

POSTERIOR IMPLEMENTABILITY IN A TWO-PERSON DECISION PROBLEM¹

BY JERRY R. GREEN AND JEAN-JACQUES LAFFONT

When a decision rule is implemented using a Bayesian incentive compatible mechanism in which the messages are publicly observable, the players' information is augmented by their observation of each others' strategies. In this paper we study the set of Bayesian implementable decision rules which have the further property that the information conveyed in the process of their implementation does not invalidate the optimality of the players' strategies. We call such rules *posterior implementable*.

We concentrate on a two-person problem with two possible decisions, and, for this problem, we obtain a complete characterization of the set of posterior implementable decision rules. Our main result entails that a posterior implementable social decision rule can take essentially only two values throughout the range of observations of the two players. The domains over which each of these two outcomes is realized can be characterized by the fact that the boundary between them is a step function satisfying a certain set of equations.

The motivation for studying posterior implementable rules is that they represent the outcomes of a two-stage cooperative process. In the first stage, communication takes place but no binding commitments can be made. The second stage consists of ratifying ("signing") the agreement obtained at the first stage. At this state, no further information is conveyed. Both parties must be satisfied with the nonbinding commitments obtained at the first stage, so that these actions are actually carried out.

Possible applications of this theory are given. A constructive method for finding the posterior implementable rules is presented and the set of such rules is contrasted, in an example, with the full set of Bayesian implementable rules.

KEYWORDS: Collective decision making, incentives, Bayesian implementation, posterior implementation, information, commitment, mechanism design.

INTRODUCTION

WE STUDY A MODEL of collective decision-making in which each of two parties has private information that is relevant to the value of a joint action that they are considering. The particular economic environment we examine and the solution concept we employ are both new. It is probably best to begin with a general discussion of the context and objectives of this paper before going on to the details.

In many realistic instances, the process of forming contractual arrangements takes place in two stages. The parties first exchange information in nonbinding, informal negotiations. These negotiations are open-ended. They terminate only when no further voluntary communication is forthcoming. Second, they sign the actual contract that will govern the structure and conduct of their relationship. At the signing stage, no further information is exchanged. Indeed to introduce new information or further consideration at that time would be to extend the negotiation phase. The formal signing of the amended contract, representing the binding agreement, would be done subsequently and voluntarily. This paper is

¹ This research was supported by Commissariat au Plan, Paris and National Science Foundation Grants IST-8310118 and SES-8207338 to Harvard University. The authors would like to thank Dilip Abreu for helpful comments.

an attempt to ascertain which decision rules can emerge from such a two-step process.

Having described our objectives in general terms, let us now present the model by means of a concrete application to the case of two individuals who are deciding whether or not to undertake a joint venture. Before making this decision they will each gather information that can be used to evaluate the success of the venture, if it were undertaken. They will then meet and pool their information, but only to the extent that each of them is willing to reveal what he has learned. Finally, they will decide whether or not actually to undertake the venture. At this last stage they will each have whatever information has been revealed by the other player in the previous stage, in addition to their own observations. The outcome of this process can be described by the induced decision rule. It gives the probability that the venture will be undertaken, as a function of the observed private information. We provide a criterion for the implementability of a decision rule which we call *posterior implementability*. It is designed to capture the idea that binding agreements cannot be made until all information has been exchanged.

Let us now try to contrast the concept of posterior implementability with the rest of the theory of incentive compatibility, from which it is quite distinct. There are two grounds for comparison. The first concerns the information available to the players and to the mechanism designer. The second concerns the ability or inability of players to make binding commitments during the play of the game associated with the mechanism.

Any mechanism must specify the strategies available to each agent and the outcome that will result if each possible combination of strategies is played. The way in which this game is played depends upon the information available to the participants. Incentive theory quite properly recognizes that this information may be different than that available to the mechanism designer at the time at which commitment to use the mechanism is made. If the agents know only their own private information, the appropriate concept is the Bayesian equilibrium of the game. If they actually know all about the other players in the game, then the Nash equilibrium is more natural. In either case, however, two crucial assumptions are made. First the information is exogenously given to the agents before they must play. And second, a binding commitment to abide by the result of the mechanism is possible.

In our model both of these presumptions are dropped. We assume that each agent acquires all additional information that is endogenously transmitted in the course of playing the mechanism. And precisely because no binding commitments can be made, they are allowed to modify their play after having received this endogenous information. Only when it is common knowledge that all agents are going to remain satisfied with their play, given the play of all others, do we imagine that the outcome specified by the mechanism is actually implemented.

Thus, our concept of implementability lies somewhere between Bayesian and Nash implementability in its informational basis. We emphasize, however, that in its assumption of the inability of agents to commit themselves to their play, it departs significantly from the mainstream of this literature.

Before turning to a formal treatment of our model, we offer some further remarks about the economic environment as it pertains to the structure of preferences and information. In most of the models for which Nash implementation has been used as the solution concept, the information of each agent pertains only to his own preferences over the set of actions. It is irrelevant to others. Therefore agents would need to know this information only because it determines the strategy that other agents will play. For any given strategic choice others' information does not affect their preferences. In environments with this informational structure the definitions of Nash implementability and posterior implementability coincide.

In our model, and in other models where mutually payoff-relevant information is present, Nash implementable decision rules will be a proper subset of posterior implementable rules. In a Nash implementation, all private information is revealed, whereas in a posterior implementation the information to be revealed only need be fine enough to render the decision rule measurable—that is, fine enough so that a unique action can actually be determined. The additional information available in the Nash case may destroy the optimality of the strategies played in the posterior implementation.

The revelation principle, stating that Bayesian implementable rules can be implemented via truth-telling strategies, is not valid for the concepts of posterior and Nash implementability. Nash implementability, however, can be regarded as a special case of Bayesian implementability, for which the revelation principle is valid, if we endow everyone with the information available to all. Viewed in this way, it is quite natural to construct message spaces for each agent which consist simply of all the possible observations available to all agents in the system. Because of the endogeneity of the information the revelation principle cannot similarly be applied to posterior implementability. As information is initially private and is only exchanged in the process of playing the mechanism, it turns out that it is often important to keep the strategy spaces fairly small so as not to adversely influence incentives. Indeed, the strategy spaces in our model turn out to be far smaller than each individual agent's private information, rather than far larger as would be generally the case with Nash.

The requirement of posterior implementability places strong restrictions on the decision rules. In this paper we examine a two-person decision problem with two possible social choices. For this case we are able to give a complete characterization of the posterior implementable decision rules. Smooth variation of the social decision, or smoothly changing randomizations over the two possible decisions, cannot be achieved. As the private information varies, the social decision can take only two possible values. Moreover, the sets of pairs of private observations over which each value of the social decision is realized must satisfy some stringent restrictions. The set of posterior implementable decision rules is much smaller than the set of Bayesian implementable rules.

In Section 1 we lay out the model and our assumptions. Section 2 is devoted to the proof of the characterization of posterior implementable decision rules. Finally, Section 3 characterizes the qualitative nature of posterior implementable

rules and offers various comments on the construction of mechanisms implementing these rules. The relationship of this research with the literature and possible extensions is also discussed.

1. THE CONCEPT OF POSTERIOR IMPLEMENTATION

We consider two individuals called players who must choose between two decisions denoted d_0 and d_1 ; d_0 is the status quo decision and its utility is set at zero for both players. Each player receives a private real valued observation that is relevant to his utility of d_1 . Let $\theta_i \in \Theta_i$ be the observation received by player i , $i = 1, 2$, and let $v_i(\theta_1, \theta_2)$ be the expected utility of d_1 for player i given θ_1 and θ_2 . The joint distribution of (θ_1, θ_2) over $\Theta_1 \times \Theta_2$, which is common knowledge, is given by the density $f(\theta_1, \theta_2)$.²

Throughout the paper we make the following assumptions:

ASSUMPTION A1: Θ_i is a closed interval of \mathbb{R} , $i = 1, 2$: $\Theta_i = [\theta_{i \min}, \theta_{i \max}]$.

ASSUMPTION A2: v_i is continuous and strictly increasing in both arguments, $i = 1, 2$. There are at most a finite number of pairs (θ_1, θ_2) for which $v_1(\theta_1, \theta_2) = v_2(\theta_1, \theta_2) = 0$.

ASSUMPTION A3: f is continuous and strictly positive on $\Theta_1 \times \Theta_2$. The conditional densities $f_1(\cdot | \theta_1)$ and $f_2(\cdot | \theta_2)$ are strictly positive on their domains, Θ_2 and Θ_1 , respectively. Moreover, for any subinterval $\hat{\Theta}_i$ of Θ_i , the conditional distributions $F_j(\cdot | \theta_j, \theta_i \in \hat{\Theta}_i)$ are monotonic increasing in θ_j in the partial ordering of first order stochastic dominance, for $i = 1, 2, j \neq i$.

Assumption A1 is a technical assumption. It avoids certain problems related to disconnected domains of private information, analogous to those discussed by Holmstrom (1979) in the context of dominant strategy implementation.

Assumption A2 represents the idea that higher values of θ_1 and θ_2 are, for both players, "good news" with respect to the value of d_1 .

Assumption A3 asserts a monotone likelihood ratio property of f which is crucial for the analysis. With A2 it gives the monotonicity of the integrals, $\int_{\hat{\Theta}_j} v_i(\theta_1, \theta_2) f_i(\theta_j | \theta_i) d\theta_j$ for $i = 1, 2$ and $j \neq i$, and any subintervals $\hat{\Theta}_1, \hat{\Theta}_2$.

An alternative derivation of the same structure would be to assume that the true ultimate values of d_1 to the two players, V_1 and V_2 , are jointly distributed with θ_1 and θ_2 in such a way that the vector-valued random variable $(V_1, V_2, \theta_1, \theta_2)$ is affiliated (see Milgrom (1981)). The fact that the support of the conditional distribution $F_1(\cdot | \theta_1)$ (resp. $F_2(\cdot | \theta_2)$) is independent of θ_1 (resp. θ_2) simplifies the analysis but could easily be generalized.

Incentive theory has introduced the notion of a mechanism to describe the strategic interaction through which a decision is reached.

²The distribution functions associated with density functions (f) are identified by capital letters (F).

A mechanism is a pair of measurable message spaces M_1, M_2 and a measurable outcome function g from $M_1 \times M_2$ into $[0, 1]$.

The interpretation is that the players choose messages, $m_1 \in M_1$ and $m_2 \in M_2$, respectively; then the decision d_1 is chosen with probability $g(m_1, m_2)$.

Most of the literature on incentive compatible group decision making has focused on the case in which the players choose $m_i \in M_i$ based on their a priori beliefs and their private information $\theta_i \in \Theta_i$. Allowing for the possibility of randomization, a strategy of agent i is a measurable function from Θ_i into the family of distributions over M_i . This can be expressed by the conditional distributions $s_i(m_i | \theta_i)$, for $\theta_i \in \Theta_i$. If the strategies s_1 and s_2 are used in the mechanism (M_1, M_2, g) the resulting social choice can be described by the function

$$\phi : \Theta_1 \times \Theta_2 \rightarrow [0, 1]$$

defined by

$$\phi(\theta_1, \theta_2) = \int_{M_1 \times M_2} g(m_1, m_2) ds_1(m_1 | \theta_1) ds_2(m_2 | \theta_2).$$

The value of ϕ is the probability that the decision d_1 will be taken when (θ_1, θ_2) is the information received by the players.

Given s_2 , a message $m_1 \in M_1$ is an optimal message for player 1 if it maximizes

$$\int_{\Theta_2} \int_{M_2} v_1(\theta_1, \theta_2) g(m_1, m_2) ds_2(m_2 | \theta_2) dF_1(\theta_2 | \theta_1).$$

A strategy for player 1 is *optimal* if for almost every $\theta_1 \in \Theta_1$, $s_1(\cdot | \theta_1)$ assigns zero probability to the set of nonoptimal messages.

If, for the given strategy s_1 , the strategy s_2 is an optimal strategy for player 2 and conversely, then we say that (s_1, s_2) constitute a *Bayesian equilibrium* of the mechanism and that ϕ is *Bayesian incentive compatible*. The adjective Bayesian refers to the fact that in verifying the optimality of his choice m_i given θ_i , player i uses his posterior distribution $F_i(\cdot | \theta_i)$ as well as his knowledge of $s_j(\cdot | \cdot)$ to compute the joint distribution of (m_j, θ_j) that is relevant to his action.

When there exists a mechanism (M_1, M_2, g) and a Bayesian equilibrium (s_1, s_2) of this mechanism resulting in $\phi : \Theta_1 \times \Theta_2 \rightarrow [0, 1]$, we say that ϕ is *implementable*. The set of implementable functions is one of the basic objects studied in the theory of incentives. It is generally thought to represent the feasible set of alternatives for this pair of individuals, and any optimization or normative discussion of mechanism design is then carried out within this set.

In this paper we add a further requirement beyond the implementability of ϕ as defined above. We will now describe the idea before offering the formal definition. Suppose that ϕ has been implemented by (M_1, M_2, g) and a Bayesian equilibrium (s_1, s_2) . As each player i knows the strategies being played, he can compute the posterior distribution of the other's observation θ_j given his play $m_j \in M_j$. At this point it is often in neither player's interest to play the m_i that he has selected. Assuming that the other player will persist in his decision, each would like to revise his own. The main idea of this paper is to impose as a further

condition that at every situation to which the Bayesian equilibrium of the mechanism could arrive, both players are satisfied with their choices.

Formally, let ϕ be implemented via (M_1, M_2, g) in the Bayesian equilibrium (s_1, s_2) . Let $\mu(m_1, m_2, \theta_1, \theta_2)$ be the joint distribution over $M_1 \times M_2 \times \Theta_1 \times \Theta_2$ generated in this equilibrium. Let $\mu_i(\theta_j)$ be the marginal distribution of m_i given θ_j for $j \neq i$. Define $F_2(\theta_1 | m_1, \theta_2)$ and $F_1(\theta_2 | m_2, \theta_1)$ to be the conditional distributions that the two players would hold about each other's observation after having observed the other's choice m_i and their own observation θ_j . They are defined for every θ_j and $\mu_i(\theta_j)$ -almost every m_i . We will say that (s_1, s_2) is a *posterior optimal* pair of strategies if, for μ -almost every $(m_1, m_2, \theta_1, \theta_2)$,

$$m_1 \text{ maximizes } \int_{\Theta_2} v_1(\theta_1, \theta_2) g(m'_1, m_2) dF_1(\theta_2 | m_2, \theta_1)$$

over $m'_1 \in M_1$, and

$$m_2 \text{ maximizes } \int_{\Theta_1} v_2(\theta_1, \theta_2) g(m_1, m'_2) dF_2(\theta_1 | m_1, \theta_2)$$

over $m'_2 \in M_2$.

A function ϕ is said to be *posterior implementable* if it is implementable via (M_1, M_2, g) with posterior optimal strategies (s_1, s_2) .

A basic result of the theory of incentives, known as the *revelation principle*, states that if ϕ is implementable then ϕ can be implemented by $M_1 = \Theta_1$, $M_2 = \Theta_2$, and $g = \phi$, with the strategies (s_1, s_2) that are each degenerate distributions concentrated on the truth. When ϕ is implemented in this way, both players will know, *ex post*, the exact observation received by the other. Therefore, ϕ is posterior implemented through such a truthful implementation if for all (θ_1, θ_2) , the truthful statements θ_1, θ_2 are a Nash equilibrium of the game with payoff functions $v_1(\theta_1, \theta_2)g(\tilde{\theta}_1, \tilde{\theta}_2)$ and $v_2(\theta_1, \theta_2)g(\tilde{\theta}_1, \tilde{\theta}_2)$ where $\tilde{\theta}_1, \tilde{\theta}_2$ are the strategies. This means that the truthful strategies are dominant strategies in this game. And this implies that, unless g is a constant, each v_i can have only one sign, throughout Θ_i . In any interesting problem this will generally not be the case. We can conclude from this discussion that truthful strategies reveal too much information for them to be useful in a posterior implementation.

In designing a mechanism so as to satisfy the stronger property of posterior implementability, one faces the following type of tradeoff. The smaller the message spaces, the less information will be conveyed in equilibrium and the easier it will be to achieve posterior optimality. But at the same time, smaller message spaces allow for less flexibility in varying g to accommodate the players' mutual interests.

In the next section we characterize the posterior-implementable choice functions.

2. CHARACTERIZATION OF POSTERIOR-IMPLEMENTABLE CHOICE RULES

In this section we prove, through a sequence of lemmas, that posterior-implementable rules must have a very special form. There can be at most two

possible outcomes of the social decision process as (θ_1, θ_2) varies. The regions in which these two outcomes obtain are formed by a partitioning of $\Theta_1 \times \Theta_2$ by a step function which must satisfy certain restrictions. The strength of this result derives from the fact that implementable rules, in general, can be much more complex. There may be many outcomes and they may depend in a more elaborate way on the private information of the agents.

The sequence of steps used to prove our characterization theorem is as follows. We show (Lemma 1) that it is never necessary to set up a mechanism in which there are some messages that are never sent or in which there are any redundant messages, that is distinct messages which lead to identical outcomes. Then we show (Lemma 2) that the mapping from individual's private observations to the messages they send is of a very special form. The interval Θ_i is simply partitioned into a collection of subintervals, each of which is mapped nonstochastically into a single message. Using this result we show (Lemma 3) that holding other players' messages fixed, the outcome function can take on at most three values as any one player's message varies. The locus of discontinuity points in the outcome function is then characterized in a sequence of steps (Lemmas 4, 5, 6, and 7) and is shown to be a step function solving a certain system of integral equations. From this result, the next step (Lemma 8) reduces from three to two the maximum number of outcomes that a player can reach by varying his own message. Putting these results together, we arrive at the Theorem sketched above.

We first introduce the concept of minimal implementation. The choice rule ϕ is said to be *minimally posterior implemented* via (M_1, M_2, g) if it is posterior implemented via (M_1, M_2, g) and if (i) there do not exist distinct m_1^1, m_1^2 in M_1 such that $g(m_1^1, m_2) = g(m_1^2, m_2)$ for all $m_2 \in M_2$, (ii) there do not exist distinct m_2^1, m_2^2 in M_2 such that $g(m_1, m_2^1) = g(m_1, m_2^2)$ for all $m_1 \in M_1$.

In a minimal implementation there is no redundant message. To characterize implementable choice rules there is no loss of generality in restricting the analysis to minimal posterior implementation as shown by the following lemma.

LEMMA 1: *If ϕ is posterior implementable, then it is minimally posterior implementable.*

PROOF: Let ϕ be posterior implemented via (M_1, M_2, g) and the strategies (s_1, s_2) . Define an equivalence relation on M_1 by identifying points m_1^1, m_1^2 such that $g(m_1^1, m_2) = g(m_1^2, m_2)$ for all m_2 . Let M_1^* be the set of such equivalence classes, and likewise for M_2^* . Let $g^*(m_1^*, m_2^*) = g(m_1, m_2)$ for any $m_1 \in m_1^*, m_2 \in m_2^*$. The measurability of (M_1^*, M_2^*, g^*) is inherited from (M_1, M_2, g) in a straightforward way.

Properties (i) and (ii) in the definition of minimal implementability are satisfied by construction.

Define the strategies $s_i^*(\cdot | \theta_i)$ on M_i^* by

$$s_i^*(A_i^* | \theta_i) = s_i(A_i(A_i^*) | \theta_i)$$

for every measurable $A_i^* \subseteq M_i^*$, where $A_i(A_i^*) = \{m_i | m_i \in m_i^* \text{ for some } m_i^* \in A_i^*\}$.

It suffices to show that $s_i^*(\cdot|\theta_i)$ are posterior optimal under the mechanism (M_1^*, M_2^*, g^*) .

Let $F_2^*(\theta_1|m_1^*, \theta_2)$ be the beliefs of player 2 after the observation of θ_2 and m_1^* , defined for all possible observations of m_1^* given the strategy s_1^* . Let $F_2(\theta_1|m_1, \theta_2)$ be defined as in Section 1. Given θ_2 consider any optimal message m_2 and its equivalence class $m_2^* \in M_2^*$. By definition of m_1^* , the choice of m_2 is optimal when beliefs are $F_2(\theta_1|m_1, \theta_2)$ for any $m_1 \in m_1^*$. Since $F_2^*(\theta_1|m_1^*, \theta_2)$ is a convex combination of $F_2(\theta_1|m_1, \theta_2)$ for $m_1 \in m_1^*$, the choice of m_2^* remains posterior optimal in the mechanism (M_1^*, M_2^*, g^*) after the observation of θ_2 and m_1^* . Hence, s_2^* constitutes a posterior optimal strategy. Q.E.D.

By virtue of Lemma 1 we can confine attention to minimal posterior implementations. Let us define the correspondence $T_i: M_i \rightarrow \Theta_i$ by

$$T_i(m_i) = \{\theta_i | m_i \text{ is an optimal message given } \theta_i\}.$$

Given $s_i(\cdot|\theta_i)$, there is a joint distribution of $(m_i, \theta_i) \in M_i \times \Theta_i$. The set $T_i(m_i)$ is the support of the induced conditional distribution of θ_i given m_i .

We next derive some properties of the correspondences T_i .

LEMMA 2: *In any minimal posterior implementation of ϕ via (M_1, M_2, g) , the image sets $T_i(m_i)$ are convex and if $m_1^1 \neq m_1^2$ then $T_i(m_1^1) \cap T_i(m_1^2)$ contains at most one point.*

PROOF: Let

$$W(\theta_1, m_1, m_2) = \int_{\Theta_2} v_1(\theta_1, \theta_2) g(m_1, m_2) dF_1(\theta_2|m_2, \theta_1).$$

In choosing a best response m_1^* to m_2 given θ_1 , player 1 will want to maximize $g(m_1, m_2)$ when $\int_{\Theta_2} v_1(\theta_1, \theta_2) dF_1(\theta_2|m_2, \theta_1)$ is positive, minimize it if this integral is negative, and will be indifferent to the value of $g(m_1, m_2)$ if this integral is zero.

First we will show that $T_i(m_i)$ is convex. Suppose that $\theta_1^1 \leq \theta_1^2 \leq \theta_1^3$ and that $\theta_1^1, \theta_1^3 \in T_i(m_1^*)$ and $\theta_1^2 \notin T_i(m_1^*)$. Suppose $\theta_1^2 \in T_i(m_1^{**})$. Consider any \bar{m}_2 such that $g(m_1, \bar{m}_2)$ is not a constant with respect to m_1 . To ease the notation, let $I(\theta_1) \equiv \int_{\Theta_2} v_1(\theta_1, \theta_2) dF_1(\theta_2|\bar{m}_2, \theta_1)$. By positive affiliation and the monotonicity of $v_1(A2, A3)$, $I(\theta_1^1) \geq 0$ implies $I(\theta_1^2) > 0$, $I(\theta_1^3) > 0$; and $I(\theta_1^1) \leq 0$ implies $I(\theta_1^2) < 0$, $I(\theta_1^3) < 0$. Since $\theta_1^1, \theta_1^3 \in T_i(m_1^*)$, either (i) $I(\theta_1^1) \geq 0$, $I(\theta_1^3) > 0$ and m_1^* maximizes $g(m_1, \bar{m}_2)$ or (ii) $I(\theta_1^1) < 0$, $I(\theta_1^3) \leq 0$ and m_1^* minimizes $g(m_1, \bar{m}_2)$. As m_1^{**} is assumed to be an optimal message given θ_1^2 , in either case $g(m_1^{**}, \bar{m}_2) = g(m_1^*, \bar{m}_2)$, as they must both maximize $g(\cdot, \bar{m}_2)$ (case (i)) or minimize $g(\cdot, \bar{m}_2)$ (case (ii)) when θ_1^2 is observed.

If \bar{m}_2 is such that $g(m_1, \bar{m}_2)$ is a constant with respect to m_1 , then in particular, $g(m_1^{**}, \bar{m}_2) = g(m_1^*, \bar{m}_2)$. Therefore, for all $\bar{m}_2 \in M_2$, $g(m_1^{**}, \bar{m}_2) = g(m_1^*, \bar{m}_2)$, contradicting the minimality of the implementation. This proves that $T_i(m_1^*)$ is convex, that is, it is an interval.

Moreover, the above argument shows that if $\theta_1 \in \text{int } T_1(m_1^*)$, then $\theta_1 \notin T_1(m_1^*)$ for any $m_1^{**} \neq m_1^*$. To prove that if $m_1^1 \neq m_1^2$ then $T_1(m_1^1) \cap T_1(m_1^2)$ contains at most one point, we apply this observation. If they did intersect in a nondegenerate interval, we could take θ_1 in the interior of this interval, and as it would be in the interior of both $T_1(m_1^1)$ and $T_1(m_1^2)$, we would have a contradiction. Q.E.D.

Given any minimal implementation of ϕ , the last lemma tells us that the information conveyed to each player, *ex post*, consists of knowing that the θ_i of his opponent lies in some interval, possibly degenerate. The implication of this condition is that when posterior beliefs differ because of a difference in the opponent's play, the change in beliefs is an unambiguous shift, upward or downward, in the range of possible values of his θ_i . Therefore, the player's evaluation of his expected utility will shift monotonically with the opponent's play, by virtue of the fact that preferences are monotone and that the observations are affiliated.

The next lemma draws the implications of this monotonicity for the nature of the functions g that can be used in a minimal implementation. Let us define an ordering on M_i that orders the points of M_i in the same way as the intervals Θ_i to which each $m_i \in M_i$ is associated. This ordering will be called the *informational ordering of M_i in the minimal implementation of ϕ via (M_1, M_2, g)* . The equilibrium (s_1, s_2) , or simply the *informational ordering* when the context is otherwise clear.

LEMMA 3: *In any minimal posterior implementation of ϕ , for each $m_1 \in M_1$, $g(m_1, m_2)$ can take at most three values as m_2 varies over M_2 . Moreover, $g(m_1, \cdot)$ is monotonically nondecreasing in the informational ordering on M_2 .*

PROOF: Fix $m_1 \in M_1$. Let $\bar{\theta}_2$ be such that

$$0 = \int_{\theta_1} v_1(\theta_1, \bar{\theta}_2) dF_2(\theta_1 | m_1, \bar{\theta}_2).$$

Assume $g(m_1, m_2)$ takes at least two values as m_2 varies over M_2 . There must be $m_2 \in M_2$ such that $\bar{\theta}_2 \in \text{int } T_2(m_2)$. For if so, there would be some $\theta_2 \in T_2(m_2)$ for which $g(m_1, \cdot)$ should be minimized, and others for which it should be maximized, contradicting the fact that the same m_2 is optimal.

Therefore, for each m_2 , either

(i) $T_2(m_2) \subseteq [\theta_2, \infty)$

or

(ii) $T_2(m_2) \subseteq (-\infty, \bar{\theta}_2]$.

All m_2 such that (i) holds must give rise to the same value of g , i.e., $g(m_1, m_2) = \bar{\theta}_2$, in which case any value of g is acceptable.

Likewise, all m_2 such that (ii) holds must give rise to the same value of g , i.e., $g(m_1, m_2) = \bar{\theta}_2$. Therefore, there can be at most three values of $g(m_1, m_2)$.

at maximal value for m_2 such that $T_2(m_2) \cap (\bar{\theta}_2, \infty) \neq \phi$, a minimal value for m_2 such that $T_2(m_2) \cap (-\infty, \bar{\theta}_2) \neq \phi$, and a value between them which is possible only when $T_2(m_2) = \{\bar{\theta}_2\}$.

This argument has also demonstrated the stated monotonicity property. Q.E.D.

From the previous lemma we know that for each player i there is a set of intervals of Θ_i which overlap at most at their endpoints, describing the messages sent in a minimal implementation.

Define a function $t_1: M_2 \rightarrow \Theta_1$ as follows: If there exists θ_1 such that

$$\int_{\Theta_2} v_1(\theta_1, \theta_2) dF_1(\theta_2 | m_2, \theta_1) = 0$$

set $t_1(m_2) = \theta_1$. If this integral is negative for all θ_1 , set $t_1(m_2) = \theta_{1\max}$, and if positive for all θ_1 , set $t_1(m_2) = \theta_{1\min}$. The function t_1 is strictly monotonically decreasing in the informational ordering on M_2 . We can use t_1 to define a correspondence from Θ_2 to Θ_1 that will be very useful in the analysis of posterior-implementable decision rules.

Let $Z_1: \Theta_2 \rightarrow \Theta_1$ be defined by

$$Z_1(\theta_2) = \{\theta_1 | \theta_1 = t_1(m_2) \text{ for some } m_2 \in M_2 \text{ such that } \theta_2 \in T_2(m_2)\}.$$

The graph of Z_1 consists of intervals where m_2 corresponds to an interval of values of θ_2 and isolated points where m_2 corresponds to a unique value of θ_2 . The intervals may be open or closed at their endpoints, but this will be of no consequence in what follows.

Let \bar{Z}_1 be generated by the correspondence Z_1 by taking the closure of the graph of Z_1 and then taking the convex hull of the values of the resulting correspondence for each $\theta_2 \in \Theta_2$. Let ζ_1 be the graph of \bar{Z}_1 .

The graph ζ_1 consists of a set of "steps". For each m_2 , there is a piece of the graph of the form $\{t_1(m_2)\} \times T_2(m_2)$, which we call a "vertical" segment. These vertical segments are connected by "horizontal" segments which are generated by the convexification procedure described above. (See Figure 1.)

The same construction for player 2 results in a graph ζ_2 . In ζ_2 the nature of the horizontal and vertical segments is reversed: vertical segments are obtained by convexification and horizontal segments result from the cutoff values of θ_2 for each $m_1 \in M_1$.

Note that, in principle, some points of ζ_1 could be on both horizontal and vertical segments, as for example any "corner" or any point of the form $(t_1(m'_2), T_2(m'_2))$ where m'_2 is a message sent after the observation of a single value of θ_2 (see Figure 1). If all points θ_2 in a nondegenerate interval $I_2 \subseteq \Theta_2$ were of this type, then ζ_1 would be a smooth decreasing curve in this region because information would have been completely revealed. We rule out this possibility in the next lemma.

LEMMA 4: *In any minimal posterior implementation of ϕ , ζ_1 cannot be strictly decreasing over any nondegenerate interval.*

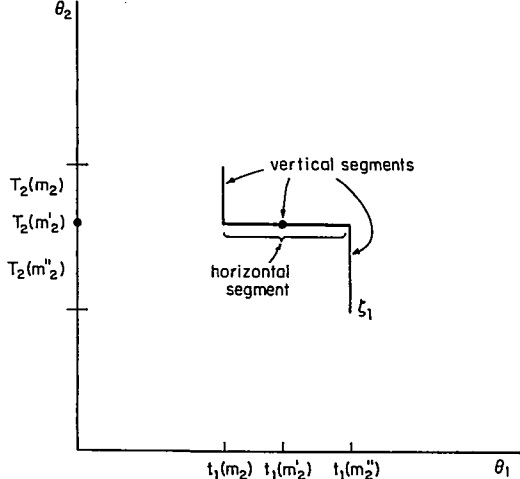


FIGURE 1

PROOF: Assume that ζ_1 is strictly decreasing over some nondegenerate interval I_2 . There exists a nondegenerate interval I_1^0 such that ζ_1 and ζ_2 are distinct within $I_1^0 \times I_2$. This is true because if they were to coincide over any rectangle, $I_1 \times I_2$, and thus both be strictly decreasing, both players would have perfect information about each other, and hence ζ_i would be the graph of $v_i = 0$, for both i , which would contradict the fact that $v_1 = v_2 = 0$ for only finitely many pairs (θ_1, θ_2) (A2). Whenever they fail to coincide over $I_1 \times I_2$ we can find a subinterval of I_1, I_1^0 , such that they are distinct on $I_1^0 \times I_2$.

Thus let $\theta_{1A}, \theta_{1B}, \theta_{2A}, \theta_{2B}$ be such that $\theta_{1A} < \theta_{1B}, \theta_{2A} < \theta_{2B}, \theta_{1A} < Z_1(\theta_{2B}) < \theta_{1B}, \theta_{2A} > Z_2(\theta_{1A})$, and $\theta_{2A} > Z_2(\theta_{1B})$, without loss of generality. (See Figure 2.)

Let $m_{1A}, m_{1B}, m_{2A}, m_{2B}$ be messages associated with these respective θ 's.

From Lemma 3,

$$g(m_{1A}, m_{2A}) = g(m_{1A}, m_{2B}) \quad \text{and} \quad g(m_{1B}, m_{2A}) = g(m_{1B}, m_{2B}).$$

Also,

$$g(m_{1A}, m_{2A}) = g(m_{1B}, m_{2A}).$$

Therefore,

$$g(m_{1A}, m_{2B}) = g(m_{2A}, m_{2B}).$$

Call their common value \bar{g} .

Also, if $m_1 < m_{1A}$ in the informational ordering,

$$g(m_1, m_{2A}) = \bar{g}, \quad g(m_1, m_{2B}) = \bar{g}$$

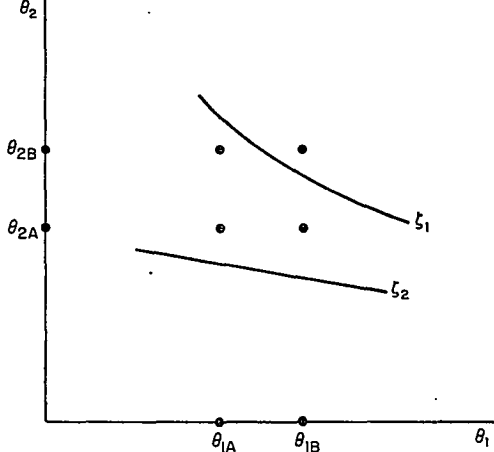


FIGURE 2

and likewise if $m_1 > m_{1B}$. Finally, by monotonicity, if $m_{1A} < m_1 < m_{1B}$,

$$g(m_1, m_{2A}) = g(m_1, m_{2B}) = \bar{g}.$$

Therefore, m_{2A} and m_{2B} are redundant messages, and hence ϕ is not minimal. Q.E.D.

Let $\{\bar{\theta}_1\} \times I_2$ be a vertical segment of ζ_1 , where I_2 is an interval in Θ_2 . This segment will be called terminal if either (see Figure 3) $(\theta_1, \theta_2) \in \zeta_1$ implies $\theta_1 \geq \bar{\theta}_1$ or $(\theta_1, \theta_2) \in \zeta_1$ implies $\theta_1 \leq \bar{\theta}_1$.

LEMMA 5: *Let ϕ be a nonconstant decision rule that is minimal posterior implementable. Let (θ_1, θ_2) be a point in ζ_1 which is not in ζ_2 . Then (θ_1, θ_2) is in a terminal segment of ζ_1 .*

PROOF: Assume that it is in a segment of ζ_1 that is not terminal. We can find two nonintersecting horizontal segments, one in ζ_1 and one in ζ_2 , that have in common a nondegenerate interval I_1 of θ_1 values, and we can find two nonintersecting vertical segments, one in ζ_1 and one in ζ_2 , having a common nondegenerate interval I_2 of θ_2 values (see Figure 4). Let $R = I_1 \times I_2$. By the minimality of the implementation, and by the measurability of ϕ with respect to $M_1 \times M_2$, $\phi(\theta_1, \theta_2)$ is constant over $\text{int } R$. Take $(\theta_1^*, \theta_2^*) \in \text{int } R$.

By monotonicity of g from Lemma 3 applied to ζ_1 , $\phi(\theta_1, \theta_2^*) = \phi(\theta_1^*, \theta_2^*)$ for all $\theta_1 \leq \theta_1^*$. And by the monotonicity of ζ_2 and this same lemma, $\phi(\theta_{1\min}, \theta_{2\min}) = \phi(\theta_1^*, \theta_2^*)$.

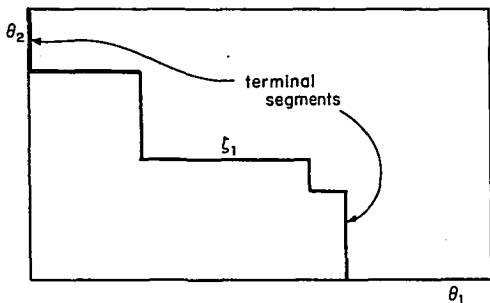
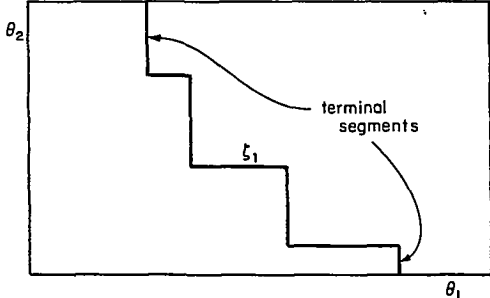


FIGURE 3

Similarly, $\phi(\theta_1^*, \theta_2) = \phi(\theta_1^*, \theta_2^*)$ for all $\theta_2 \geq \theta_2^*$. And $\phi(\theta_{1\max}, \theta_{2\max}) = \phi(\theta_1^*, \theta_2^*)$.

Thus ϕ is a constant.

Q.E.D.

Consider the common part of the loci ζ_1 and ζ_2 and their upper left and lower right endpoints a and b . It follows from monotonicity (Lemma 3) that the value of ϕ at any point which is dominated by a point on this locus (in the vector ordering) is equal to that at any other such point. Let $\phi(\theta_1, \theta_2) = g_-$ on the set of dominated points. Likewise let $\phi(\theta_1, \theta_2) = g_+$ whenever (θ_1, θ_2) dominates some point on this locus.

Thus ϕ is known to have exactly two values, except perhaps on this locus and in the two "corners" northwest of a and southeast of b . We will now show that the value in these corners can be determined uniquely. It will be either g_+ or g_- .

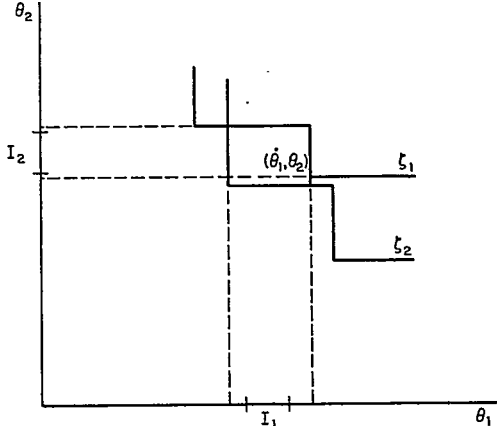


FIGURE 4

according to the shape of this locus near a and b . In each corner we have one of the three possible cases described in Figure 5.

LEMMA 6: *Case 3 on Figure 5 is impossible.*

PROOF: Since a' and b' are the endpoints of terminal segments, ζ_1 and ζ_2 must coincide in a neighborhood of a' when θ_1 is above a'_1 and θ_2 is below a'_2 . We first show that there must be an infinite number of segments in ζ_1 and ζ_2 in any neighborhood of a' . If there were a finite number of them, then because $a' \neq a$ and $b' \neq b$, we could find a vertical segment of ζ_2 with b' as an endpoint and a horizontal segment of ζ_1 with a' at an endpoint. This would contradict the fact that ζ_1 and ζ_2 coincide in this region. Therefore, there is an accumulation point of horizontal segments and of vertical segments converging to a' .

We know that the locus of $v_1 = 0$ cuts every nondegenerate vertical segment in its interior, by definition of $t_1(m_2)$. Therefore, $v_1 = 0$ holds at a' , by continuity. But $v_1 = 0$ is a strictly monotonic decreasing locus. Hence, it cannot cut the segment aa' in its interior; hence, $a' = a$. Q.E.D.

LEMMA 7: *In the corner defined by case 1, $\phi = g_-$ and in the corner defined by case 2, $\phi = g_+$.*

PROOF: *Case 1.* Suppose $\phi = \tilde{g} > g_-$ which is the value taken by ϕ below bb' from above. Then by measurability, agent 2 would need two different messages, one above b and one below b . But, then a, b', a' could not be on the same vertical line. Therefore, by monotonicity $\tilde{g} = g_-$. *Case 2* is treated similarly. Q.E.D.

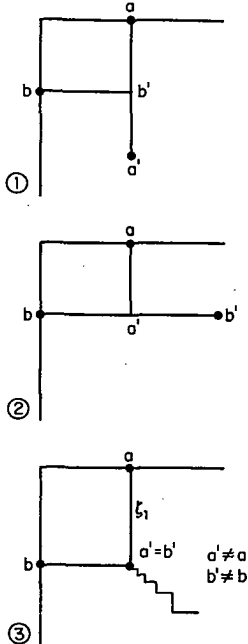


FIGURE 5

To complete the characterization of implementable choice rules, we examine the behavior of g along $\zeta_1 = \zeta_2$.

LEMMA 8: *Let ϕ be minimally implemented by (M_1, M_2, g) . Suppose that θ_1 is not an accumulation point of intervals $T_1(m_1)$ for $m_1 \in M_1$. Let $\theta_1 = T(\bar{m}_1)$. Then $g(\bar{m}_1, m_2)$ takes at most two values as m_2 varies in M_2 .*

PROOF: Consider first a message $m_1 \in M_1$ such that $T_1(m_1)$ is a nondegenerate interval and to which are associated three values of g as m_2 varies. To this message corresponds a cutoff point $\bar{\theta}_2$ for agent 2 and a horizontal segment AB of $\zeta_1 = \zeta_2$. By measurability there is a specific message \bar{m}_2 associated with $\bar{\theta}_2$; thus the associated cutoff point $t_1(\bar{m}_2)$ for agent 1 is in the interior of AB .

The third value of g can only occur on AB , but since g is not constant, g must take two different values on AB , a contradiction.

Consider now a message $m_1 \in M_1$ such that $\theta_1 = T_1(m_1)$ which belongs to $T_1(m_1^1)$ and $T_1(m_1^2)$ and to which are associated three values of g as m_2 varies. Then the same argument as above applies to a vertical segment if θ_1 is not $\theta_{1\min}$ or $\theta_{1\max}$.

Take now the case where $\theta_1 = \theta_{1\min}$. Immediately to the right of θ_1 there is a message m_1' . Consider Figure 6.

Let $\bar{\theta}_2 = t_2(m_1)$ be the cutoff point for m_1 and $\underline{\theta}_2 = t_2(m_1')$ be the cutoff point for m_1' and suppose that we have these values for $g(m_1, m_2)$ as m_2 varies over M_2 .

To $\bar{\theta}_2$ is associated a message m_2 and below $\bar{\theta}_2$ there is another message m_2' . But since $g = g_-$ on the interval $(\bar{\theta}_2, \underline{\theta}_2)$, it implies that θ_{\min} is the cutoff point for m_2' . But then the cutoff point for m_2 should be strictly to the left of $\theta_{1\min}$ which is impossible. Q.E.D.

In an accumulation point of nondegenerate intervals (θ_1, θ_2) , θ_1 and θ_2 can give rise to specific messages m_1 and m_2 , respectively, and g can take any value between g_- and g_+ at (m_1, m_2) .

Since v_1 must cut each vertical segment of $\zeta_1 = \zeta_2$ and v_2 must cut each horizontal segment of $\zeta_1 = \zeta_2$, an accumulation point can only arise at a point of intersection of v_1 and v_2 (see Figure 7).

We can now describe several features that posterior implementable ϕ -functions necessarily have.

Almost everywhere ϕ is nonstochastic and almost everywhere takes only one of two values, namely, g_+ if (θ_1, θ_2) is above $\zeta_1 = \zeta_2$ and g_- if (θ_1, θ_2) is below $\zeta_1 = \zeta_2$.

The behavior of ϕ on $\zeta_1 = \zeta_2$ is more complex. Throughout the interior of a given horizontal (or vertical) segment of $\zeta_1 = \zeta_2$, ϕ may be strictly between g_- and g_+ because $s_2(\cdot | \theta_2)$ (or $s_1(\cdot | \theta_1)$) is a random variable taking values on two distinct messages, leading to g_- and g_+ with positive probability. These random variables may differ on the different segments.

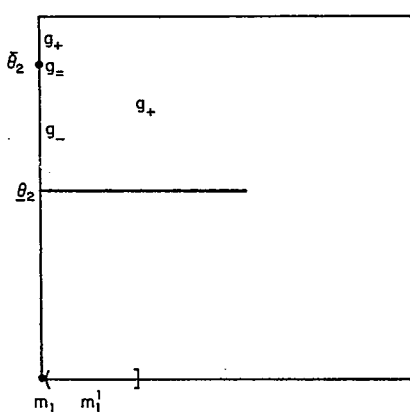


FIGURE 6

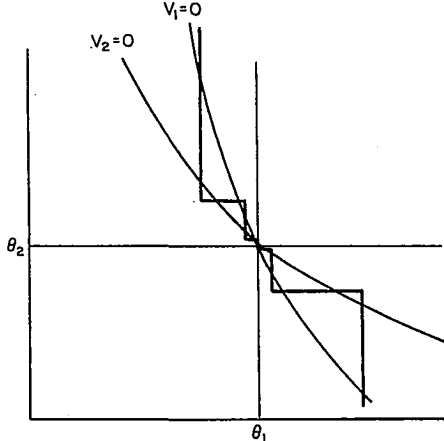


FIGURE 7

At the corners of $\zeta_1 = \zeta_2$ the value of ϕ is determined by its values on the two segments that intersect at that point because the randomizations inherent in s_1 and s_2 are performed independently.

Since the expected utility of agents is insensitive to the values of ϕ over the set of measure zero $\zeta_1 = \zeta_2$, we can ignore the values of ϕ on this locus. The characterization of implementable ϕ -functions takes then a simple form.

Let \mathcal{X} be the set of decreasing step functions which partition $\Theta_1 \times \Theta_2$ in two subsets and which have the property that any vertical segment (a, b) is such that

$$\int_{\Theta_1} v_1(\theta_1, \theta_2) dF_1(\theta_2 | \theta_1, \theta_2 \in (a, b)) = 0$$

and any horizontal segment (c, d) is such that

$$\int_{\Theta_2} v_2(\theta_1, \theta_2) dF_2(\theta_1 | \theta_2, \theta_1 \in (c, d)) = 0.$$

THEOREM: Any posterior implementable ϕ -function is such that there exists a step function ξ in \mathcal{X} and $\phi(\theta_1, \theta_2) = \phi_+$ if (θ_1, θ_2) is above ξ , and $\phi(\theta_1, \theta_2) = \phi_-$ if (θ_1, θ_2) is below ξ , where $\phi_+ \geq \phi_-$.

A major feature of this result is that ϕ is monotonically increasing and can take only two values. The dividing line between the two regions defined by the values of ϕ is a decreasing step function with a special property described by the family \mathcal{X} .

An example of posterior implementable mechanism is given in Figure 8.

The rule implemented by the mechanism in Figure 8 is one in which player 1 has three messages and player 2 has two. If player 1 sends m_{11} , $\phi = g_-$ with

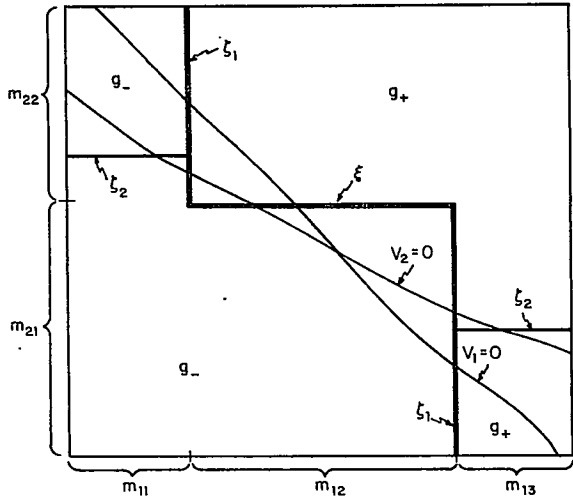


FIGURE 8

certainty; likewise, if m_{13} is sent $\phi = g_+$. In the middle range of θ_1 , where m_{12} is the message, the outcome depends on m_2 . Note that since

$$\int_{\theta_1} v_2(\theta_1, \theta_2) dF_2(\theta_2 | \theta_1, m_{12}) = 0$$

on the horizontal segment of ξ , player 2 will want to choose his message as indicated after m_{12} . Since player 1's other messages determine the outcome, player 2 is indifferent in those cases.

Similarly,

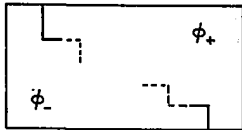
$$\int_{\theta_2} v_1(\theta_1, \theta_2) dF_1(\theta_2 | \theta_1, m_{2k}) = 0$$

for $k=1$ and 2 , so that player 1 is choosing optimally given each of player 2's messages.

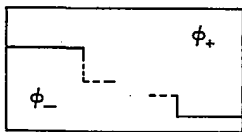
3. COMMENTS

A. The Four Forms of Implementable Rules

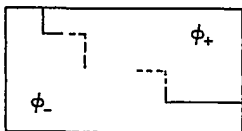
The step function ξ referred to in the last theorem, which separates the region of ϕ_+ from ϕ_- , connects either the upper ($\theta_2 = \theta_{2\max}$) or right ($\theta_1 = \theta_{1\max}$) boundary of Θ to the lower ($\theta_2 = \theta_{2\min}$) or left ($\theta_1 = \theta_{1\min}$) boundary. Qualitatively there are four cases as shown in Figure 9.



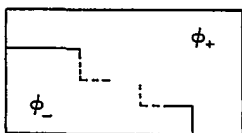
case 1



case 2



case 3



case 4

FIGURE 9

In the first two, one of the two players can dictate either g_+ or g_- —his lowest message mandating the latter, and his highest mandating the former, for all possible messages of his opponent. Between these messages there is cooperation in the sense that both players can influence the outcome. In the third case, either player can mandate g_- by playing his lowest message, while in the fourth case either can force g_+ .

Note that in the first case the number of messages available to player 1 is one greater than that for player 2, in case 2 it is one less, and in cases 3 and 4 their message spaces are of equal cardinality. This characterization of the four cases will be used below to provide a fixed point method for finding implementable rules.

B. The Construction of Posterior Implementable Decision Rules Using Fixed Point Methods

The results above characterize the posterior implementable decision rules, and four qualitatively different forms of these rules have been displayed in Figure 9.

It is of course important to find posterior optimal rules by a mathematically constructive procedure. That is the goal of this subsection.

For a given type of rule we might want to know, for example, whether there are posterior implementable rules with some specified number of messages being used. Intuitively, the more distinct messages are used, the more steps there will be in the associated curve Γ separating the region where $\phi = \phi_+$ from $\phi = \phi_-$, and the more likely it becomes that the decision will be ex post efficient. (This is merely meant to be suggestive; any result along these lines would have to depend on assumptions about the v_i much stronger than monotonicity.)

Below we will show how all posterior implementable rules can be found by examining the fixed points of a collection of functions.

Let $(\underline{\theta}_1, \bar{\theta}_1)$ be an arbitrary interval in Θ_1 . Define the function h_2 mapping intervals in Θ_1 into Θ_2 as follows. Consider the expression

$$(3.1) \quad \int_{\underline{\theta}_1}^{\bar{\theta}_1} v_2(\theta_1, \theta_2) dF(\theta_1 | \theta_2).$$

If there exists $\theta_2 \in \Theta_2$ such that (3.1) is zero, set

$$h_2(\underline{\theta}_1, \bar{\theta}_1) = \theta_2.$$

If (3.1) is positive for all θ_2 , set

$$h_2(\underline{\theta}_1, \bar{\theta}_1) = \theta_{2\min}$$

and if (3.1) is negative for all θ_2 , set

$$h_2(\underline{\theta}_1, \bar{\theta}_1) = \theta_{2\max}.$$

The monotonicity arguments used above in Section 2 show that (3.1) has at most one zero in Θ_2 .

Define $h_1(\underline{\theta}_2, \bar{\theta}_2)$ in an analogous fashion.

Let $\theta_i^1 < \dots < \theta_i^{K-1}$ be points in Θ_i and let $\theta_i^0 = \theta_{i\min}$, $\theta_i^K = \theta_{i\max}$ for $i = 1, 2$.

Define the function H mapping vectors $(\theta_1^1, \dots, \theta_1^{K-1}, \theta_2^1, \dots, \theta_2^{K-1})$ into vectors $(t_1^1, \dots, t_1^K, t_2^1, \dots, t_2^K)$ by

$$t_i^k = h_i(\theta_j^{k-1}, \theta_j^k) \quad \text{for } i = 1, 2; k = 1, \dots, K-1; j \neq i.$$

Clearly H has all the properties necessary for the application of Brouwer's fixed-point theorem. Moreover, $t_i^k \leq t_i^{k+1}$ for all k , by virtue of monotonicity considerations.

If $(\theta_1^{*1}, \dots, \theta_1^{*K-1}, \theta_2^{*1}, \dots, \theta_2^{*K-1})$ is a fixed point of H in which all adjacent points $\theta_i^{*k}, \theta_i^{*(k+1)}$ are distinct, then there is a posterior implementable decision rule with each player sending K distinct messages, each m_{ik} corresponding to an interval $(\theta_i^{*(k-1)}, \theta_i^{*k})$ in Θ_i . The associated g function is given by

$$(3.2) \quad g(m_{1k_1}, m_{2k_2}) = g_+ \quad \text{if } k_1 + k_2 \geq K + 1 \\ = g_- \quad \text{if } k_1 + k_2 < K + 1.$$

The meaning of this fixed point is very intuitive and can be seen in Figure 10 for the case of $K = 2$. If $\theta_{1\min} < \theta_1 < \theta_1^{*1}$, player 1 sends the message m_{11} . Player 2's response to this will be to maximize or minimize the value of $g(m_{11}, \cdot)$ according to the sign of (3.1) where $\underline{\theta}_1 = \theta_{1\min}$, $\bar{\theta}_1 = \theta_1^{*1}$. The cutoff value of θ_2 for this decision is therefore θ_2^{*1} . Note that if $\theta_1 > \theta_1^{*1}$, so that player 1 sends m_{12} , player 2 would want to use the cutoff $h_2(\theta_1^{*1}, \theta_{1\max})$. In particular, he would want to minimize $g(m_{12}, \cdot)$ in this circumstance, whenever $\theta_2 < h_2(\theta_1^{*1}, \theta_{1\max})$. However, as $g(m_{12}, m_2) = g_+$ for all m_2 (by (3.2)), the choice of m_2 is irrelevant to him, and therefore the strategy defined by the cutoff $\theta_2^{*1} = h_2(\theta_{1\min}, \theta_1^{*1})$ is optimal.

Note that this construction always gives a rule of the form of Case 4 of Figure 9. To attain rules as in Case 3, for example, one could define H by

$$t_i^k = h_i(\theta_j^{k-1}, \theta_j^k) \quad \text{for } i = 1, 2; k = 2, \dots, K; j \neq i;$$

and the corresponding mechanism by

$$\begin{aligned} g(m_{1k_1}, m_{1k_2}) &= g_+ & \text{if } k_1 + k_2 > K + 1, \\ &= g_- & \text{if } k_1 + k_2 \leq K + 1. \end{aligned}$$

Details of the other two cases are precisely the same and are left to the interested reader.

C. Comparison with Bayesian Incentive Compatibility

In general, the set of Bayesian incentive compatible decision rules is much larger than the set of posterior implementable rules. The main reason for this,

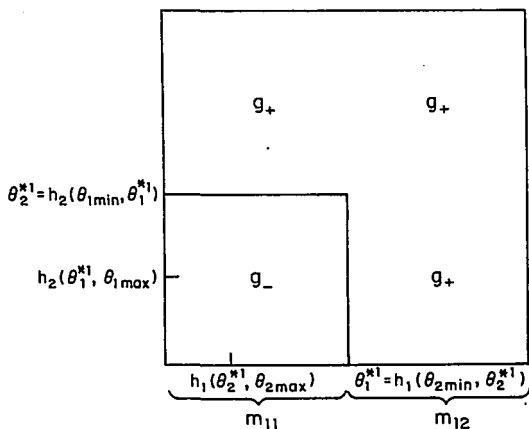


FIGURE 10

quite obvious, is that there is a single incentive-compatibility constraint for each private observation instead of separate constraints for each possible response by the other player.

For the class of environments studied in this paper, the possibility of correlation between θ_1 and θ_2 makes the complete characterization of the Bayesian incentive compatible rules very difficult (see Laffont and Maskin (1979) and Aspremont and Gérard-Varet (1982)). When θ_1 and θ_2 are independently distributed, any piecewise differentiable incentive compatible rule is characterized by the first and second order conditions that truthful strategies by each player are optimal, given that the other player is truthful.

To illustrate the restriction imposed by posterior implementability to Bayesian incentive compatible rules, we consider an example with a discrete set of characteristics. The extension of our theory to this case would be straightforward and would allow for weakly monotonic evaluation functions.

The matrix of payoffs is defined below where in each entry the first (second) number denotes agents 1's (2) payoff if the decision is d_1 :

	$\bar{\theta}_1$	$\bar{\bar{\theta}}_1$	
$\bar{\theta}_2$	1, -1/2	1, 1	
$\bar{\bar{\theta}}_2$	-1, -1	-1/2, 1	

The four pairs of observations are equally likely.

Mechanisms can be assumed to give each player two possible messages, without loss of generality. The g function can be represented by the matrix $(g(\theta_1, \theta_2))$. Figure 11 shows the ex ante utilities obtained by the two players under various mechanisms. The Bayesian incentive compatible mechanisms which are ex ante Pareto optimal generate the utilities described by the heavy line of Figure 11. They are generated by the extreme points: E is generated by the best mechanism for an utilitarian decision maker:

$$\begin{bmatrix} 2/3 & 1 \\ 0 & 2/3 \end{bmatrix};$$

B is generated by the dictatorship of agent 2:

$$\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix};$$

C is generated by the dictatorship of agent 1:

$$\begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}.$$

The set of all ex ante attainable utilities under Bayesian incentive compatibility is the convex hull of these three points with the trivial mechanism that always chooses d_0 and hence gives both players a zero utility.

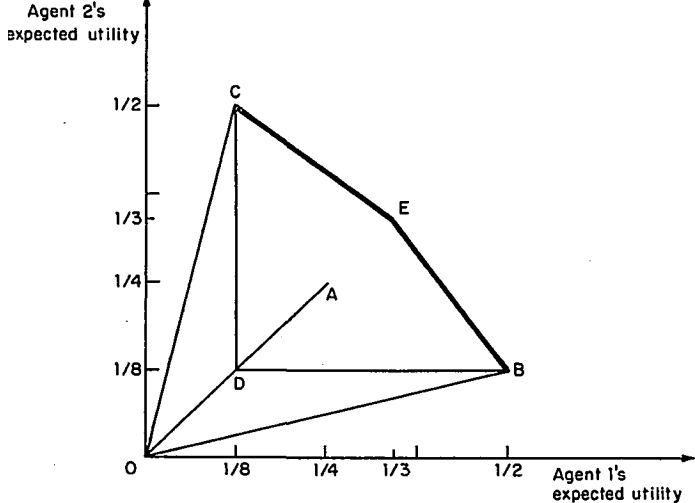


FIGURE 11

It is easy to see, however, that the point E cannot be sustained by a posterior implementable rule. It would require that all information be leaked and this would provide an incentive for one of the agents to change his message whenever the randomized outcome is realized.

Points B and C , being dictatorial rules, are of course posterior implementable. In addition, the trivial mechanism that always chooses d_1 generates the expected utilities at D .

The only other posterior implementable mechanism is the unanimity mechanism

$$\begin{bmatrix} 0 & \alpha \\ 0 & 0 \end{bmatrix} \text{ for } \alpha \in [0, 1].$$

Expected utilities are monotonic in α and at the best such mechanism, where $\alpha = 1$, they are shown as A in Figure 11.

The set of expected utilities generated by posterior implementable rules is not convex. For example, mixing the unanimity mechanism with the dictatorship of individual 1 yields mechanisms of the form

$$\begin{bmatrix} 0 & 1 \\ 0 & \alpha \end{bmatrix}.$$

Here individual 2 would not continue to tell the truth at $\theta_2 = \bar{\theta}_2$ after learning that $\theta_1 = \bar{\theta}_1$.

The set of expected utilities generated by posterior implementable mechanisms are shown by the light lines in Figure 11. They are generated by mechanisms that are convex combinations of the two trivial mechanisms (choosing d_0 or d_1) with unanimity or with either of the two dictatorial rules.

D. Relations with the Literature

The concept of posterior implementability was used by Pratt and Zeckhauser (undated) in a model with a transferable resource. They confined their attention to the problem of whether truthful revelation of the private parameter value would remain optimal, given a knowledge of other players' observations. For the particular class of environments they considered, the (full-information) socially optimal decision rule could be implemented using this equilibrium. It was therefore unnecessary for Pratt and Zeckhauser to define the more general, nontruthtelling concept of posterior implementability treated in this paper.

Milgrom (1981) mentions a property related to, but distinct from, posterior implementability in the context of auction design. In the Vickrey auction, where an indivisible object is sold to the highest bidder at the price offered by the second-highest bidder, the revelation of this second-highest price would not cause any of the bidders to wish to change their bids. However, if all the bids were to be mutually observable, then, in general, some of the participants would want to change their bids. Prospectively, under rational expectations, the first round of bidding would then be different from what it would have been had the bids been secret and binding offers. Such dynamic issues are not treated by Milgrom, nor is it obvious that our concept of posterior implementation is the appropriate construct. We do believe, however, that there is an interesting relationship and perhaps even an equivalence between rational expectations equilibria in dynamic models with nonbinding, observable offers and the equilibria established in a posterior implementation. We intend to explore this in future work.

On a related point, the work of Farrell (1982) should be mentioned. Farrell studies the decision rules that can be achieved if a preliminary phase of voluntary communication is used before the actual play of a game of incomplete information. During the communication phase, the agents may announce the moves they plan to make later, hoping to coordinate with their opponents. The perfect equilibria of these dynamic games of communication, which are characterized by Farrell, may be related to the posterior implementable decision rules. Again, this is a question requiring future work.

Posterior implementability can be compared to the concept of durability of a mechanism, as introduced by Holmstrom and Myerson (1983) and as further explored by Crawford (1985). A decision rule is durable if it is the Bayesian equilibrium of a mechanism and if, moreover, after the private information has been observed, no substitute mechanism can be introduced by any player that will be accepted by all players as superior to the one they are using. Durability, like posterior implementability, reduces the set of implementable decision rules. However, the two criteria are different in several respects. When implementing

a durable decision rule, the information revealed may make agents' indicated plays nonoptimal, and thus the decision rule is not posterior implementable. On the other hand, it is not obvious that every posterior implementable decision rule can be implemented in a durable fashion. The difference really turns on the unanimity requirement for overturning nondurable rules and on the fact that the mechanism selected to implement a posterior implementable rule is chosen before, rather than after, the private information is observed.

Finally, mention should be made of work which places related restrictions on the set of Bayesian incentive compatible mechanisms. Cramton (1984) uses the idea that the players' lack of ability to commit themselves to stop bargaining, although they can perceive a mutual benefit to continuing, may affect the conduct of their earlier negotiations. The set of incentive compatible bargaining mechanisms subject to this "no involuntary stopping" constraint is in general much smaller than the set of all incentive compatible mechanisms. Cramton calls them the perfect bargaining mechanisms, by analogy with the perfect equilibria of extensive games. This idea is related to posterior implementability, but is different in some important respects. Cramton's agents cannot recant an agreement to trade, once this agreement has been reached. In a posterior implementable mechanism, entire strategies can be revised because no binding agreements at all can be made. It should be noted, however, that whenever trade is consummated in Cramton's model it is mutually beneficial. Therefore, except insofar as the price represents a zero-sum transfer between the buyer and seller, neither party would actually want to alter or delay an agreement to trade.

E. Extensions

In this paper we assumed that all messages sent by the agents were publicly observable. An important extension is to examine cases in which only some aspects of these messages are observable. Specifically, suppose that the messages are sent to a central agent who uses them to select a public choice. It may therefore be reasonable that the agents will come to learn and draw inferences from this choice, but cannot see the details of all the messages. To the extent that the public choice is not a sufficient statistic for the messages, agents' posterior beliefs would not be as refined as we have assumed in this paper.

A related extension concerns cases where the public choice is a parameter (such as a price vector) in a subsequent decision problem faced by the agents. This is the situation in rational expectations models of the learning from prices variety (see Green (1973), Grossman (1981) and Laffont (1985)). Laffont (1985) has shown that rational expectations equilibria, when they are not completely revealing, may be inefficient within the class of all Bayesian implementable decision rules. It would be of interest to examine the efficiency issue within the smaller class of rules that are implementable subject to mutual knowledge of whatever "public" parameters are needed to control the system.

In the narrower context of extensions of the present model, there are several obvious directions to pursue, for example, cases of more than two players, more

than two possible decisions, or the inclusion of a transferable resource. We also would like to determine whether there are welfare relationships among the posterior implementable decision rules and whether the number of messages used by each of the players is indicative of the utility levels attainable.

Department of Economics, Harvard University. Cambridge, MA 02138, U.S.A.
and
GREMAQ, Université des Sciences Sociales de Toulouse, Toulouse, France

Manuscript received December, 1984; final revision received February, 1986.

REFERENCES

- ASPREMONT, C. D', AND L. A. GÉRARD-VARET (1982): "Bayesian Incentive-Compatible Beliefs," *Journal of Mathematical Economics*, 10, 83-104.
- CRAMTON, P. (1984): "Sequential Bargaining Mechanisms," in "Two Papers on Sequential Bargaining," IMSSS Technical Report No. 444, Stanford.
- CRAWFORD, V. (1985): "Efficient and Durable Decision Rules: A Reformulation," *Econometrica*, 53, 817-836.
- FARRELL, J. VON R. (1982): "Communication in Games," Mimeo, Massachusetts Institute of Technology.
- GREEN, J. (1973): "Information, Efficiency and Equilibrium," Discussion Paper No. 284, Harvard Institute of Economic Research.
- GROSSMAN, S. (1981): "An Introduction to the Theory of Rational Expectations Under Asymmetric Information," *Review of Economic Studies*, 48, 541-560.
- HOLMSTRÖM, B. (1979): "Groves Schemes on Restricted Domains," *Econometrica*, 47, 1137-1144.
- HOLMSTRÖM, B., AND R. MYERSON (1983): "Efficient and Durable Decision Rules with Incomplete Information," *Econometrica*, 51, 1799-1820.
- LAFFONT, J. J., AND E. MASKIN (1979): "A Differential Approach to Expected Utility Maximizing Mechanisms," in *Aggregation and Revelation of Preferences*, ed. by J. J. Laffont. Amsterdam: North-Holland, pp. 289-308.
- LAFFONT, J. J. (1985): "On the Welfare Analysis of Rational Expectations Equilibria with Asymmetric Information," *Econometrica*, 53, 1-30.
- MILGROM, P. (1981): "Rational Expectations, Information Acquisition, and Competitive Bidding," *Econometrica*, 49, 921-944.
- PRATT, J., AND R. ZECKHAUSER (undated), "Honest Information for Group Decision," Mimeo, Harvard University.