Mechanism Design with Incomplete Information: A Solution to the Implementation Problem

Author(s): Thomas R. Palfrey and Sanjay Srivastava

Source: *Journal of Political Economy*, Vol. 97, No. 3 (Jun., 1989), pp. 668-691
Published by: University of Chicago Press
Stable URL: http://www.jstor.org/stable/1830460
Accessed: 07-03-2016 23:24 UTC

Mechanism Design with Incomplete Information: A Solution to the Implementation Problem

Thomas R. Palfrey

California Institute of Technology

Sanjay Srivastava

Carnegie-Mellon University

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 rule can be made the unique equilibrium outcome to a mechanism. We derive a general necessary condition for unique implementation that implies that the positive result for private-value models applies with considerably less generality to common-value settings.

I. 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 that can be achieved by institutions. Mechanism design theory studies precisely this problem.

We thank the National Science Foundation for financial support under grant SES-8608118 and are grateful to seminar participants at Carnegie-Mellon University, Stanford University, the University of Chicago, and California Institute of Technology for comments. Palfrey gratefully acknowledges the financial support of the Sloan Foundation and the Exxon Education Foundation to the Center for Advanced Study in the Behavioral Sciences. We also wish to thank the referees and Robert Townsend for useful suggestions.

[*Journal of Political Economy*, 1989, vol. 97, no. 3]

An institutional design problem arises whenever a group of individuals with conflicting interests have to make a collective decision. One example of such a problem is the allocation of public goods. Other examples include the design of auctions, constitutional design questions, organized markets such as stock exchanges, and, more generally, contractual agreements between parties such as labor contracts and agency contracts. Since most economic, social, and political activity is organized around institutions, a fundamental problem is the characterization of what institutions can achieve, that is, exactly which collective choices can be attained by institutions.

An important practical reason for studying this characterization stems from the observation that most changes in policy actually change the institutional settings that govern activity. Changes often take place when the outcomes generated by existing institutions are perceived to be undesirable according to some welfare criterion. The question being posed here can be restated as, Given a welfare criterion, does there exist an institution that generates only outcomes that are satisfactory according to the welfare criterion? This question is also related to the line of reasoning employed in the Coase theorem, which asserts that if institutional arrangements are inadequate in the sense of leading to (Pareto) undesirable outcomes, rational agents will move toward an alternative institution that does not have undesirable outcomes. Our analysis can then be viewed as precisely characterizing when such alternative institutions exist. If they do exist, then the outcomes associated with the welfare criterion are said to be implementable. More generally, we are interested in discovering the class of welfare criteria whose outcomes are implementable.

The well-known difficulty in the design problem is that information relevant for determining a satisfactory outcome may be dispersed among the individuals involved. Consequently, in order to achieve an allocation rule that depends on this information (about preferences, endowments, priors over payoff-relevant states of the world, etc.), the rules of the institution must provide appropriate incentives for individuals to share their information. This implies the basic principle of mechanism design with incomplete information, that any outcome that is a Bayesian Nash equilibrium outcome to a mechanism (institution) must satisfy an incentive compatibility condition (Myerson 1979; Harris and Townsend 1981).

This principle further implies an important second idea, known as the *revelation principle,* that any incentive-compatible allocation rule can be made an equilibrium outcome of a very simple type of game: a direct game in which each individual is requested to report his private information. The outcome is then determined by the allocation rule, based exactly on the reported private information of all the individuals. Incentive compatibility is simply the property that, for each individual, the best thing to do in this particular direct game is to report private information *truthfully* as long as all other individuals are also truthfully reporting their private information.

This fundamental insight into mechanism design with incomplete information has allowed many allocation problems to be analyzed and forms the basis for the modern theory of second-best welfare analysis (Holmstrom and Myerson 1983; Prescott and Townsend 1984; Laffont 1985). This insight provided a major technical breakthrough because the analysis of Pareto-optimal allocations in economies with private information could be tractably formulated as a standard programming problem: maximizing a planner's objective function subject to the usual constraints, augmented by an additional set of *incentive* constraints. Furthermore, the revelation principle suggested properties of actual institutions that would be capable of producing these optimal allocations. In this way, the formal analysis of welfare economics and institutions was brought under a single unified approach.

Unfortunately, there is a serious caveat to the "revelation principle" link between institutions and welfare analysis. Incentive compatibility does not imply any restrictions on individual incentives in the direct revelation game if other individuals *are not telling the truth* (Postlewaite and Schmeidler 1987). Consequently, there can (and often will) exist *other* equilibrium outcomes to the direct game that are undesirable (e.g., by the criterion of Pareto optimality). There exist several prominent examples of this problem (Demski and Sappington [1984] and Bhattacharya [1987] in reference to incentive contracts, Milgrom [1981] in auctions, and Palfrey and Srivastava [1987] in the implementation of rational expectations equilibria) that threaten the value of this whole approach to mechanism design.

These recent examples illustrate that the implementation problem has two equally important aspects. In order to implement an allocation rule, a mechanism must be constrained not only by the property that it has *an* equilibrium that produces desirable outcomes but also by the property that *other undesirable outcomes do not arise as equilibria.* Thus the work associated with the revelation principle has elegantly proved that incentive compatibility is a necessary condition for implementation, but the examples cited above indicate that incentive compatibility may not be a sufficient condition.

In this paper, we show that this multiplicity problem can be solved in the large and important class of environments in which private information is of the "private-values" variety; that is, each individual's utility depends only on the outcome and his or her own private information. To achieve this result, a mild refinement of Bayesian Nash equilibrium is adopted and more complex institutions than "direct games" are required. Thus we simultaneously provide a sufficiency proof of incentive compatibility for unique implementation in a broad class of environments and also, via a constructive proof, indicate how our solution may have implications for the details of institutional design.

II. Relation to the Literature

In attempts 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., Selten 1975; Grossman and Perry 1986; Kohlberg and Mertens 1986; Banks and Sobel 1987; Cho and Kreps 1987). 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 1985; Postlewaite 1985; Postlewaite and Schmeidler 1986; Palfrey and Srivastava 1987, in press).

This paper continues a line of inquiry followed by Palfrey and Srivastava (1986) and Moore and Repullo (1988) that merged these two approaches and asked whether flexibility in mechanism design together with a refined equilibrium concept could resolve the multiplicity problem when problems of asymmetric information are absent (i.e., in complete information environments). 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 *asymmetric* information, *all* multiplicity problems can be resolved with a simple strengthening of Bayesian Nash equilibrium: equilibrium in which no individual uses a weakly dominated strategy. This is a mild condition since a weakly dominated strategy is always (weakly) inferior to some other strategy regardless of the strategies employed by the other players and is strictly inferior for some strategies others might use. An important reason for using this refinement is that Bayesian Nash equilibrium places insufficient restrictions on behavior, leading to the implausible use of weakly dominated strategies. This is illustrated clearly by example 2 of Section IV and is precisely the type of behavior excluded by our refinement.

The domain restrictions we impose are that no agent is ever completely indifferent over all alternatives, values are private, and 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 in augmenting a direct mechanism and specifying outcomes so that the desired incentive-compatible allocation rule 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 (in press), 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), many "nice" allocation rules 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 rule that 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, together with a no veto power condition and at least three agents, is also sufficient (Saijo 1988). Monotonicity is satisfied by many economically interesting *sets* of allocation rules. For example, the correspondence that associates each pure exchange neoclassical economy with the set of Pareto-optimal redistributions is monotonic, as is the (constrained) Walrasian correspondence. However, most allocation rules (i.e., single-valued correspondences) are not monotonic and thus not Nash implementable. Moore and Repullo (1988) (see also Abreu and Sen 1986) show that the class of implementable allocation rules expands significantly if the mechanism is played sequentially and subgame perfection is imposed on the equilibrium. Palfrey and Srivastava (1986) have since shown that if there are at least three players and complete indifference is ruled out, then *all* allocation rules 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 environments with private values. 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 environments, in which an agent's preferences may depend on other agents' types, 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 environments and provide an example with common values, which illustrates the strength of the necessary condition, highlights the difficulties arising in these situations, and indicates why positive results in this domain will be more limited.

The private-values model is described in Section III. In Section IV, we provide examples to show why direct mechanisms are generally not sufficient for implementation and also why we need to use refinements of Bayesian Nash equilibrium. Our central possibility result is given in Section V, while extensions to common values are considered in Section VI.

III. 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. The term 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 $T = T^1 \times T^2 \times \ldots \times T^I$. An *allocation rule* is a function x: $T \rightarrow A$. Let $X = \{x: T \rightarrow A\}$ be the set of all allocation rules.

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. To simplify notation, we assume that the support of $q^i(t_i|t_{-i})$ equals 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 narrowed down.

Given an allocation rule $x \in X$, the (interim) expected utility to i conditional on t_i is denoted by

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

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

The term M^i is the *message space* of i, while g is the *outcome function*. If $M^i = T^i$ for all i, then (M, g) is a *direct* mechanism.

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

Given a joint strategy $\sigma = (\sigma^1, \ldots, \sigma^I)$, 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 rule, say x, does there exist a mechanism that has x as its *unique* equilibrium outcome? Following the implementation literature, if there exists such a mechanism, we say that the allocation rule is *implementable*.

We will study implementation using two concepts of equilibrium. These are as follows. 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,

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

(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 exist i, t_i, and $\tilde{\sigma}^i$: $T^i \rightarrow M^i$ such that $V^i[g(\tilde{\sigma}^{-i}, \tilde{\sigma}^i(t_i)), t_i] \geq V^i[g(\tilde{\sigma}^{-i}, \sigma^i(t_i)), t_i]$ 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 at t_i by using $\tilde{\sigma}^i(t_i)$ instead of $\sigma^i(t_i)$, while for some strategy combination of the others, he does strictly better at t_i by using $\tilde{\sigma}^i(t_i)$.

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 rule that 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: $T \rightarrow A$ is *incentive compatible* if for all i, for all t_i,

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

The following result is well known.

THEOREM 1. If x is implementable, then x is incentive compatible.

IV. Eliminating Equilibria by Indirect Mechanisms

To begin our analysis, we consider implementation using Bayesian equilibrium as the solution concept. We start with an example showing how indirect mechanisms help alleviate the multiple equilibrium problem.

Example 1.—Consider a pure exchange economy with two goods, an aggregate endowment $\bar{w} \in R^2_+$, and two agents. Agent 1 can be of two types, $T^1 = \{t_1, t_1'\}$, while agent 2 has only one type, so $T^2 = \{t_2\}$. Preferences are as in figure 1, and each type of agent 1 is equally likely. Consider the allocation rule given in figure 1. It is easy to check that x is incentive compatible, and it is also (ex post) Pareto optimal. The direct mechanism is $M^1 = T^1$, $M^2 = T^2$, so the game can be written as follows:

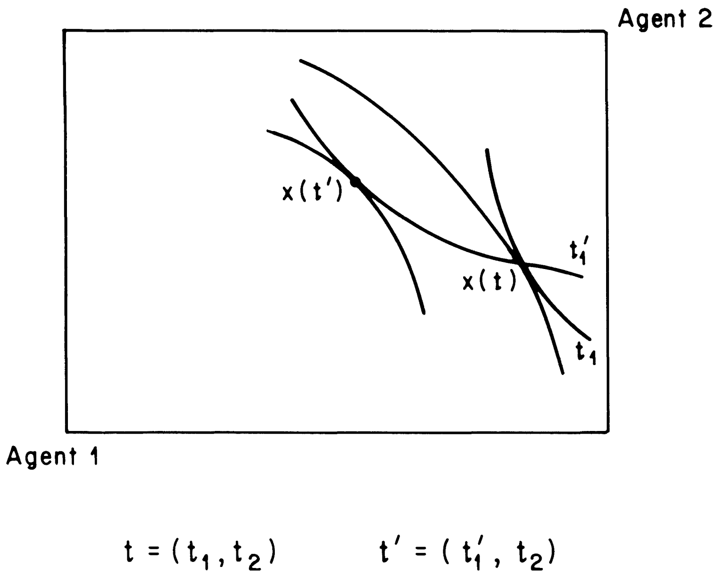

FIG. 1

Agent 1	Agent 2
	t_2
t_1	$x(t)$
t_1'	$x(t')$

Truth telling is clearly an equilibrium to this game, yielding x as the truthful equilibrium outcome. However, this game has another equilibrium, one in which agent 1 says t_1 independently of his type. This is an equilibrium because at t_1' agent 1 is indifferent between $x(t)$ and $x(t')$. Unfortunately, agent 2 is not indifferent between this strategy and the truthful one: he strictly prefers the outcome when agent 1 reports truthfully. Further, if agent 1 always reports t_1, the outcome at t' is inefficient.

In order to overcome this problem, we can attempt to expand the strategy sets of the agents (use an indirect mechanism) or refine the concept of equilibrium (or both). In this particular example, refinements such as undominated equilibrium, (trembling-hand) perfect equilibrium, or proper equilibrium do not rule out the bad equilibrium in the direct mechanism since they would all rely on possible mistakes made by agent 2. Since agent 2 has only one strategy, the refinements do not help. We now show that a simple indirect mechanism can eliminate the problem in this example.

Consider adding a strategy for agent 2, say N, with the following outcomes:

	Agent 2	
Agent 1	t_2	N
t_1	$x(t)$	$x(t')$
t_1'	$x(t')$	$x(t)$

Now note that if 1 always plays t_1, 2 should play N, in which case 1 should play t_1' when he is of type t_1. Thus the bad equilibrium has been eliminated. It can be verified that there are two equilibria to this game:

$$\sigma^1(t_1) = t_1, \quad \sigma^1(t_1') = t_1', \sigma^2 = t_2$$

and

$$\sigma^1(t_1) = t_1', \quad \sigma^1(t_1') = t_1, \sigma^2 = N.$$

In either equilibrium, the outcome at t is $x(t)$, and that at t' is $x(t')$.

In this example, then, a simple extension of the mechanism implemented the desired allocation rule. This naturally raises the question of when indirect mechanisms by themselves are sufficient to implement desirable allocation rules.

To answer this question, consider an incentive-compatible allocation rule x. The associated direct mechanism is $M^i = T^i$ for all i, and $g(t) = x(t)$ for all t. Incentive compatibility ensures that truth telling is an equilibrium to this direct game, yielding x as the truthful equilibrium outcome. As in the example, however, there may be other equilibria to the direct mechanism, and the question is whether these can be eliminated by expanding the mechanism. To examine what these equilibria might be, we first need to examine all possible strategies agents might use. In a direct mechanism, any strategy for agent i is a function from T^i into T^i, say $\alpha^i: T^i \rightarrow T^i$. Truth telling is simply the identity function. We call α^i a *deception* by i, the interpretation being that when i is of type t_i, he acts as if he is of type $\alpha^i(t_i)$.

In the example, the "bad" equilibrium strategy by agent 1 was the deception $\alpha^1(t_1) = t_1, \alpha^1(t_1') = t_1$. With this notation, incentive compatibility can be rewritten as, for all i, for all t_i, and for all $\alpha^i: T^i \rightarrow T^i$,

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

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 *not* say what is a best response when other agents are playing deceitfully.

Let $\alpha = (\alpha^1, \ldots, \alpha^I)$ and $\alpha^{-i} = (\alpha^1, \alpha^2, \ldots, \alpha^{i-1}, \alpha^{i+1}, \ldots, \alpha^I)$ so that $\alpha = (\alpha^{-i}, \alpha^i)$. Then every candidate for equilibrium in the direct game is a (joint) deception α. If α is being used, the outcome to the direct game is x_α, where $x_\alpha(t) = x(\alpha(t))$ for all t. If α is an equilibrium and $x_\alpha \neq x$, then we have an undesired outcome to the direct game. The question being posed can now be rephrased to ask when it is possible to add strategies to the direct game so that any α with $x_\alpha \neq x$ is not an equilibrium. To answer this, fix α such that $x_\alpha \neq x$. For any agent i, consider giving him an additional message, say m^i. For each t, let $y(t_{-i}, \alpha^i(t_i)) = g(t_{-i}, m^i)$, the outcome when i plays m^i and the other agents play t_{-i}. To ensure that α is not an equilibrium, we want agent i to play m^i when the other agents are playing α^{-i}. If i plays m^i and the others use α^{-i}, the outcome at t is $y(\alpha^{-i}(t_{-i}), \alpha^i(t_i))$. If i uses α^i and the others use α^{-i}, the outcome at t is $x(\alpha^{-i}(t_{-i}), \alpha^i(t_i))$. Thus x_α is an equilibrium outcome unless, at some t_i,

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

If there exist i, t_i, and y such that the inequality above is satisfied, then α cannot be an equilibrium since agent i will deviate to m^i. However, we must also be careful that introducing m^i does not lead to x not being an equilibrium outcome; that is, we still want truth telling to be an equilibrium. Thus we must also have that, *for all* $t_i' \in T^i$,

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

Defining $y_\alpha(t) = y(\alpha^{-i}(t_{-i}), \alpha^i(t_i))$, we arrive at the condition called Bayesian monotonicity, which is necessary for implementation.

DEFINITION 7. x: $T \rightarrow A$ satisfies *Bayesian monotonicity* if, for any deception α such that $x_\alpha(t) \neq x(t)$ for some t, there exist i, t_i, and an allocation rule y: $T \rightarrow A$ such that

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

and

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

The next example shows that appealing to indirect mechanisms alone will generally not be enough to solve the implementation problem.

Example 2.—$I = 3$, $A = \{a, b\}$, and $T^i = \{t_a, t_b\}$ for all i. Types are independently drawn with $q(t_b) = q$ for all i and $q^2 > .5$. Preferences are as follows: type t_a strictly prefers a to b, while type t_b strictly prefers 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 rule, x:

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

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

This allocation rule has many nice properties and, indeed, is the only *reasonable* allocation rule 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 (1983); (iii) $x(t)$ is the (unique) majority winner at t; (iv) it maximizes an Arrow social welfare function; and (v) it can be implemented in dominant strategies by a direct mechanism.

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 that 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 since $x_\alpha \equiv b$ violates properties ii, iii, and iv.

To show that Bayesian monotonicity is not satisfied requires us to prove that there does not exist y: $T \rightarrow A$ that satisfies the first set of inequalities in definition 7, 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 rule. 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 $y_\alpha(t) = a$ for all t. Since a is the worst element for type t_b, the inequality

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

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 7 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 7. Since the problem is symmetric, we need consider only agent 1. The expected utility from x at t_a is $1 - q^2$ while that from

$y(t_{-1}, \alpha^1(t_a))$ at t_a is

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

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

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

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

Palfrey and Srivastava (1987) present several more examples of reasonable allocation rules that are not Bayesian implementable even in pure exchange economies. These include allocation rules defined by various notions of optimality and by various notions of equity and fairness. In the next section, we show that these problems may be solved when indirect mechanisms are used together with our mild refinement of Bayesian equilibrium.

V. Undominated Bayesian Equilibrium

In this section, using Bayesian equilibria that do not involve the use of weakly dominated strategies, we prove the central result of the paper: *Any allocation rule that satisfies incentive compatibility can be made the unique equilibrium outcome to a mechanism in a large class of models.*

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 if two types are different, then their preferences over some pair of alternatives must be different.

DEFINITION 8. *Value-distinguished types.*—For all i, t_i, t_i', and $t_i \neq t_i'$, 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 allocation rules. This will be the case if, for example, the difference between types is the difference in risk aversion. In this case, types are value distinguished on the set of lotteries over A. Therefore, one may think of A more generally as a set of lotteries and the $U^i(\cdot)$ as preferences over lotteries.

Our sufficiency result requires us also to impose the following two mild restrictions on the domain of possible types. The first states that there is no type for whom all alternatives give equal utility. The second is that each type has a most preferred and a least preferred alternative. The latter is implied, for example, if each U^i is continuous and A is compact.

DEFINITION 9. (i) *No complete indifference.*—For all i and t_i, there exist $a, b \in A$ with $U^i(a, t_i) > U^i(b, t_i)$. (ii) *Existence of best and worst elements.*—For all i and t_i, there exist $b(t_i)$, $w(t_i) \in A$ with $U^i(b(t_i), t_i) \geq U^i(a, t_i)$ for all $a \in A$ and $U^i(a, t_i) \geq U^i(w(t_i), t_i)$ for all $a \in A$.

THEOREM 2. Assume that $I \geq 3$, that there is no complete indifference, that best and worst elements exist, 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.

The Appendix contains a formal proof of theorem 2 and a detailed construction of a general implementing mechanism. Here we give the intuition behind the construction of the mechanism and explain how it works.

Following the intuition behind the examples of the previous section, we see that the mechanism is, effectively, a direct mechanism with some additional strategies appended in a way that eliminate undesirable equilibria. Each agent submits a message that has four components. The first component is from the "direct" part of the message space: $M_1^i = T^i$. The second component is either a second report of one's own type or a report of someone else's type. The third component of the message space is a half-open real interval that is used in the mechanism to break ties. The fourth component is a requested outcome. Formally, let

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

where $M_1^i = T^i$, $M_2^i = \cup_j T^j$, $M_3^i = [0, I + 2)$, and $M_4^i = A$. The key aspect of the mechanism is that, except for specific isolated portions of M, the outcome function, g, is essentially direct in that it depends only on the first component of each agent's report. Calling this region D_0, we have $g(m) \equiv x(m_1)$ for all $m \in D_0$. The remainder of M is divided into I subregions indexed by i. In such a region, D^i, $m_2^j = \hat{t}_i \in T^i$ for all $j \neq i$. The outcome function in such a region is given in table 1.

The entries in the table are to be interpreted as follows. The four-tuples defining columns and rows are strategy choices by agent i (columns) and by all other agents (rows). We have denoted by a_{i*} the outcome requested by the agent (i*) who wins the tie-breaking procedure as determined by m_3. The outcomes $y_1(t_i, t_i')$ and $y_2(t_i, t_i')$ are a pair of allocations for which i's preferences differ depending on whether i is type t_i or t_i'. The existence of the pair is guaranteed by the assumption of value-distinguished types. In fact, the proof given in the Appendix is only for *strictly* value-distinguished types. Therefore,

TABLE 1

OUTCOME FUNCTION RESTRICTED TO REGION D^i

Message of All $j \neq i$	Message of Agent 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 $(t_i, t_j, k_i, a_i), j \neq i$
1. (t_j, t_i, k_j, a_j), $k_j \in [I + 1, I + 2)$	a_i	a_i	a_{i*}
2. (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_{i*}
3. (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_{i*}
4. (t_j, t_i, k_j, a_j), $k_j \in [i, i + 1)$	a_i	a_i	$w(t_i)$
5. (t_j, t_i'', k_j, a_j), $k_j \in [i, i + 1)$, $t_i'' \neq t_i$	$w(t_i'')$	$w(t_i'')$	$w(t_i'')$
6. All other messages with $m_2^j = \hat{t}_i \in T^i$	a_{i*}	a_{i*}	a_{i*}

$y_1(t_i, t_i')$ and $y_2(t_i, t_i')$ have the property that $U^i(y_1(t_i, t_i'), t_i) > U^i(y_2(t_i, t_i'), t_i)$ but $U^i(y_2(t_i, t_i'), t_i') > U^i(y_1(t_i, t_i'), t_i')$. Straightforward methods for extending the mechanism to account for *weak* value distinction are contained in Palfrey and Srivastava (1986).

The proof then proceeds in three steps: (1) all equilibria must lie in D_0, (2) all equilibria must involve "truthful" reports (i.e., $m_1^i = t_i$ for all i, t_i), and (3) the joint strategy where $\sigma^i(t_i) = (t_i, t_i, 0, b(t_i))$ for all i is an equilibrium.

To prove step 1, we show that no equilibrium can lie in D^i for any i. Suppose that agent i is of type t_i. Note first that reporting $m_4^i = a$ 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 that $m_4^i = b(t_i)$.

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

A similar argument applies for i if $m_2^i \neq m_1^i$. We conclude that all equilibria must lie in D_0, with $m_3^j = 0$ and $m_1^j = m_2^j$ 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 alters the outcome only 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 change the outcome only from $x(t)$ to $x(t_{-i}, t_i')$ at t. To see that it is not weakly dominated, we have to consider each possible deviation by i and all possible strategies by others. 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 differ only in the last component of the message space.

VI. Common Values

The most significant assumption in theorem 2 is private values. Even though a large majority of applications to date of Bayesian games to economic problems and applications of the revelation principle to mechanism design have used this assumption, it is clearly quite restrictive. Our general possibility result does not apply with nearly the same force in settings with common values, which 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 we still assume that, at t, i observes only t_i and that there is no moving support. Economic examples of common-value allocation problems include oil lease auctions studied by Wilson (1977) and others and oligopoly with private information about demand studied by Palfrey (1985) and others. In the auction, n bidders submit competitive bids for the right to drill for oil at a specified location. The oil they drill for has a common value to all bidders, but they differ in their (correlated) private estimates of how much oil will be found and recovered. In the oligopoly setting, firms face a common demand curve for a homogeneous product but have different (correlated) estimates of the parameters of the demand curve. These estimates are privately known. In both of these examples, t_i corresponds to an individual estimate, and U^i corresponds to a conditional expected value of the oil or output, net of an accepted bid or production costs. This conditional expectation will generally be different when conditioned on the entire vector of estimates rather than being conditioned only on one's own private estimate. For this reason U^i is a function of t instead of just t_i. The definition of incentive compatibility is now modified accordingly.

DEFINITION 6$'$. x is *incentive compatible* if for all i, for all t_i,

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

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

THEOREM 3. If x is implementable in undominated Bayesian equilibrium, then x is incentive compatible, and for any α: $T \to T$, $x_\alpha(t) \neq x(t)$ for some t implies that at least one of the following conditions holds: (a) There exist i, t_i, and $y \in X$ with

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

for all $t_i' \in T_i$ and

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

(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 the Appendix.

With private values, parts b and c of this result reduce to the statement that types are value distinguished. For example, consider part 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 not that different from (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 3.—$A = \{a, b\}$, $I = 3$, $T^i = \{t_a, t_b\}$ for all i, and types are independently drawn with $q^i(t_b) = q$ for all i and $q^2 > .5$. The agents have "majoritarian" preferences, given by

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

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

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

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

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

Surprisingly, this allocation rule 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.

Appendix

Proof of Theorem 2

We divide the message space as follows:

$$D_1 = \{m | m_2^j \in T^j \text{ for all } j\},$$

$$D_2 = \{m | \text{there does not exist } i \text{ and } t_i \in T^i \text{ with } m_2^j = t_i \ \forall j \neq i\}.$$

Let

$$D^i = \{m \mid \text{there exists } t_i \in T^i \text{ with } m_2^j = t_i \; \forall j \neq i\}.$$

Note that $M \setminus (D_1 \cup D_2) = \cup_{i=1}^{I} D^i$. Let

$$D_{3A}^i = \{m \in D^i \mid \forall j \neq i,\ m_2^j \neq m_1^i,\ m_3^i \in [I+1, I+2);\ m_1^i = m_2^i,\ m_3^i = 0\},$$

$$D_{3B}^i = \{m \in D^i \mid \forall j \neq i,\ m_2^j = m_1^i,\ m_3^i \in [I+1, I+2);\ m_1^i = m_2^i,\ m_3^i = 0\},$$

$$D_{4A}^i = \{m \in D^i \mid \forall j \neq i,\ m_2^j = m_2^i,\ m_3^i \in [I+1, I+2);\ m_1^i \neq m_2^i\},$$

$$D_{4B}^i = \{m \in D^i \mid \forall j \neq i,\ m_2^j \neq m_1^i,\ m_2^j \neq m_2^i,\ m_3^i \in [I+1, I+2);\ m_1^i \neq m_2^i\},$$

$$D_5^i = \{m \in D^i \mid \forall j \neq i,\ m_2^j = m_1^i,\ m_3^j \in [i, i+1)\},$$

$$D_{5A}^i = \{m \in D_5^i \mid m_1^i \neq m_2^i \text{ or } m_1^i = m_2^i \text{ and } m_3^i \neq 0\},$$

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

For $m \in D_6$, let i^* be the smallest i such that $m_3^i \geq m_3^j$ for all j, and let $a_{i^*} = m_4^{i^*}$. The outcome function is given by

$$g(m) = \begin{cases} x(t) & \text{if } m \in D_1 \text{ and } m_1 = t \\ x(t) & \text{if } m \in D_2 \text{ and } m_1 = t \\ y_1(t_i, t_i') & \text{if } m \in D_{3A}^i \text{ and } m_1^i = t_i,\ m_2^j = t_i' \ \forall j \neq i \\ a_i & \text{if } m \in D_{3B}^i \text{ and } m_4^i = a_i \\ y_2(t_i, t_i') & \text{if } m \in D_{4A}^i \text{ and } m_1^i = t_i,\ m_2^j = t_i' \ \forall j \neq i \\ a_i & \text{if } m \in D_{4B}^i \text{ and } m_4^i = a_i \\ a_i & \text{if } m \in D_{5A}^i \text{ and } m_4^i = a_i \\ w(t_i) & \text{if } m \in D_5^i \setminus D_{5A}^i \text{ and } m_2^j = t_i \ \forall j \neq i \\ a_{i^*} & \text{if } m \in D_6. \end{cases}$$

We start by showing that $\sigma^i(t_i) = (t_i, t_i, 0, b(t_i))$ for all i and t, which lies in D_1 for all t, is a Bayesian equilibrium. This can be seen by noting that a unilateral deviation by i from this strategy affects the outcome only if i changes m_1^i. (Note that this would not be true if $I = 2$ since in that case $D_2 \cap (D_3^i \cup D_4^i) \neq \varnothing$.) If, at t_i, i instead reports $m_1^i = t_i'$, 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 agent's favor. Without loss of generality, therefore, we assume that $m_4^i = b(t_i)$ 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) $m_1^i \neq t_i$: In this case, i is strictly worse off when $m^j = (t_j, t_i, i, a_i)$ for all $j \neq i$ 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 $m^j = (t_j, t_i, i, a_j)$ for all $j \neq i$. (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 $m^j = (t_j, t_i, i, a_j)$ for all $j \neq i$ 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 $m_3^i > 0$ for some i. To see this, let J be an integer such that $J \leq m_3^i < J + 1$. Then, reporting $\hat{m}^i = m^i$ except $\hat{m}_3^i = (m_3^i + J + 1)/2$ 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 $g(m^{-i}, \hat{m}^i) = m_4^i = b(t_i)$, and no configuration of messages such that $U^i(g(m), t_i) > U^i(g(m^{-i}, \hat{m}^i), t_i)$. Second, $\sigma^i(t_i) = (t_i', t_j, 0, b(t_i))$ is weakly dominated by $\bar{\sigma}^i(t_i) = (t_i', t_i, ½, b(t_i))$ and $\sigma^i(t_i) = (t_i', t_i'', 0, b(t_i))$ with $t_i' \neq t_i''$ is weakly dominated by $\bar{\sigma}^i(t_i) = (t_i', t_i'', ½, b(t_i))$. This leaves only $\sigma^i(t_i) = (t_i', t_i', 0, b_i)$, where b_i is a best element at t_i. We claim that $m^i = (t_i', t_i, k_i, b_i)$ weakly dominates this strategy. The outcome changes only 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 $U^i(y_2(t_i', t_i), t_i) > U^i(y_1(t_i', t_i), t_i)$; i is no worse off in D_6. This concludes the proof of theorem 2.

Proof of Theorem 3

The revelation principle implies that 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 t. 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 condition a.

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

$$\int U^i(g(\tilde{\sigma}^{-i}, \bar{m}^i), t)dq(t_{-i}|t_i) \geq \int U^i(g(\tilde{\sigma}^{-i}, \sigma_\alpha^i(t_i)), t)dq(t_{-i}|t_i) \qquad \text{(A1)}$$

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

Let $\bar{\sigma}^i(t_i') = \bar{m}^i$ for all t_i'. Since σ is not weakly dominated at $\alpha^i(t_i)$, we get either

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

or

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

Let $y_1 = g(\bar{\sigma}^{-i}, \sigma^i)$ and $y_2 = g(\bar{\sigma}^{-i}, \bar{\sigma}^i)$.

Case i

Substituting for y_1 and y_2 in inequality i yields $V^i(y_1, \alpha^i(t_i)) > V^i(y_2, \alpha^i(t_i))$. By hypothesis, $\sigma_\alpha^i(t_i) = \sigma^i(\alpha^i(t_i))$ is weakly dominated by $\bar{m}^i$ at t_i. Let $y_{1\alpha}^i = g(\bar{\sigma}^{-i}, \sigma_\alpha^i)$, $y_{2\alpha}^i = g(\bar{\sigma}^{-i}, \bar{\sigma}_\alpha^i)$. Here, $y_{1\alpha}^i$ is the outcome when i plays σ_α^i and all other agents play $\bar{\sigma}^{-i}$, and $y_{2\alpha}^i$ is the outcome when i plays $\bar{\sigma}_\alpha^i$ and all other agents play $\bar{\sigma}^{-i}$. Note that $\bar{\sigma}_\alpha^i = \bar{\sigma}^i$ since $\bar{\sigma}^i$ is a constant strategy.

Replacing $\tilde{\sigma}^{-i}$ with $\bar{\sigma}^{-i}$ in (A1) then yields $V^i(y_{1\alpha}^i, t_i) \leq V^i(y_{2\alpha}^i, t_i)$. For any

β^{-i}, let $y^i_{1\alpha\beta} = g(\bar{\sigma}_\beta^{-i}, \sigma^i_\alpha)$, $y^i_{2\alpha\beta} = g(\bar{\sigma}_\beta^{-i}, \bar{\sigma}_\alpha^{-i})$, so that $y^i_{1\alpha\beta}$ is the outcome when i plays σ^i_α and all other agents use deceptions β^j, and similarly for $y^i_{2\alpha\beta}$.

Replacing $\tilde{\sigma}^{-i}$ with $\bar{\sigma}_\beta^{-i}$in (A1) yields $V^i(y^i_{1\alpha\beta}, t_i) \leq V^i(y^i_{2\alpha\beta}, t_i)$ for all β^{-i}. Next, note that for any β with $\beta^i = \alpha^i$, $y_{1\beta} = y^i_{1\alpha\beta}$ and $y_{2\beta} = y^i_{2\alpha\beta}$. We have thus shown the existence of y_1 and y_2 with

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

and

$$V^i(y_{1\beta}, t_i) \leq V^i(y_{2\beta}, t_i) \quad \text{for all } \beta \text{ with } \beta^i = \alpha^i.$$

To complete case i, it must also be the case that $V^i(g(\tilde{\sigma}^{-i}, \bar{\sigma}^i), t_i) > V^i(g(\tilde{\sigma}^{-i}, \sigma^i_\alpha), t_i)$ for some $\tilde{\sigma}^{-i}$. Let $z_1 = g(\tilde{\sigma}^{-i}, \bar{\sigma}^i)$ and $z_2 = g(\tilde{\sigma}^{-i}, \sigma^i_\alpha(t_i))$. 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, and 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 $\beta^i = \alpha^i$,

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

and

$$V^i(z_{1\beta}, t_i) \geq V^i(y_{2\beta}, t_i)$$

for any deception β with $\beta^i = \alpha^i$. This is precisely the requirement in condition b.

Case ii

In this case,

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

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

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

and

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

Letting $y_2 = g(\bar{\sigma}^{-i}, \bar{\sigma})$ and $y_1 = g(\bar{\sigma}^{-i}, \sigma^i_\alpha)$, 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 condition c and concludes the proof.

Proof of Claim in Example 3

We prove the claim in two parts: (A) If σ is a Bayesian equilibrium to (M, g) with $g(\sigma) = x$, then σ_α is a Bayesian equilibrium. (B) If σ is an undominated Bayesian equilibrium to (M, g), with $g(\sigma) = x$, then σ_α is undominated.

Proof of Part A

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

Proof of Part B

This is more involved and requires meticulous checking that none of the conditions a, b, and c of theorem 3 are satisfied. Since part A of the proof already implies that condition a is not satisfied, we need only show that conditions b and c are not satisfied. In this simple example, the proof reduces to showing that there does not exist a pair of allocations y_1, y_2 such that

$$V^i(y_1, t_b) > V^i(y_2, t_b) \text{ but } V^i(y_{1\beta}, t_a) \leq V^i(y_{2\beta}, t_a) \quad \forall\, \beta \text{ with } \beta^i = \alpha^i. \quad \text{(A2)}$$

This can be proved in a series of steps. First, without loss of generality, fix $i = 3$.

Step 1.—It suffices to show that there do not exist allocation rules y: $T^{-3} \rightarrow A$ (i.e., allocation rules that are constant in player 3's type) such that $V^3(y_1, t_b) > V^3(y_2, t_b)$ but $V^3(y_{1\beta^{-3}}, t_a) \leq V^3(y_{2\beta^{-3}}, t_a)$ for all β^{-3}.

Proof.—This follows immediately from the fact that both inequalities of (A2) hold the argument of y_1 and y_2 corresponding to player 3's type fixed at t_b. Q.E.D.

Thus we are reduced to a search of all pairs y_1 and y_2 that can be represented by 2 × 2 outcome matrices, as follows:

Player 1's type	Player 2's type: t_a	Player 2's type: t_b
t_a	$y_1(t_a, t_a)$	$y_1(t_a, t_b)$
t_b	$y_1(t_b, t_a)$	$y_1(t_b, t_b)$

y_1

Player 1's type	Player 2's type: t_a	Player 2's type: t_b
t_a	$y_2(t_a, t_a)$	$y_2(t_a, t_b)$
t_b	$y_2(t_b, t_a)$	$y_2(t_b, t_b)$

y_2

The remainder of the proof involves a demonstration that there is no way to fill in the matrices above with a's and b's in such a way that (A2) is satisfied.

Step 2.—If some entry in y_2 is a, the corresponding entry in y_1 is a; that is, $y_2(t_{-i}) = a \Rightarrow y_1(t_{-i}) = a$ for all t_{-i}.

Proof.—Suppose that $y_2(\hat{t}_{-i}) = a$ but that $y_1(\hat{t}_{-i}) = b$ for some $\hat{t}_{-i}$. Then the second inequality of (A2) is violated for the β^{-i} that projects to $\hat{t}_{-i}$; that is, $\beta^{-i}(t_{-i}) = \hat{t}_{-i}$ for all t_{-i}. Q.E.D.

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 (A2) is violated. Q.E.D.

Step 4.—$y_2(t_{-i}) = b$ for all t_{-i}.

Proof.—Suppose that $y_2(t_a, t_a) = a$. 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)

y_1	t_a	t_b
t_a	a	·
t_b	a	·

y_2	t_a	t_b
t_a	a	·
t_b	b	·

(II)

	t_a	t_b
t_a	a	a
t_b	·	·

	t_a	t_b
t_a	a	b
t_b	·	·

(III)

	t_a	t_b
t_a	a	·
t_b	·	a

	t_a	t_b
t_a	a	·
t_b	·	b

Case I violates the second inequality of (A2) for β^{-i} given by

$$\beta^1\colon\ \beta^1(t_a) = t_b,\ \beta^1(t_b) = t_a \quad \text{(player 1 [row] flips),}$$

$$\beta^2\colon\ \beta^2(t_a) = t_a,\ \beta^2(t_b) = t_a \quad \text{(player 2 [column] projects to } t_a\text{).}$$

Similarly, case II violates the second inequality of (A2) for β^{-i} given by

$$\beta^1(t_a) = t_a,\ \beta^1(t_b) = t_a \quad \text{(player 1 [row] projects to } t_a\text{),}$$

$$\beta^2(t_a) = t_b,\ \beta^2(t_b) = t_a \quad \text{(player 2 [column] flips).}$$

To see that case III cannot occur, we know from I and II that case III 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 (A2) for β^{-i} given by

$$\beta^1(t_a) = \beta^2(t_a) = t_b,\ \beta^1(t_b) = \beta^2(t_b) = t_a \quad \text{(both players flip).}$$

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). Q.E.D.

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

Proof.—If not, then $y_1 = y_2$. Q.E.D.

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

Proof.—Suppose that $y_1(t_a, t_a) = a$. 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. Q.E.D.

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 (A2). Therefore, there do not exist any (y_1, y_2) pairs satisfying (A2), so x is not implementable. Q.E.D.

References

Abreu, Dilip, and Sen, A. "Subgame Perfect Implementation." Manuscript. Princeton, N.J.: Princeton Univ., 1986.

Banks, Jeffrey S., and Sobel, Joel. "Equilibrium Selection in Signaling Games." *Econometrica* 55 (May 1987): 647–64.

Bhattacharya, Sudipto. "Tournaments, Termination Schemes and Forcing Contracts." Paper no. 34. Morristown, N.J.: Bell Communications Res., 1987.

Cho, In-Koo, and Kreps, David M. "Signaling Games and Stable Equilibria." *Q.J.E.* 102 (May 1987): 179–221.

Crawford, Vincent P. "A Procedure for Generating Pareto-Efficient Egalitarian-Equivalent Allocations." *Econometrica* 47 (January 1979): 49–60.

Dasgupta, Partha S.; Hammond, Peter J.; and Maskin, Eric S. "The Implementation of Social Choice Rules: Some General Results on Incentive Compatibility." *Rev. Econ. Studies* 46 (April 1979): 185–216.

Demski, Joel S., and Sappington, David. "Optimal Incentive Contracts with Multiple Agents." *J. Econ. Theory* 33 (June 1984): 152–71.

Grossman, Sanford J., and Perry, Motty. "Perfect Sequential Equilibrium." *J. Econ. Theory* 39 (June 1986): 97–119.

Harris, Milton, and Townsend, Robert M. "Resource Allocation under Asymmetric Information." *Econometrica* 49 (January 1981): 33–64.

Holmstrom, Bengt, and Myerson, Roger B. "Efficient and Durable Decision Rules with Incomplete Information." *Econometrica* 51 (November 1983): 1799–1819.

Kohlberg, Elon, and Mertens, Jean-François. "On the Strategic Stability of Equilibria." *Econometrica* 54 (September 1986): 1003–37.

Laffont, Jean-Jacques. "On the Welfare Analysis of Rational Expectations Equilibria with Asymmetric Information." *Econometrica* 53 (January 1985): 1–29.

Maskin, Eric S. "Nash Equilibrium and Welfare Optimality." Manuscript. Cambridge: Massachusetts Inst. Tech., 1977.

———. "The Theory of Implementation in Nash Equilibrium: A Survey." In *Social Goals and Social Organization: Essays in Memory of Elisha Pazner*, edited by Leonid Hurwicz, David Schmeidler, and Hugo Sonnenschein. Cambridge: Cambridge Univ. Press, 1985.

Milgrom, Paul R. "Rational Expectations, Information Acquisition, and Competitive Bidding." *Econometrica* 49 (July 1981): 921–43.

Moore, John, and Repullo, Rafael. "Subgame Perfect Implementation." *Econometrica* 56 (September 1988): 1191–1220.

Moulin, Hervé. "Dominance Solvable Voting Schemes." *Econometrica* 47 (November 1979): 1337–51.

Myerson, Roger B. "Incentive Compatibility and the Bargaining Problem." *Econometrica* 47 (January 1979): 61–73.

Palfrey, Thomas R. "Uncertainty Resolution, Private Information Aggregation, and the Cournot Competitive Limit." *Rev. Econ. Studies* 52 (January 1985): 69–83.

Palfrey, Thomas R., and Srivastava, Sanjay. "Nash Implementation Using Undominated Strategies." Manuscript. Pittsburgh: Carnegie-Mellon Univ., 1986.

———. "On Bayesian Implementable Allocations." *Rev. Econ. Studies* 54 (April 1987): 193–208.

———. "Implementation with Incomplete Information in Exchange Economies." *Econometrica* (in press).

Postlewaite, Andrew. "Implementation via Nash Equilibria in Economic Environments." In *Social Goals and Social Organization: Essays in Memory of Elisha Pazner*, edited by Leonid Hurwicz, David Schmeidler, and Hugo Sonnenschein. Cambridge: Cambridge Univ. Press, 1985.

Postlewaite, Andrew, and Schmeidler, David. "Implementation in Differential Information Economies." *J. Econ. Theory* 39 (June 1986): 14–33.

———. "Differential Information and Strategic Behavior in Economic Environments: A General Equilibrium Approach." In *Information, Incentives and Economic Mechanisms: Essays in Honor of Leonid Hurwicz*, edited by Theodore Groves, Roy Radner, and Stanley Reiter. Minneapolis: Univ. Minnesota Press, 1987.

Prescott, Edward C., and Townsend, Robert M. "Pareto Optima and Competitive Equilibria with Adverse Selection and Moral Hazard." *Econometrica* 52 (January 1984): 21–45.

Reichelstein, S. "A Note on Feasible Implementations." Manuscript. Berkeley: Univ. California, 1985.

Saijo, T. "Strategy Space Reduction in Maskin's Theorem: Sufficient Conditions for Nash Implementation." *Econometrica* 56 (May 1988): 693–700.

Selten, Reinhard. "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games." *Internat. J. Game Theory* 4, no. 1 (1975): 25–55.

Wilson, Robert. "A Bidding Model of Perfect Competition." *Rev. Econ. Studies* 44 (October 1977): 511–18.