DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA 91125

MECHANISM DESIGN WITH INCOMPLETE INFORMATION: A SOLUTION TO THE IMPLEMENTATION PROBLEM

Thomas R. Palfrey
California Institute of Technology

Sanjay Srivastava
Carnegie-Mellon University

SOCIAL SCIENCE WORKING PAPER 658

December 1987

Abstract

The main result of this paper is that the multiple equilibrium problem in mechanism design can be avoided in private value models if agents do not use weakly dominated strategies in equilibrium. We show that in such settings, any incentive compatible allocation can be made the unique equilibrium outcome to a mechanism. We derive a general necessary condition for unique implementation which implies that the positive result for private value models applies with considerably less generality to common value settings and to situations in which an agent's information does not index the agent's preferences.

1. Introduction

Institutions play a fundamental role in the organization of economic, political, and social activity. A central problem in the theory of institutions is the characterization of outcomes which can be achieved by institutions. Mechanism design theory studies precisely this problem.

A basic principle of mechanism design with incomplete information is that any outcome which is a Bayesian Nash equilibrium outcome to a mechanism (institution) must satisfy an incentive compatibility condition (Harris and Townsend [1981], Myerson [1979]). Implementation theory seeks to characterize those outcomes which are unique equilibrium outcomes to mechanisms, and incentive compatibility is thus a necessary condition for implementation. In this paper, we show that with a slight refinement of Bayesian Nash equilibrium, incentive compatibility is also sufficient for implementation in a large class of models. Our main result, is that in a large class of models, for any incentive compatible allocation there exists an institution to which the allocation is the unique equilibrium outcome.

Since an incentive compatible allocation can always be made an equilibrium outcome to a direct mechanism, the heart of the implementation problem is to design more complex institutions to which there are no other equilibrium outcomes. In attempting to resolve problems of multiple equilibria in games, two approaches have been followed in the literature. One approach attempts to eliminate multiple equilibria by refining the notion of equilibrium (e.g., Banks and Sobel [1987], Cho and Kreps [1987], Grossman and Perry [1986], Kohlberg and Mertens [1986], Selten [1975]). The second approach asks whether, given an equilibrium concept, the mechanism being played by the

agents can be designed so as to eliminate undesirable equilibria while retaining desirable ones (see Dasgupta, Hammond and Maskin [1979], Maskin [1986], Postlewaite [1986], Postlewaite and Schmeidler [1986], Palfrey and Srivastava [1985,1987]).

This paper continues a line of inquiry followed by Moore and Repullo [1986] and Palfrey and Srivastava [1986] which merges these two approaches and asks whether flexibility in mechanism design together with a refined equilibrium concept can resolve the multiplicity problem. Earlier applications of this approach to specific complete information settings can be found in Crawford [1979], Moulin [1979], and Reichelstein [1985]. Our result is that in a large class of settings with incomplete information, *all* multiplicity problems can be resolved with an extremely mild refinement of Bayesian Nash equilibrium: equilibria which do not involve the use of weakly dominated strategies. This result is obtained in a private values model in which agents are incompletely informed about the preferences of other agents. The restrictions we impose are that no agent is ever completely indifferent over all alternatives and that there are at least three agents. We do not require a "no veto power" condition (as in e.g. Maskin [1977] and Abreu and Sen [1986]). The proof consists of augmenting a direct mechanism and specifying outcomes so that a given incentive compatible allocation is the unique equilibrium outcome to the game.

Our possibility result stands in sharp contrast to previous results on implementation with incomplete information. Palfrey and Srivastava [1985], extending the earlier analysis of Postlewaite and Schmeidler [1986], show that a condition called Bayesian monotonicity is necessary for implementation in (unrefined) Bayesian Nash equilibrium. As shown in Palfrey and Srivastava

[1987], most "nice" allocations do not satisfy this condition even if the domain of application is restricted to the set of pure exchange economies. In Section IV of this paper, we provide the even more striking example of an allocation which is implementable in dominant strategies but not in Bayesian Nash equilibrium.

With complete information, several positive results have been obtained. Maskin [1977] showed that a condition termed monotonicity is necessary for Nash implementation and that together with a no veto power condition and at least three agents, is also sufficient (Saijo [1985]). Monotonicity is satisfied by many economically interesting sets of allocations. For example, the correspondence which associates each pure exchange neoclassical economy with the set of Pareto optimal redistributions is monotonic, as is the (constrained) Walrasian correspondence. However, most allocations (i.e. single valued correspondences) are not monotonic, and thus not Nash implementable. Moore and Repullo [1986] (see also Abreu and Sen [1986]) show that the class of implementable allocations expands significantly if the mechanism is played sequentially and subgame perfection is imposed on the equilibrium. Palfrey and Srivastava [1986] show that if there are at least 3 players and complete indifference is ruled out, then all allocations are implementable in Nash equilibrium if weakly dominated strategies are not used. This paper is then an extension of our previous results to incomplete information. What is surprising is that our previous results extend in a straightforward manner, in contrast to the failure of positive Nash implementation results to extend to Bayesian Nash implementation (Palfrey and Srivastava [1987]).

Our general possibility result with private values does not extend easily

to common value models or to models in which an agent's type only indexes the agent's information about other agents. We derive a necessary condition for unique implementation in general models and provide an example with common values, both of which highlight the difficulties arising in these situations and which indicate that positive results in this domain will be more limited.

The private values model is described in the next section, while the central possibility result is presented in Section 3. These results are compared to implementation with (unrefined) Bayesian Nash equilibrium in Section 4. General domains (e.g. common values) are considered in Section 5.

2. The Model

We employ the widely used private values model in which agents are incompletely informed about the preferences of other agents. There are I agents, and T^i denotes the set of possible types for agent i. A type for agent i, t_i, specifies the preferences of i and also i's information about other agents. A is an arbitrary set of alternatives, and $U^i(\cdot,t_i)$ the utility function of agent i if he is of type $t_i \in T^i$. Let $T = T^1 x T^2 x \ldots x T^I$. An *allocation* is a function $x : T \rightarrow A$. Let

$$X = \{ x: T \rightarrow A \}$$

be the set of all allocations.

Each agent is assumed to know his own type but not necessarily that of any other agent. The prior distribution over types is given by a distribution q on T. Given a type t_i for i, we denote by $q^i(t_{-i} | t_i)$ the posterior distribution of i over the types of the other agents. To simplify notation, we make a no moving support assumption, i.e. that $\text{support}(q^i(t_i | t_{-i})) = T^i$ for all i and t. This implies that the type of any agent is purely private information in the sense that even by pooling the information of all agents except i, i's type cannot be determined exactly.

Given an allocation $x \in X$, the expected utility to i is denoted by

$$V^i(x,t_i) = \int U^i(x(t_{-i},t_i),t_i) dq^i(t_{-i} | t_i).$$

Definition 1: A *mechanism* is a pair (M,g), $M = M^1 x\ M^2 x \ldots x M^I$ and $g : M \rightarrow A$.

M^i is the message space of i, while g is the outcome function.

Definition 2: A *strategy* for agent i is a function $\sigma^i: T^i \rightarrow M^i$.

Given a strategy σ, we denote by $g(\sigma)$ the outcome generated by σ, where the

outcome at t is $g(\sigma(t))$. The question being posed in this paper can now be formulated precisely: given an equilibrium concept and an allocation, *does there exist a mechanism which has* x *as its* *unique* *equilibrium outcome ?* Following the implementation literature, if there exists such a mechanism, we say that the allocation is *implementable*.

3. *Undominated Bayesian Equilibrium*

In this section, we study the implementation question using Bayesian equilibria which do not involve the use of weakly dominated strategies. We show that any allocation which satisfies the standard incentive compatibility condition can be made the unique equilibrium outcome to a mechanism in a large class of models.

Let $\sigma^{-i} = (\sigma^1,\ldots,\sigma^{i-1},\sigma^{i+1},\ldots,\sigma^I)$, so $\sigma = (\sigma^{-i},\sigma^i)$.

Definition 3:

(i) σ^i is a *best response* for i to σ^{-i} if for all t_i,

$$V^i(g(\sigma^{-i},\sigma^i),t_i) \geq V^i(g(\sigma^{-i},\tilde{\sigma}^i),t_i) \text{ for all } \tilde{\sigma}^i : T^i \rightarrow M^i$$

(ii) σ is a *Bayesian equilibrium* if σ^i is a best response to σ^{-i} for all i.

Definition 4: σ is *weakly dominated* if there exists i, t_i and $\tilde{\sigma}^i : T^i \rightarrow M^i$ such that $V^i(g(\tilde{\sigma}^{-i},\tilde{\sigma}^i),t_i) \geq V^i(g(\tilde{\sigma}^{-i},\sigma^i),t_i)$ for all $\tilde{\sigma}^{-i}$ with strict inequality for some $\tilde{\sigma}^{-i}$.

This says that no matter what strategies are used by the others, agent i does at least as well by using $\tilde{\sigma}^i$ instead of σ^i, and for some strategy combination of the others, he does strictly better.

Definition 5: σ is an *undominated Bayesian equilibrium* if σ is a Bayesian equilibrium and σ is not weakly dominated.

It is clear that any allocation which can be made the unique equilibrium outcome to a mechanism must satisfy an incentive compatibility condition. This is immediate from the literature on Bayesian incentive compatibility (e.g. Myerson [1979], Harris and Townsend [1981]).

Definition 6: x: T → A is *incentive compatible* if for all i, for all t_i,

$$\int U^i(x(t_{-i},t_i),t_i)dq^i(t_{-i}|t_i) \geq \int U^i(x(t_{-i},t_i'),t_i)dq^i(t_{-i}|t_i) \text{ for all } t_i' \in T^i.$$

We state the following well known result for completeness.

Theorem 1: If x is implementable then x is incentive compatible.

The next definition summarizes a restriction on the environment. It says that there are no redundant preference-types for any agent in the sense that types index preferences.

Definition 7 (Value-Distinguished types): For all i, t_i, t_i', $t_i \neq t_i'$, either there exist $y^i,z^i \in A$ with $U^i(y^i,t_i) \geq U^i(z^i,t_i)$ and $U^i(y^i,t_i') < U^i(z^i,t_i')$ or there exist $y^i,z^i \in A$ with $U^i(y^i,t_i) > U^i(z^i,t_i)$ and $U^i(y^i,t_i') \leq U^i(z^i,t_i')$.

In some applications, value distinction may require us to consider random allocations. This will be the case if, for example, the difference between types is difference in risk aversion. In this case, types are value distinguished on the set of lotteries over A.

Our sufficiency result requires us to also impose the following mild restrictions on the domain of possible types.

Definition 8:

(i) (No Complete Indifference) For all i, t_i, there exists a,b ∈ A with $U^i(a,t_i) > U^i(b,t_i)$.

(ii) (Existence of best and worst elements) For all i, t_i, there exist $b(t_i)$, $w(t_i) \in A$ with $U^i(b(t_i),t_i) \geq U^i(a,t_i)\ \forall\ a \in A$ and $U^i(a,t_i) \geq U^i(w(t_i),t_i)\ \forall\ a \in A$.

Theorem 2: Assume $I \geq 3$, no complete indifference, the existence of best and worst elements, and that types are value distinguished. If x is incentive compatible, then x can be made the unique undominated Bayesian equilibrium outcome to a mechanism.

Proof: We construct a mechanism for the case of strictly value distinguished types, i.e. $t_i \neq t_i'$ implies there exist $y_1(t_i,t_i') \in A$ and $y_2(t_i,t_i') \in A$ with $U^i(y_1(t_i,t_i'),t_i) > U^i(y_2(t_i,t_i'),t_i)$ and $U^i(y_1(t_i,t_i'),t_i') < U^i(y_2(t_i,t_i'),t_i')$. Straightforward methods for extending the mechanism to account for indifference in value distinction are contained in Palfrey and Srivastava [1986]. Let

$M^i = M_1^i \times M_2^i \times M_3^i \times M_4^i$ where

$M_1^i = T^i$

$M_2^i = \bigcup_j T^j$

$M_3^i = [0,I+2)$

$M_4^i = A.$

In the appendix, we prove that the outcome function can be defined so that the only undominated Bayesian equilibrium is of the type $\sigma^i(t_i) = (t_i,t_i,0,b_i)$ for all i and t, where b_i is a best element for i at t_i. Here, we give the intuition behind the construction of the mechanism and how it works. The allocation rule, g, is, effectively, a direct mechanism almost everywhere in M. By this we mean that, except for specific isolated portions of M, g only uses the information contained in the first component of everyone's message. Calling this region, R_o, we have $g(m) = x(m_1)\ \forall\ m \in R_o$. The remainder of M is divided into regions indexed by i. In these regions, $m_2^j = \hat{t}_i \in T^i$ for all $j \neq i$. The allocation rule for such a region, denoted R_i, is given in the following table.

	Message of Agent i		
Message of all $j \neq i$	$(t_i,t_i,0,a_i)$	(t_i,t_i',k_i,a_i)	(t_i,t_i,k_i,a_i), $k_i>0$ or (t_i,t_j,k_i,a_i), $j \neq i$
(t_j,t_i,k_j,a_j) $k_j \in [I+1,I+2)$	a_i	a_i	$w(t_i)$
(t_j,t_i',k_j,a_j) $k_j \in [I+1,I+2)$	$y_1(t_i,t_i')$	$y_2(t_i,t_i')$	$a_*(k)$
(t_j,t_i'',k_j,a_j) $k_j \in [I+1,I+2)$ $t_i'' \notin (t_i,t_i')$	$y_1(t_i,t_i'')$	a_i	$a_*(k)$
(t_j,t_i,k_j,a_j) $k_j \in [i,i+1)$	a_i	a_i	$w(t_i)$
(t_j,t_i'',k_j,a_j) $k_j \in [i,i+1)$ $t_i'' \neq t_i$	$w(t_i'')$	$w(t_i'')$	$w(t_i'')$
All other messages with $m_2^j = \hat{t}_i \in T^i$	$a_*(k)$	$a_*(k)$	$a_*(k)$

Table 1

Allocation rule restricted to region R_i

We have denoted by $a_*(k)$ the outcome requested in m_4 by the smallest j such that $k_j \geq k_\ell$ for all ℓ. The strategies for agent i are given by the columns of the matrix, while the strategies of all agents other than i are given by the rows.

The initial step in the proof is to show that there is no equilibrium in R_i. Suppose agent i is of type t_i. Note first that reporting $m^i_4 = a$ with $U^i(a,t_i) < U^i(b(t_i),t_i)$ is weakly dominated; changing a to $b(t_i)$ is strictly better for i at several m_{-i}, and, if the rest of m^i is unaltered, i is never worse off. Without loss of generality, then, suppose $m^i_4 = b(t_i)$.

Next, we note that there is no equilibrium with $k_j > 0$ for some j. To see this, suppose $J \leq k_j < J+1$ for some nonegative integer $J \leq I+1$ and $k_j \neq 0$. Then, $(k_i + J+1)/2$ weakly dominates, since j is strictly better off somewhere in the bottom row of the table and no worse off anywhere.

A similar argument applies for i if $k_i \neq 0$ or if $m^i_2 \neq m^i_1$. We conclude that all equilibria must lie in R_o, with $m^j_3 = 0$ and $m^j_1 = m^j_2$ for all j.

The next step is to observe that at t_i', playing $(t_i, t_i, 0, b(t_i'))$ with $t_i' \neq t_i$ is weakly dominated by $(t_i, t_i', k_i, b(t_i'))$. This change only alters the outcome in rows 2,3, and 6. In row 2, the outcome changes from $y_1(t_i, t_i')$ to $y_2(t_i, t_i')$. By construction, $U^i(y_2(t_i,t_i'),t_i') > U^i(y_1(t_i,t_i'),t_i')$, so i is better off. In rows 3 and 6, i is never worse off.

Hence the only possible equilibrium is $\sigma^i(t_i) = (t_i, t_i, 0, b(t_i))$ for all i, and t_i. To see that this is indeed an undominated Bayesian equilibrium, we first note that incentive compatibility implies that when all $j \neq i$ play σ^j, σ^i is a best response for i, since a unilateral deviation by i can only change the outcome from $x(t)$ to $x(t_{-i}, t_i')$ at t. To se that it is not weakly dominated, we have to consider each possible deviation by i. These cases are covered in detail in the Appendix, and are easily checked by inspection of the table.

To conclude, the only equilibria are $\sigma^i(t_i) = (t_i, t_i, 0, b_i)$ for all i and t_i where b_i is a best element at t_i, and all these equilibria yield x as the outcome. Hence, this mechanism implements x. If some individual has more than one best element then there are multiple equilibria, but all equilibria produce x as the outcome. Furthermore, the equilibrium strategies are "interchangeable", since they only differ in the last component of the message space.

4. *Unrefined Bayesian Equilibrium*

In this section, we compare the results of Section 3 with implementation using Bayesian equilibrium as the solution concept. It is shown in Palfrey and Srivastava [1985] that a necessary condition for unique implementation of an allocation can be written in terms of "deceptive" strategies.

A *deception* by agent i is a function $\alpha^i : T^i \to T^i$. The interpretation is that when i is of type t_i, i acts as if he is of type $\alpha^i(t_i)$. In a direct mechanism, the set of all possible α^i is the set of strategies available to i, while truth telling by i is the identity α^i. With this notation, incentive compatibility can be rewritten as : for all i, for all t_i,

$$\int U^i(x(t_{-i},t_i),t_i)dq^i(t_{-i}|t_i) \geq \int U^i(x(t_{-i},\alpha^i(t_i)),t_i)dq^i(t_{-i}|t_i)$$

for all $\alpha^i : T^i \to T^i$. This is the standard incentive compatibility condition and says that if in a direct mechanism, all other agents are using truthful strategies, then the truthful strategy does at least as well for agent i as any deception. Notice that incentive compatibility does <u>not</u> say what is a best response when other agents are playing deceitfully. Implementation essentially requires us to study which nontruthful α's can arise as equilibria. In contrast, Theorem 1 told us that, at least for a certain class of models, one can design mechanisms such that these alternative candidates for equilibrium rely on the use of dominated strategies. The following condition characterizes those α's which will be sustained as Bayesian equilibria.

Let $\alpha = (\alpha^1,\ldots,\alpha^I)$, $\alpha^{-i} = (\alpha^1,\alpha^2,\ldots,\alpha^{i-1},\alpha^{i+1},\ldots,\alpha^I)$ so that $\alpha = (\alpha^{-i},\alpha^i)$. For any α, denote $x_\alpha(t) = x(\alpha(t))$.

<u>Definition 9</u>: $x : T \to A$ satisfies <u>*Bayesian monotonicity*</u> if for any α such that $x_\alpha(t) \neq x(t)$ for some t, there exist i, t_i, $y : T \to A$ such that

$\int U^i(x(t_{-i},t_i'),t_i')dq^i(t_{-i}|t_i') \geq \int U^i(y(t_{-i},\alpha^i(t_i)),t_i')dq^i(t_{-i}|t_i')$ <u>for all</u> t_i'

and $\int (x_\alpha(t_{-i},t_i),t_i)dq^i(t_{-i}|t_i) < \int U^i(y_\alpha(t_{-i},t_i),t_i)dq^i(t_{-i}|t_i)$.

It follows almost immediately that if x can be made the unique Bayesian equilibrium outcome to a mechanism, then it must satisfy Bayesian monotonicity. To see this suppose (M,g) implements x, and α is such that $x_\alpha(t) \neq x(t)$ for some t. If σ is a Bayesian equilibrium with $g(\sigma) = x$, then we must have $V^i(g(\sigma^{-i},\sigma^i),t_i') \geq V^i(g(\sigma^{-i},\tilde{\sigma}^i),t_i')$ for all t_i' and for all $\tilde{\sigma}^i$. But σ_α, defined by $\sigma_\alpha^i(t_i) = \sigma^i(\alpha^i(t_i))$ for all i and t_i, cannot be an equilibrium, since $g(\sigma_\alpha) = x_\alpha$ and x_α is not an equilibrium outcome. This implies there must exist some agent, say i, a message, m^i, and some type for this agent, say t_i, such that

$$V^i(g(\sigma_\alpha^{-i},m^i),t_i) > V^i(g(\sigma_\alpha),t_i).$$

Let $y(t) = g(\sigma^{-i}(t), m^i)$ for all t. Then, $g(\sigma_\alpha^{-i},m^i) = y_\alpha$ while $g(\sigma_\alpha) = x_\alpha$ so we get $\int (x_\alpha(t_{-i},t_i),t_i)dq^i(t_{-i}|t_i) < \int U^i(y_\alpha(t_{-i},t_i),t_i)dq^i(t_{-i}|t_i)$.
However, since σ^i is a best response to σ^{-i}, we must have
$\int U^i(x(t_{-i},t_i'),t_i')dq^i(t_{-i}|t_i') \geq \int U^i(y(t_{-i},\alpha^i(t_i)),t_i')dq^i(t_{-i}|t_i')$ <u>for all</u> t_i'.
We conclude that Bayesian monotonicity is necessary for implementation.■

The following example shows that most "nice" allocations do not satisfy Bayesian monotonicity.

<u>Example 1</u>: $I=3$, $A=\{a,b\}$, $T^i = \{t_a,t_b\}$ for all i. Types are independently drawn with $[q^i(t_b)]^2 > .5$ for all i. Preferences are as follows: type t_a strictly prefers a to b, while type t_b strictly preferes b to a. Normalize utility so that $U^i(a,t_a) = 1 > 0 = U^i(b,t_a)$ and $U^i(b,t_b) = 1 > 0 = U^i(a,t_b)$.
Consider the following allocation, x :

2 is

		t_a	t_b
1 is	t_a	a	a
	t_b	a	b

3 is t_a

2 is

		t_a	t_b
1 is	t_a	a	b
	t_b	b	b

3 is t_b

This allocation has many nice properties and indeed, is the only reasonable allocation in that:

(i) It is incentive compatible.

(ii) It is ex-ante efficient, interim efficient, durable, and ex-post efficient in the sense of Holmstrom and Myerson [1981].

(iii) $x(t)$ is the (unique) majority winner at t.

(iv) It maximizes an Arrow social welfare function.

(v) It can be implemented in dominant strategies (and, of course, in undominated Bayesian equilibrium).

Remarkably, x is not implementable in Bayesian equilibrium: let $\alpha^i(t_i) = t_b$ for all i, so $x(\alpha(t)) = b$ for all t. We show below that there do not exist i, y, and t_i which satisfy the inequalities required by Bayesian monotonicity. Consequently, in any game in which σ is a Bayesian equilibrium with $g(\sigma) = x$, σ_α is also a Bayesian equilibrium with $g(\sigma_\alpha) = x_\alpha$. This has severe welfare implications, as $x_\alpha = b$ violates (ii), (iii), and (iv).

To show that Bayesian monotonicity is not satisfied requires us to prove that there does not exist $y: T \to A$ which satisfies the first set of inequalities in Definition 6, with y_α simultaneously satisfying the second inequality. To see this, note first that since α is a "projection" to t_b, y_α is a constant allocation. Furthermore, if $y_\alpha(t) = b$ for all t, then $x_\alpha = y_\alpha$, in which case the second inequality could not be satisfied, so we can limit attention to y's such that $y_\alpha(t) = a \; \forall \; t$. Since a is the worst element for type t_b, the inequality:

$$\int U^i(x_\alpha(t_{-i},t_i),t_i)dq^i(t_{-i}|t_i) < \int U^i(y_\alpha(t_{-i},t_i),t_i)dq^i(t_{-i}|t_i)$$

implies $t_i = t_a$. Further, since $\alpha(t) = (t_b,t_b,t_b)$, we must have $y(t_b,t_b,t_b) = a$. By our choice of $y(t_b,t_b,t_b)$, the second inequality of Definition 9 is satisfied for all i when i is type t_a. We need to show that the other elements of y cannot be picked to satisfy the first inequality of Definition 9. Since the problem is symmetric, we need consider only agent 1. The expected utility from x at t_a is $1-q(t_b)^2$ while that from $y(t_{-1},\alpha^1(t_a))$ at t_a is:

$$q(t_a)^2U^1(y(t_b,t_a,t_a),t_a)+2q(t_b)q(t_a)U^1(y(t_b,t_b,t_a),t_a)+q(t_b)^2U^1(y(t_b,t_b,t_b),t_a)$$

Since $y(t_b,t_b,t_b) = a$, this reduces to

$$q(t_b)^2 + q(t_a)^2 U^1(y(t_b,t_a,t_a),t_a) + 2q(t_b)q(t_a) U^1(y(t_b,t_b,t_a),t_a)$$

The minimum value of this last expression over y is $q(t_b)^2$, which is greater than $1-q(t_b)^2$, so the first inequality of Bayesian monotonicity must be violated when agent 1 is of type t_a. Hence, x is not implementable.■

5. *Extensions*

In this section, we discuss the assumptions underlying Theorem 1, and consider generalizations of our model.

A. Common Values

The most significant assumption in Theorem 1 is private values. Even though a large majority of applications to date of Bayesian games to economic problems and virtually all applications of the revelation principle to mechanism design have used this assumption, it is clearly a restrictive assumption. Curiously, our general possibility result does not apply with nearly the same force in settings with common values, as we now discuss.

The model itself is easily modified to incorporate common values. To do this, we write the utility function of agent i at t as $U^i(\cdot,t)$ instead of $U^i(\cdot,t_i)$, but still assume that at t, i observes only t_i, and that there is no moving support. Incentive compatibility is now stated as follows.

__Definition 6′__: x is __incentive compatible__ if for all i, for all t_i,
$\int U^i(x(t_{-i},t_i),t)\,dq^i(t_{-i}|t_i) \geq \int U^i(x(t_{-i},t_i'),t)\,dq^i(t_{-i}|t_i)$ for all $t_i' \in T^i$.

Let $V^i(y,t_i) = \int U^i(y(t_{-i},t_i),t)\,dq^i(t_{-i}|t_i)$. The following theorem yields a necessary condition for implementing an allocation.

__Theorem 3__: If x is implementable in undominated Bayesian equilibrium, then x is incentive compatible, and for any $\alpha:T \to T$, $x_\alpha(t) \neq x(t)$ for some t implies that at least one of the following conditions hold:

(a) there exist i, t_i, and $y \in X$ with

$$\int U^i(x(t_{-i},t_i'),t_{-i},t_i')\,dq^i(t_{-i}|t_i') \geq \int U^i(y(t_{-i},\alpha^i(t_i)),t_{-i},t_i')\,dq^i(t_{-i}|t_i')$$

__for all__ $t_i' \in T_i$ and

$$\int U^i(x_\alpha(t_{-i},t_i),t_{-i},t_i)\,dq^i(t_{-i}|t_i) < \int U^i(y_\alpha(t_{-i},t_i),t_{-i},t_i)\,dq^i(t_{-i}|t_i).$$

(b) there exist i, t_i, and $y_1, y_2, z_1, z_2 \in X$ with

$V^i(y_1,\alpha^i(t_i)) > V^i(y_2,\alpha^i(t_i))$
$V^i(y_{1\beta},t_i) \leq V^i(y_{2\beta},t_i)$ for all deceptions β with $\beta^i = \alpha^i$
$V^i(z_1,t_i) > V^i(z_2,t_i)$ and
$V^i(z_{1\beta},t_i) \geq V^i(z_{2\beta},t_i)$ for all deceptions β with $\beta^i = \alpha^i$

(c) there exist i, t_i, and $y_1,y_2 \in X$ with

$V^i(y_1,\alpha^i(t_i)) = V^i(y_2,\alpha^i(t_i))$
$V^i(y_1,t_i) < V^i(y_2,t_i)$ and
$V^i(y_{1\beta},t_i) \leq V^i(y_{2\beta},t_i)$ for all deceptions β with $\beta^i = \alpha^i$.

__Proof__ : See Appendix.

With private values, parts (b) and (c) of this result reduce to the statement that types are value distinguished. For example, consider (b). In this case, we must have $U^i(y_1(t_{-i},t_i'),t_i') > U^i(y_2(t_{-i},t_i'),\ t_i')$ for some t_{-i} where $t_i' = \alpha^i(t_i)$. Now, consider $\beta^{-i}(t_{-i}') = t_{-i}$ for all t_{-i}', yielding $U^i(y_1(t_{-i},t_i'),t_i') \leq U^i(y_2(t_{-i},t_i')t_i')$, which says that t_i and t_i' are value distinguished. The assumption of no complete indifference yields the existence of z_1 and z_2 satisfying the requirements of the condition.

Except in private values models, conditions (b) and (c) appear to be very strong, in fact sufficiently strong that they seem unlikely to be satisfied in general applications. This suggests that undominated Bayesian implementation is essentially equivalent to (unrefined) Bayesian implementation once one moves beyond private value domains with value-distinguished types.

The following example, which is a variant of our earlier example, illustrates the difficulties arising with common values.

__Example 2__: $A = \{a,b\}$, $I = 3$, $T^i = \{t_a,t_b\}\ \forall i$, and types are independently drawn with $(q^i(t_b))^2 > .5$ for all i. The agents have "majoritarian" preferences, given by

$$U^i(a,t) = \begin{cases} 1 & \text{if at least 2 agents are type } t_a \\ 0 & \text{otherwise} \end{cases}$$

$$U^i(b,t) = \begin{cases} 0 & \text{if at least 2 agents are type } t_b \\ 1 & \text{otherwise.} \end{cases}$$

With this structure of preferences, all agents are ex-post identical. The following incentive compatible allocation, x, is (uniquely) efficient in all senses and for each t, picks out the unanimous socially preferred outcome:

		2 is	
		t_a	t_b
1 is	t_a	a	a
	t_b	a	b

3 is t_a

		2 is	
		t_a	t_b
1 is	t_a	a	b
	t_b	b	b

3 is t_b

Surprisingly, this allocation is not implementable in undominated Bayesian equilibrium. To see this, consider $\alpha^i(t_i) = t_b$ for all i, so $x_\alpha(t) = b$ for all t, as in Example 1. We claim that for any mechanism, if x is an undominated Bayesian equilibrium outcome, then x_α is also an undominated Bayesian equilibrium outcome. A proof is given in the Appendix.

B. Value distinguished types

We turn next to the assumption of value distinguished types. This assumption rules out the case where an agent's type determines only his information about other agents. This case, of purely "information distinguished types" is not covered by Theorem 2. In fact, with private values, Theorem 3 reduces to saying that if x is implementable, then either x must satisfy Bayesian monotonicity or types must be value distinguished. We are unaware of specific economic applications in the literature in which types are not value distinguished, though such problems have been studied in game theory under the heading of "games with incomplete information on one-and-one-half sides."

6. Summary

The main result of this paper is that the multiple equilibrium problem in mechanism design can be solved in private value models, if ~~as~~ players do not to use weakly dominated strategies. The result applies less generally to common value situations and to situations in which types are not value distinguished.

Appendix

Proof of Theorem 2: We divide the message space as follows.

$D_1 = \{m \mid m^j = (t_j, t_j', k_j, a_j)\ \forall_j\}$

$D_2 = \{m \notin D_1 \mid$ there do not exist i, $t_i'' \in T_i$ with $m_2^j = t_i''\ \forall\ j \neq i\}$

$$\begin{aligned} D^i_{3A} = \{m \mid\ & m^j = (t_j, t_i', k_j, a_j)\ \forall\ j \neq i,\ k_j \in [I+1, I+2)\ \forall\ j \neq i\} \\ & m^i = (t_i, t_i, 0, a_i),\ t_i \neq t_i'\} \end{aligned}$$

$$\begin{aligned} D^i_{3B} = \{m \mid\ & m^j = (t_j, t_i, k_j, a_j)\ \forall\ j \neq i,\ k_j \in [I+1, I+2)\ \forall\ j \neq i,\} \\ & m^i = (t_i, t_i, 0, a_i)\} \end{aligned}$$

$$\begin{aligned} D^i_{4A} = \{m \mid\ & m^j = (t_j, t_i', k_j, a_j)\ \forall\ j \neq i,\ k_j \in [I+1, I+2)\ \forall\ j \neq i, \\ & m^i = (t_i, t_i', k_i, a_i),\ t_i \neq t_i\} \end{aligned}$$

$$\begin{aligned} D^i_{4B} = \{m \mid\ & m^j = (t_j, t_i'', k_j, a_j)\ \forall\ j \neq i,\ k_j \in [I+1, I+2)\ \forall\ j \neq i, \\ & m^i = (t_i, t_i', k_i, a_i),\ t_i' \neq t_i'',\ t_i' \neq t_i\} \end{aligned}$$

$$\begin{aligned} D^i_{5A} = \{m \mid\ & m^j = (t_j, t_i, k_j, a_j)\ \forall\ j \neq i,\ k_j \in [i, i+1)\ \forall\ j \neq i, \\ & m^i = (t_i, t_i', k_i, a_i),\ t_i' \neq t_i, \text{ or } m^i = (t_i, t_i, 0, a_i))\} \end{aligned}$$

$$\begin{aligned} D^i_{5B} = \{m \mid\ & m^j = (t_j, t_i, k_j, a_j)\ \forall\ j \neq i,\ k_j \in [i, i+1)\ \forall\ j \neq i, \\ & m^i \neq (t_i, t_i', k_i, a_i),\ t_i' \neq t_i, \text{ and } m^i \neq (t_i, t_i, 0, a_i))\} \end{aligned}$$

$D_6 = \{$all other $m\}$.

For $m \in D_6$, let $i*$ be the smallest i such that $m_3^i \geq m_3^j\ \forall\ j$, and let $a_{i*} = m_4^{i*}$.

The outcome function is given by

$$g(m) = \begin{cases} x(t) & \text{if } m \in D_1 \\ x(t) & \text{if } m \in D_2 \\ y_1(t_i, t_i') & \text{if } m \in D^i_{3A} \\ a_i & \text{if } m \in D^i_{3B} \\ y_2(t_i, t_i') & \text{if } m \in D^i_{4A} \\ a_i & \text{if } m \in D^i_{4B} \\ a_i & \text{if } m \in D^i_{5A} \\ w(t_i) & \text{if } m \in D^i_{5B} \\ a_i * & \text{if } m \in D_6 \end{cases}$$

We start by showing that $\sigma^i(t_i) = (t_i, t_i, 0, b(t_i))$ is a Bayesian equilibrium. This can be seen by noting that a unilateral deviation by i from from this strategy only affects the outcome if i changes m^i_1. (Note: this would not be true if $I=2$, since in that case $D_2 \cap (D^i_3 \cup D^i_4) \neq \emptyset$.) If, at t_i, i instead reports $m^i_1 = t_i'$, the outcome at t is $x(t_{-i}, t_i')$ instead of $x(t_{-i}, t_i)$. Incentive compatibility now directly implies that σ is a Bayesian equilibrium.

Next, we argue that σ is not weakly dominated. To see this, note first that not reporting a best element in the fourth component of the message is always weakly dominated, since the report in this component is always used in an agents favor. Without loss of generality, therefore, we assume that $m^i_4 = b(t_i)$ for all i, t_i.

Next, we consider four possible types of deviations by i at t_i and show that none of these deviations weakly dominates $(t_i, t_i, 0, b(t_i))$.

(i) $m^i_1 \neq t_i$.

In this case, i is strictly worse off when $m^j = (t_j, t_i, i, a_i)$ $\forall j \neq i$ since the outcome moves from $b(t_i)$ to $w(t_i)$.

(ii) $m^i = (t_i, t_i, k_i, b(t_i))$, $k_i > 0$. Again, i is strictly worse off when $m^j = (t_j, t_i, i, a_j)$ $\forall j \neq i$.

(iii) $m^i = (t_i, t_j, k_i, b(t_i))$, $j \neq i$.

In this case, i is again strictly worse off when $m^j = (t_j, t_i, i, a_j)$ $\forall j \neq i$, since the outcome changes from $b(t_i)$ to $w(t_i)$

(iv) $m^i = (t_i, t_i', k_i, b(t_i))$, $t_i' \neq t_i$. Here, i is strictly worse off when $m^j = (t_j, t_i', I+1, a_j)$ since the outcome changes from $y_1(t_i, t_i')$ to $y_2(t_i, t_i')$.

We conclude that σ is an undominated Bayesian equilibrium, yielding x as the outcome.

We now argue that there are no other equilibria, thereby concluding that x is the unique equilibrium outcome. This is argued in two steps : first, that all undominated equilibria are of the form $\sigma^i(t_i) = (t_i', t_i', 0, b(t_i))$, and second, that $t_i' \neq t_i$ is weakly dominated.

First, note that there is no equilibrium at t with $m^i_3 > 0$ for some i. To see this, let J be an integer such that $J \leq m^i_3 < J+1$. Then, reporting $\hat{m}^i = m^i$ except $\hat{m}^i_3 = (m^i_3 + J + 1)/2$ weakly dominates reporting m^i since there is a configuration of messages in D_6 such that $g(m^{-i}, m^i) = w(t_i)$ but $g(m^{-i}, \hat{m}^i) = m^i_4 = b(t_i)$, and no configuration of messages such that $U^i(g(m), t_i) > U^i(g(m^{-i}, \hat{m}^i), t_i)$. Second, $\sigma^i(t_i) = (t_i, t_j, 0, b(t_i))$ and $\sigma^i(t_i) = (t_i', t_i'', 0, b(t_i))$ for $t_i' \neq t_i''$ are both weakly dominated by the same argument. This only leaves $\sigma^i(t_i) = (t_i', t_i', 0, b_i)$ where b_i is a best element at t_i. We claim $m^i = (t_i', t_i, k_i, b_i)$ weakly dominates this strategy. The outcome only changes in D_3 and D_6. In D_3, the outcome changes from $y_1(t_i', t_i)$ to $y_2(t_i', t_i)$, so i is strictly better off since $U^i(y_2(t_i', t_i), t_i) > U^i(y_1(t_i', t_i), t_i)$. It is clear that i is no worse off in D_6.

This concludes the proof of Theorem 2.■

Proof of Theorem 3: The revelation principle implies x is incentive compatible. Let (M,g) implement x, let σ be an equilibrium with $g(\sigma) = x$, and let $x_\alpha \neq x$ for some α. Then, σ_α, yielding x_α as the outcome, is not an undominated Bayesian equilibrium. Two cases arise : either σ_α is not a Bayesian equilibrium or it is one. In the first case the argument showing Bayesian monotonicity is necessary for Bayesian implementation yields (a).

Suppose, then, that σ_α is a Bayesian equilibrium. Then, it must be weakly dominated, so there exist i, t_i, and $\bar{\sigma}^i$ such that

$$\int U^i(g(\bar{\sigma}^{-i},\bar{\sigma}^i),t),dq^i(t_{-i}|t_i) \geq \int U^i(g(\bar{\sigma}^{-i},\sigma^i_\alpha(t_i)),t)\ dq^i(t_{-i}|t_i)$$

for all $\bar{\sigma}^{-i}$ with strict inequality holding for some $\bar{\sigma}^{-i}$. Note that $\alpha^i(t_i) \neq t_i$, since otherwise, $\sigma^i_\alpha(t_i) = \sigma^i(t_i)$, which would imply that σ is weakly dominated, a contradiction.

Since σ is not weakly dominated at $\alpha^i(t_i)$, we get either

(i) $\int[U^i(g(\bar{\sigma}^{-i},\sigma^i),\alpha^i(t_i)) - U^i(g(\bar{\sigma}^{-i},\bar{\sigma}^i),\alpha^i(t_i))]dq^i(t_{-i}|t_i) > 0$ for some $\bar{\sigma}^{-i}$

or

(ii) $\int[U^i(g(\bar{\sigma}^{-i},\sigma^i),\alpha^i(t_i)) - U^i(g(\bar{\sigma}^{-i},\bar{\sigma}^i),\alpha^i(t_i))]dq^i(t_{-i}|t_i) = 0$

for all $\bar{\sigma}^{-i}$.

Consider case (i), and let $\bar{\sigma}^i(t_i') = \bar{\sigma}^i(\alpha^i(t_i))$ for all t_i'. Then, letting $y_1 = g(\bar{\sigma}^{-i},\sigma^i(\alpha^i(t_i)))$, $y_2 = g(\bar{\sigma}^{-i},\bar{\sigma}^i)$, we get $V^i(y_1,\alpha^i(t_i)) > V^i(y_2,\alpha^i(t_i))$. Since $\bar{\sigma}^i$ weakly dominates σ^i_α at t_i, we must have $V^i(y_1,t_i) \leq V^i(y_2,t_i)$. Further, for any deception β with $\beta^i = \alpha^i$, we must have $V^i(y_{1\beta},t_i) \leq V^i(y_{2\beta},t_i)$, since $y_{1\beta}$ is the outcome when all agents except i play $\bar{\sigma}_{\beta^{-i}}$ and i plays σ^i_α while $y_{2\beta}$ is the outcome when all $j \neq i$ play $\bar{\sigma}_{\beta^{-i}}$ and i plays $\bar{\sigma}^i$. To complete case (i), it must also be the case that $V^i(g(\bar{\sigma}^{-i},\bar{\sigma}^{-i}),t_i) > V^i(g(\bar{\sigma}^{-i},\sigma^i_\alpha),t_i)$ for some $\bar{\sigma}^{-i}$. Let $z_1 = g(\bar{\sigma}^{-i},\bar{\sigma}^i)$, $z_2 = g(\bar{\sigma}^{-i},\sigma^i_\alpha(t_i))$. Then,

$V^i(z_1,t_i) > V^i(z_2,t_i)$. Repeating the argument above, we get $V^i(z_{1\beta},t_i) \geq V^i(z_{2\beta},t_i)$ for all β with $\beta^i = \alpha^i$.

We have thus shown that there exist i, t_i, y_1, y_2, z_1, z_2 such that

$V^i(y_1,\alpha^i(t_i)) > V^i(y_2,\alpha^i(t_i))$,

$V^i(y_{1\beta},t_i) \leq V^i(y_{2\beta},t_i)$ for any deception β with $\beta^i = \alpha^i$.

$V^i(z_1,t_i) > V^i(z_2,t_i)$

$V^i(z_{1\beta},t_i) \geq V^i(z_{2\beta},t_i)$. for any deception β with $\beta^i = \alpha^i$

This is precisely the requirement in (b).

Consider next case (ii). Here,

$V^i(g(\bar{\sigma}^{-i},\sigma^i),\alpha^i(t_i)) = V^i(g(\bar{\sigma}^{-i},\bar{\sigma}^i),\ \alpha^i(t_i))$

for all $\bar{\sigma}^{-i}$. Since $\bar{\sigma}^i$ weakly dominates σ^i_α, we must have

$V^i(g(\bar{\sigma}^{-i},\bar{\sigma}^i),t_i) > V^i(g(\bar{\sigma}^{-i},\sigma^i_\alpha),t_i)$ for some $\bar{\sigma}^{-i}$, and

$V^i(g(\bar{\sigma}_{\beta^{-i}},\bar{\sigma}^i),t_i) \geq V^i(g(\bar{\sigma}_{\beta^{-i}},\sigma^i_\alpha),t_i)$ for all β with $\beta^i = \alpha^i$.

Letting $y_2 = g(\bar{\sigma}^{-i},\bar{\sigma}^i)$, $y_1 = g(\bar{\sigma}^{-i},\sigma^i_\alpha)$, we get

$V^i(y_1,\alpha^i(t_i)) = V^i(y_2,\alpha^i(t_i))$

$V^i(y_1,t_i) < V^i(y_2,t_i)$ and

$V^i(y_{1\beta},t_i) \leq V^i(y_{2\beta},t_i)$ for any deception β with $\beta^i = \alpha^i$.

This is the requirement in (c), and concludes the proof.■

Proof of Claim in Example 2: We prove the claim in two parts.

Part I: If σ is a Bayesian equilibrium to (M,g) with $g(\sigma) = x$, the σ_α is a Bayesian equilibrium.

Part II: If σ is an undominated Bayesian equilibrium to (M,g), with $g(\sigma) = x$, then σ_α is undominated.

Proof of Part I: Suppose everyone except i uses the strategy $\hat{\sigma}^{-i} = \sigma^{-i}(t_b)$ regardless of type. Then, the outcome depends (at most) only upon i's type. Regardless of i's type, i prefers b to a since the others are likely to both be t_b's. Hence $\hat{\sigma}^i = \sigma^i(t_b)$ regardless of i's type is a best response to $\hat{\sigma}^{-i}$.■

Proof of Part II: This is more involved, and requires meticulous checking that none of the conditions (a), (b), or (c) of Theorem 2 are satisfied. Since part (I) of the proof already implies that (a) is not satisfied, we need only show that (b) and (c) are not satisfied. In this simple example, (c) is obviously never satisfied since $V^i(y_1,t_b) = V^i(y_2,t_b)$ if and only if $y_1 = y_2$. Therefore the proof reduces to showing that (b) is not satisfied. That is:

There does not exist a pair of allocations, y_1,y_2 such that

(*) $V^i(y_1,t_b) > V^i(y_2,t_b)$ but $V^i(y_{1\beta},t_a) \le V^i(y_{2\beta},t_a)$ $\forall\ \beta$ with $\beta^i = \alpha^i$.

This can be proven in a series of steps.

First, without loss of generality, fix i=3.

Step 1: It suffices to show that there do not exist allocations $y: T^{-3} \to A$ (i.e. allocations which are constant in player 3's type) such that $V^3(y_1,t_b) > V^3(y_2,t_b)$ but $V^3(y_{1\beta^{-3}},t_a) \le V^3(y_{2\beta^{-3}},t_a)$ $\forall\ \beta^{-3}$.

Proof: This follows immediately from the fact that both inequalities of (*) hold the argument of y_1 and y_2 corresponding to player 3's type fixed at t_b.■

Thus we are reduced to a search of all pairs y_1 and y_2 which can be represented by 2x2 outcome matrices, as below.

Player 2's type

		t_a	t_b
Player 1's type	t_a		
	t_b		

y_1

Player 2's type

		t_a	t_b
Player 1's type	t_a		
	t_b		

y_2

The remainder of the proof involves a demonstration that there is no way to fill in the above matrices with a's and b's in such a way that (*) is satisfied.

Step 2: If some entry in y_2 is a, the corresponding entry in y_1 is a (i.e. $y_2(t_{-i}) = a \Rightarrow y_1(t_{-i}) = a\ \forall\ t_{-i}$).

Proof: Suppose $y_2(\hat{t}_{-i}) = a$ but $y_1(\hat{t}_{-i}) = b$ for some $\hat{t}_{-i}$. Then the second inequality of (*) is violated for the β^{-i} which projects to $\hat{t}_{-i}$ (i.e. $\beta^{-i}(t_{-i}) = \hat{t}_{-i}\ \forall\ t_{-i}$).■

Step 3: $y_2(t_{-i}) = b$ and $y_1(t_{-i}) = a$ for some t_{-i}.

Proof: If not, then $y_1 = y_2$ so the first inequality of (*) violated.■

Step 4: $y_2(t_{-i}) = b\ \forall\ t_{-i}$.

Proof:

Suppose $y_2(t_a,t_a) = a$. Then by step 2, $y_1(t_a,t_a) = a$, and by step 3 there exists $\hat{t}_{-i} \neq (t_a,t_a)$ such that $y_2(\hat{t}_{-i}) = b$ and $y_1(\hat{t}_{-i}) = a$. There are three possibilities:

(I)

	t_a	t_b
t_a	a	·
t_b	a	·

y_1

	t_a	t_b
t_a	a	·
t_b	b	·

y_2

(II)

	t_a	t_b
t_a	a	a
t_b	·	·

y_1

	t_a	t_b
t_a	a	b
t_b	·	·

y_2

(III)

	t_a	t_b
t_a	a	·
t_b	·	a

y_1

	t_a	t_b
t_a	a	·
t_b	·	b

y_2

(I) violates the second inequality of (*) for β_{-i} given by

β^1: $\beta^1(t_a) = t_b$ $\beta^1(t_b) = t_a$ (Player 1 (row) flips)

β^2: $\beta^2(t_a) = t_a$ $\beta^2(t_b) = t_a$ (Player 2 (column) projects to t_a)

Similarly, (II) violates the second inequality of (*) for β_{-i} given by

$\beta^1(t_a) = t_a$ $\beta^1(t_b) = t_a$ (Player 1 (row) projects to t_a)

$\beta^2(t_a) = t_b$ $\beta^2(t_b) = t_a$ (Player 2 (column) flips)

To see that (III) cannot occur, we know from (I) and (II) that the (III) case must be

	t_a	t_b
t_a	a	a
t_b	a	a

y_1

	t_a	t_b
t_a	a	a
t_b	a	b

y_2

This violates the second inequality of (*) for β_{-i} given by

$\beta^1(t_a) = \beta^2(t_a) = t_b$, $\beta^1(t_b) = \beta^2(t_b) = t_a$ (Both players flip)

Hence, $y_2(t_a,t_a) \neq a$,

Similar arguments may be used to show that

$y_2(t_{-i}) \neq a$ for $t_{-i} = (t_a,t_b),(t_b,t_a)$ and (t_b,t_b).■

Step 5 $y_1(t_{-i}) = a$ some t_{-i}.

Proof: If not, then $y_1 = y_2$.■

Step 6 $y_1(t_{-i}) = a$, all t_{-i}.

Proof: Suppose $y_1(t_a,t_a) = a$. Then it is easy to show that $y_1(t_a,t_b) = a$ and $(t_b,t_a) = a$, by arguments as in cases (I) and (II) of step 4. In fact, if $y_1(t_{-i}) = a$ for any t_{-i}, then we must have $y_1(\hat{t}_{-i}) = a$ for "adjacent" $\hat{t}_{-i}$ (i.e. $\hat{t}_{-i}$ and t_{-i} differ in only one component). Hence Step 6 follows almost immediately from Step 5.■

Steps 1-6 imply that $y_1(t_{-i}) = a$ and $y_2(t_{-i}) = b$ for all t_{-i}. However, this violates the first inequality of (*). Therefore, there do not exist any (y_1,y_2) pairs satisfying (*) so x is not implementable.■

REFERENCES

Abreu, D. and A. Sen, "Subgame Perfect Implementation," Princeton University, 1986.

Banks, J. and J. Sobel, "Equilibrium Selection in Signalling Games," *Econometrica* 55: 647-662, 1987.

Cho, I. and D. Kreps, "Signalling Games and Stable Equilibria," *Quarterly Journal of Economics* 1987, 179-221.

Crawford, V. "A Procedure for Generating Pareto-Efficient Egalitarian-Equivalent Allocations," *Econometrica* 47: 49-60, 1979.

van Damme, E., *Refinements of the Nash Equilibrium Concept*, Springer-Verlag, Berlin, 1983.

Dasgupta, P., Hammond, P. and E. Maskin, "The Implementation of Social Choice Rules: Some General Results on Incentive Compatibility," *Review of Economic Studies* 46: 185-216, 1979.

Grossman, S. and M. Perry, "Perfect Sequential Equilibrium," *Journal of Economic Theory*, 39:97-119, 1986.

Harris, M. and R. Townsend, "Resource Allocation with Asymmetric Information," *Econometrica* 49: 33-64, 1981.

Holmstrom, B. and R. Myerson, "Efficient and Durable Decision Rules with Incomplete Information," *Econometrica* 51: 1799-1819, 1983.

Kohlberg, E. and J-F. Mertens, "On the Strategic Stability of Equilibria," *Econometrica*, 54: 1003-1037, 1986.

Maskin, E., "Nash Equilibrium and Welfare Optimality," 1977, Mimeo, MIT.

Maskin, E., "The Theory of Implementation in Nash Equilibrium: A Survey", in *Social Goals and Social Organization: Volume in Memory of Elisha Pazner*, Cambridge University Press, 1986.

Moore, J. and R. Repullo, "Subgame Perfect Implementation," London School of Economics Working Paper, 1986.

Moulin, H. "Dominance Solvable Voting Schemes," *Econometrica* 47: 1337-1352, 1979.

Myerson, R., "Incentive Compatibility and the Bargaining Problem," *Econometrica* 47:61-74, 1979.

Palfrey, T. and S. Srivastava, "Implementation with Incomplete Information in Exchange Economies," Carnegie-Mellon University, 1985.

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

Palfrey, T. and S. Srivastava, "Nash Implementation Using Undominated Strategies" Carnegie-Mellon University, 1986.

Postlewaite, A., "Implementation via Nash Equilibria in Economic Environments," in *Social Goals and Social Organization: A Volume in Memory of Elisha Pazner*, ed. L. Hurwicz, D. Schmeidler, and H. Sonnenschein, 1986.

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

Reichelstein, S., "A Note on Feasible Implementations," University of California-Berkeley, 1985.

Saijo, T., "Strategy Space Reduction in Maskin's Theorem: Sufficient Conditions for Nash Implementation," Cal Tech Working Paper, 1985.

Selten, R., "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games," *International Journal of Game Theory*, 4: 25-55, 1975.