 59, 997–1021.

DUGGAN, J. (1997): "Virtual Bayesian Implementation," *Econometrica*, 65, 1175–1199.

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.

——— (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.

PALFREY, T., AND S. SRIVASTAVA (1987): "On Bayesian Implementable Allocations," *Review of Economic Studies*, 54, 193–208.

——— (1989): "Implementation with Incomplete Information in Exchange Economies," *Econometrica*, 57, 115–134.

POSTLEWAITE, A., AND D. SCHMEIDLER (1986): "Implementation in Differential Information Economies," *Journal of Economic Theory*, 39, 14–33.

SERRANO, R., AND R. VOHRA (1999): "On the Impossibility of Implementation under Incomplete Information," Working Paper No. 99-10, Brown University.