Discussion Paper No. 548

BAYESIAN EQUILIBRIUM AND INCENTIVE-COMPATIBILITY: AN INTRODUCTION

by

Roger B. Myerson

February 1983
revised, June 1983

J. L. Kellogg Graduate School of Management
Northwestern University
Evanston, Illinois 60201

Abstract. This paper is an introduction to the analysis of games with incomplete information, using a Bayesian model. The logical foundations of the Bayesian model are discussed. To describe rational behavior of players in a Bayesian game, two basic solution concepts are presented: Bayesian equilibrium, for games in which the players cannot communicate; and Bayesian incentive-compatibility, for games in which the players can communicate. The concept of virtual utility is developed as a tool for characterizing efficient incentive-compatible coordination mechanisms.

Acknowledgements. Research for this paper was supported by the Kellogg Center for Advanced Study in Managerial Economics and Decisions Sciences, and by a research fellowship from I.B.M. The author is grateful to Hugo Sonnenschein for detailed comments and discussion.

This paper is to appear in Social Goals and Social Organization, Essays in Memory of Elisha A. Pazner, edited by L. Hurwicz, D. Schmeidler, and H. Sonnenschein.

1. Introduction

Two kinds of incentive-constraints limit people's ability to reach mutually beneficial agreements in social and economic affairs. First, when one person has unverifiable private information that is not available to the others, then he cannot be compelled to reveal that information honestly unless he is given the correct incentives. Second, when a person controls some private decision variable that others cannot control or monitor, then he cannot be directed to choose any particular decision or action unless he is given the incentive to do so. That is, a social contract or coordination system may not be feasible if it gives people incentives to lie about their information or to cheat in their actions. An organization must give its members the correct incentives to share information and act appropriately. An individual cannot be relied upon to testify against himself or to exert efforts for which he will not be rewarded.

It is widely recognized by economists and other social scientists that this need to give correct incentives may be quite costly for society. In the insurance industry, for example, the inability to get individuals to reveal unfavorable information about their chances of loss is known as adverse selection, and the difficulty of getting fully insured individuals to exert efforts against their insured losses is known as moral hazard. These factors generally prevent the insurance industry from offering risk-averse individuals the full insurance that they would like to buy. Arrow [1970] has written a seminal analysis of these issues and their impact on markets for risk-bearing.

A theory of incentives must go beyond simply telling us that certain ideal forms of social organization are infeasible because they violate incentive constraints, and that incentive constraints cause losses in social

welfare. We also need to know how to minimize these losses. That is, given a social welfare function, we may want to find the best contract or social system that maximizes social welfare subject to these incentive constraints. In this paper we will see how the theory of Bayesian equilibrium and incentive-compatibility can be used to actually find such optimal contracts.

The basic object of analysis in this paper is a Bayesian game with incomplete information, as defined by Harsanyi [1967-8]. In our notation, we suppose that there are n players in the game, and that they are numbered 1, 2, ..., n. For each player i in $\{1,2,\ldots,n\}$, we let D_i denote the set of possible actions or strategic decisions available to player i in the game. We let T_i denote the set of possible types for player i. Each type t_i in T_i is a complete description of one possible state of player i's private information and beliefs about any uncertain factors relevant to the game (for example, about the preferences and abilities of various players). That is, a player's type is supposed to be a random variable that summarizes all information that he may have which is not available to the other players.

Let D denote the set of possible combinations of decisions available to the n players, and let T denote the set of possible types of the n players, so that

$$D = D_1 \times \ldots \times D_n, \tag{1.1}$$

$$T = T_1 \times \ldots \times T_n. \tag{1.2}$$

Let T_{-i} denote the set of possible combinations of types for all players other than i, so

$$T_{-i} = T_1 \times \ldots \times T_{i-1} \times T_{i+1} \times \ldots \times T_n. \tag{1.3}$$

Except in section 2, we will usually assume that D and T are finite sets.

We let $p_i(t_{-i}|t_i)$ denote the subjective probability that player i would assign to the event that t_{-i} in T_{-i} is the combination of other players' types, if i's actual type were t_i. We let $u_i(d,t)$ denote utility payoff (measured in some von Neumann-Morgenstern utility scale) that player i would get if $d = (d_1, \ldots, d_n)$ were the combination of decisions chosen by the n players and $t = (t_1, \ldots, t_n)$ were the combination of the players' types.

Thus, in general, we say that Γ is a Bayesian game iff it is of the form

$$\Gamma = (D_1, \ldots, D_n, T_1, \ldots, T_n, p_1, \ldots, p_n, u_1, \ldots, u_n) \tag{1.4}$$

where, for each i, D_i and T_i are nonempty sets, p_i is a function specifying a probability distribution $(p_i(\cdot|t_i))$ over T_{-i} for each t_i in T_i, and u_i is a function mapping $D \times T$ into the real numbers $\mathbb{R}$. In a Bayesian game, we assume that the structure of Γ in (1.4) is common knowledge among all the players, plus each player i knows his own actual type in T_i. (Following Aumann [1976], we say that a fact is common knowledge iff everyone knows it, everyone knows that everyone knows it, and so on, including every statement of the form "everyone knows that everyone knows that ... everyone knows it.")

Bayesian games are important for economic theory because they give us a general model for situations involving moral hazard and adverse selection. The goal of this paper is to provide a general introduction to the analysis of Bayesian games. In section 2, we show that the Bayesian game model is (in principle) the appropriate model for any game with incomplete information, following the work of Harsanyi [1967-8] and Mertens and Zamir [1982]. In section 3 we discuss equivalence relations between Bayesian games. In section 4, we argue that Bayesian equilibrium is the appropriate solution concept for Bayesian games, if the players cannot communicate. For games in which the players can communicate, we define Bayesian incentive compatibility

in section 5, to characterize the set of feasible coordination mechanisms for the players. An incentive-efficient mechanism is one that is Pareto-undominated within the set of incentive-compatible mechanisms. In sections 6, 7, and 8, we develop necessary and sufficient conditions which can be used to actually compute incentive-efficient mechanisms. Section 6 is devoted to the special case in which there are only informational incentive constraints (the case of pure adverse selection); section 7 is devoted to the case in which there are only strategic incentive constraints (pure moral hazard); and section 8 covers the general case.

2. Modelling games with incomplete information

We say that there is incomplete information in a game if, at the time when the players choose their strategies for playing the game, they have different private information about their preferences and abilities. This term was introduced by Von Neumann and Morgenstern [1944]. (They also used the term imperfect information, to describe games in which the players may get different private information during the course of the game, but all players begin the game with the same information. The distinction between the two terms seems to depend on whether the players actually could have planned their strategies in the game before learning their private information.) The real understanding and analysis of games with incomplete information began with the work of Harsanyi [1967-8], who introduced the basic definition of a Bayesian game and argued that it is the appropriate model for games with incomplete information. Mertens and Zamir [1982] developed a rigorous mathematical formulation of Harsanyi's argument. In this section, we review the ideas of these two important papers, using a formulation based on

(but slightly different from) that of Mertens and Zamir. Armbruster and Böge [1979] have also considered a related formulation.

A model of a game with incomplete information must include variables that describe what private information each player might have that is unavailable to other players. In Harsanyi's Bayesian games, these variables are the players' types. Thus, player i's type must specify everything that player i knows that is not common knowledge among all players. For example, if player i's only private information is his reservation wage rate, then we can let his set of possible types T_i be a subset of the real numbers, where each t_i in T_i is a possible value of player i's reservation wage. On the other hand, if some players do not know what are i's beliefs about other players' reservation wages, then player i's type must be expanded to also include parameters that specify player i's beliefs about other players' reservation wages. In this case, T_i might have to be a set of vectors, rather than a set of numbers.

The basic question to be considered in this section is the following. When we are trying to model some real-world situation in which players have incomplete information, can we always find type-sets $(T_1, \ldots, T_n)$ that are large enough to characterize all of the possible private information and beliefs that a player might have relevant to the game? To answer this question, we must consider what are the uncertainties that may arise in the structure of a game, and we must show that the players' beliefs about all these uncertainties can be specified within the type-sets of some Bayesian game.

There are several basic issues in a game about which players might have different information: how many players are actually in the game; what actions or strategic decisions are available to each player; how the outcome of the game depends on the actions chosen; and what are the players' preferences over the set of possible outcomes. Harsanyi showed that all of

these issues can be modelled in a unified way. Uncertainty about whether a particular player is "in the game" can be converted into uncertainty about the set of feasible decisions, by always including the player in the game but then giving him only one decision (= "nonparticipation") when he is supposed to be "out of the game." Uncertainty about whether a particular decision is feasible for player i can in turn be converted into uncertainty about the outcome, by saying that player i will get a very bad (negative) payoff if he uses a decision that is supposed to be infeasible. Uncertainty about outcomes and uncertainty about preferences can be unified by modelling each player's utility function directly from the space of decision-combinations into utility payoffs (representing the composition of an outcome function, that maps decision-combinations into outcomes, and a utility function, that maps outcomes into a vonNeumann-Morgenstern utility scale for the player).

So let $\{1,2,\ldots,n\}$ be the set of players, let D_i be the set of possible actions or strategic decisions for player i, and let D being the set of possible combinations of decisons, as in (1.1). To be consistent with the preceding discussion, we might say that n is the maximal number of players, and D_i is the maximal set of feasible decisions for player i.

To model the uncertainty in the game, we must put some unknown parameter $\tilde{\theta}$ into the utility functions. Thus, we let $w_i(d,\tilde{\theta})$ denote the utility payoff to player i if $d = (d_1, \ldots, d_n)$ is the combination of actions chosen by the n players and if $\tilde{\theta}$ is the value of this unknown parameter. We let H denote the set of possible values of $\tilde{\theta}$, and we refer to H as the domain of basic uncertainty in the game. If D is finite, we can assume without loss of generality that H is a subset of $\mathbb{R}^{n|D|}$, because the only role of $\tilde{\theta}$ is to specify the n utility functions from D into the real numbers $\mathbb{R}$. Furthermore, if the players' utility functions are bounded, then we can assume that H is a

subset of the $n|D|$-dimensional unit cube.

These structures $(D_1, \ldots, D_n, H, w_1, \ldots, w_n)$ are not sufficient to describe the game with incomplete information, because they do not tell us what are the players' beliefs or information about the unknown parameter $\tilde{\theta}$. The subjectivist theory of Bayesian decision-making, as developed by Savage [1954], Raiffa [1968], and others, emphasizes that any individual must have a subjective probability distribution over the possible values of any parameter that he does not know. That is, if player i does not know $\tilde{\theta}$, then he must at least have some subjective probability distribution over H that summarizes his beliefs about this unknown parameter $\tilde{\theta}$. His subjective probability distribution for $\tilde{\theta}$ can be measured by asking him questions about which gambles depending on $\tilde{\theta}$ he would prefer. (For example, to assess a player's subjective probability of the event that $\tilde{\theta}$ is in a set Ψ, where $\Psi \subseteq H$, we would ask him, for what objective probability of getting an increase of one utility-unit independently of $\tilde{\theta}$ would he be just barely willing to give up a prospect of gaining one extra utility-unit if $\tilde{\theta}$ is in Ψ.) Our description of a player as a rational decision-maker will be incomplete until we specify these subjective probabilities.

We let $\tilde{q}_i^1$ represent player i's subjective probability distribution over H. That is, for any $\Psi \subseteq H$, $\tilde{q}_i^1(\Psi)$ is i's subjective probability for the event that $\tilde{\theta} \in \Psi$. We refer to $\tilde{q}_i^1$ as the <u>first-order beliefs</u> of player i.

In a game, a player's optimal decision will generally depend on what he expects the other players to do. And what he expects the other players to do will depend on what he thinks they believe. Thus we must now ask, what does player i think are the other n-1 players' first-order beliefs? Subjectivist decision theory implies that each player i must have a subjective probability distribution for these unknown first-order beliefs $(\tilde{q}_1^1, \ldots, \tilde{q}_{i-1}^1, \tilde{q}_{i+1}^1, \ldots, \tilde{q}_n^1)$

as well as for $\tilde{\theta}$. We let $\tilde{q}_i^2$ denote this subjective probability distribution. We refer to $\tilde{q}_i^2$ as the second-order beliefs of player i. But now there are third-order beliefs (beliefs about the other players' second-order beliefs) to be assessed, and so on. We seem to be getting into an infinite regress.

Mertens and Zamir [1982] have shown that it is possible to keep track of this infinite hierarchy of beliefs within a consistent mathematical model, so that there does exist a Bayesian game with type sets that are sufficiently large to include all of a player's possible beliefs of all orders. To see how this is done, we must use some relatively sophisticated mathematics. Readers with less mathematics are encouraged to skim or even omit the rest of this section, as nothing in sections 3 through 8 will depend on it.

Given any metric space X, we let $\Delta(X)$ denote the set of all probability distributions on X that are defined on the set of Borel-measurable subsets of X. We give $\Delta(X)$ the weak topology, which is defined so that $\int f(x)p(dx)$ is a continuous function of p in $\Delta(X)$ for every bounded continuous $f:X \to \mathbb{R}$. If X is compact, then $\Delta(X)$ is also compact and metrizable. Billingsley [1968] gives a full development of this result.

Now, let Q_i^1 denote the set of i's possible first-order beliefs (probability distributions over H); that is

$$Q_i^1 = \Delta(H). \tag{2.1}$$

We can inductively define Q_i^k, the set of possible k-order beliefs of player i, for k= 2,3,4,..., by

$$Q_i^k = \Delta(H \times Q_{-i}^{k-1}) \tag{2.2}$$

where $$Q_{-i}^{k-1} = Q_1^{k-1} \times \ldots \times Q_{i-1}^{k-1} \times Q_{i+1}^{k-1} \times \ldots \times Q_n^{k-1}.$$

That is, a k-order belief for player i is a probability distribution over the possible values of $\tilde{\theta}$ and the other players' (k-1)-order beliefs. By induction, if H is compact then every Q_i^k is also a compact set (with the weak topology). We let $\tilde{q}_i^k$ denote the actual k-order beliefs of player i, in Q_i^k.

A player's k-order beliefs determine his beliefs of all orders lower than k, through a series of functions $\phi_i^{k-1}: Q_i^k \to Q_i^{k-1}$, which can be defined inductively. The function ϕ_i^1 is defined by

$$(2.3) \qquad [\phi_i^1(q_i^2)](\Psi) = q_i^2(\Psi \times Q_{-i}^1), \quad \forall q_i^2 \in Q_i^2, \ \forall \Psi \subseteq H.$$

That is, the first-order beliefs $\phi_i^1(q_i^2)$ that correspond to second-order beliefs q_i^2 are just the marginal distribution of q_i^2 on H. We inductively define $\phi_i^{k-1}(q_i^k)$, for every $k \geq 3$ and every q_i^k in Q_i^k, by

$$(2.4) \qquad (\phi_i^{k-1}(q_i^k))(\Psi) = q_i^k(\{(\theta,(q_j^{k-1})_{j\neq i}) \mid ((\theta,(\phi_j^{k-2}(q_j^{k-1}))_{j\neq i}) \in \Psi\}),$$

$$\forall \Psi \subseteq H \times Q_{-i}^{k-2}.$$

That is, the probability under $\phi_i^{k-1}(q_i^k)$ of a set of (k-2)-order beliefs is the probability under q_i^k of the (k-1)-order beliefs that are mapped into the set by the functions ϕ_j^{k-2}. By the laws of probability, each player's first-order beliefs must be the marginal distribution of his second-order beliefs on H, so $\tilde{q}_i^1 = \phi_i^1(\tilde{q}_i^2)$ for each player i. Each player i also knows that $\tilde{q}_j^1 = \phi_j^1(\tilde{q}_j^2)$ for every other player j (since i knows that j's beliefs satisfy the laws of probability), and this fact implies that $\tilde{q}_i^2 = \phi_i^2(\tilde{q}_i^3)$ for player i. Continuing inductively, we conclude that

$$\tilde{q}_i^{k-1} = \phi_i^{k-1}(\tilde{q}_i^k) \quad \forall i, \ \forall k \geq 2,$$

because it is common knowledge that every player's beliefs satisfy the laws of probability.

We let Q_i^∞ denote the set of all possible beliefs of all orders for player i, that is

$$(2.5) \qquad Q_i^\infty = \{q_i=(q_i^1,q_i^2,\ldots)\varepsilon \underset{k=1}{\overset{\infty}{\times}} Q_i^k \mid q_i^{k-1} = \phi_i^{k-1}(q_i^k), \quad \forall k\}.$$

In the terminology of Mertens and Zamir [1982], Q_i^∞ is the <u>universal belief space</u> for player i generated by H, the domain of basic uncertainty. Mertens and Zamir have shown that, if H is compact, then the universal belief space generated by H is also a compact topological space (with the product topology).

Any q_i in Q_i^∞ induces a probability distribution on $H \times Q_{-i}^\infty$, where

$$Q_{-i}^\infty = Q_1^\infty \times \ldots \times Q_{i-1}^\infty \times Q_{i+1}^\infty \times \ldots \times Q_n^\infty,$$

and we let $P_i(\cdot|q_i)$ denote this probability distribution. If Ψ is any closed subset of $H \times Q_{-i}^\infty$, then the induced probability of Ψ is

$$(2.6) \qquad P_i(\Psi|q_i) = \lim_{k\to\infty} q_i^k(\{(\theta,(q_j^{k-1})_{j\neq i})|(\theta,(q_j)_{j\neq i}) \varepsilon \Psi\}).$$

(Here q_i^k denotes the k-order component of q_i, and q_j^{k-1} denotes the (k-1)-order component of q_j.) In fact, Mertens and Zamir have shown that $P_i(|\cdot)$ is a homeomorphism between Q_i^∞ and $\Delta(H \times Q_{-i}^\infty)$. That is, player i's universal belief space Q_i^∞ includes all possible (Borel-measurable) beliefs about the basic uncertainty in H and the other players' infinite hierarchies of beliefs in Q_{-i}^∞.

Notice now that the random variable $\tilde{\theta}$ cannot directly influence any player's behavior in the game, except to the extent that players have information about $\tilde{\theta}$ that is expressed in their beliefs $(\tilde{q}_1, \ldots, \tilde{q}_n)$. So we can integrate the basic uncertainty variable out of the probability and utility functions without losing any structures relevant to predicting players' behavior. For any q_i in Q_i^∞, we let $p_i(\cdot|q_i)$ be the marginal

probability distribution of $P_i(\cdot|q_i)$ on Q_{-i}^∞. For any $q = (q_1, \ldots, q_n)$ in $\times_{j=1}^n Q_j^\infty$, we let $u_i(d,q)$ denote the conditional expectation of $w_i(d,\tilde{\theta})$, under the conditional probability distribution for $\tilde{\theta}$ induced by q_i, given that $\tilde{q}_{-i}$ (the vector of actual beliefs of players other than i) is equal to the vector $q_{-i} = ((q_1,\ldots,q_{i-1},q_{i+1},\ldots,q_n)$. That is, we may write:

$$(2.7) \qquad p_i(\Psi|q_i) = P_i(H \times \Psi|q_i), \qquad \forall i,\ \forall q_i \in Q_i^\infty,\ \forall \Psi \subseteq Q_{-i}^\infty;$$

$$(2.8) \qquad u_i(d,q) = E_{q_i}(w_i(d,\tilde{\theta})|\ \tilde{q}_{-i} = q_{-i}), \qquad \forall i,\ \forall d \in D,\ \forall q \in \times_{j=1}^n Q_j^\infty.$$

Thus at last we get the universal Bayesian game,

$$\Gamma^\infty = (D_1, \ldots, D_n, Q_1^\infty, \ldots, Q_n^\infty, p_1, \ldots, p_n, u_1, \ldots, u_n).$$

For each i and each q_i in Q_i^∞, $p_i(\cdot|q_i)$ is a probability distribution over Q_{-i}^∞, and u_i is a function from $D \times (\times_{j=1}^n Q_j^\infty)$ into $\mathbb{R}$; so Γ^∞ is indeed a Bayesian game. By construction, Q_i^∞ is large enough to include all possible private information or beliefs that player i might have about the preferences and beliefs of all players in the game.

At this point, however, we must admit that our model seems to have gotten out of hand. Compact or not, Q_i^∞ is an extremely complex mathematical object, by any standards of intuition. We started out to describe games in which players have some uncertainty about each others' preferences and beliefs. We found that, in such games, the beliefs of each player consist of an infinite sequence of subjective probability distributions over sets of probability distributions. The higher-order beliefs of a player could be critical to determining how he plays the game, so game-theoretical analysis requires that, for each player, this whole sequence of subjective probability distributions

must be specified by a variable in our model. But the set of all such sequences of probability distributions is too large for practical analysis, either by game theorists or by the players in the game! Thus, for a tractable and relevant model, the players' beliefs must be restricted to some smaller subsets of universal belief space.

The way to limit the explosion of uncertainty about beliefs about beliefs is to assume that it is common knowledge that the beliefs of each player i are in some set T_i which is a small subset of Q_i^∞. This idea is the key insight of Harsanyi's classic paper. If each set T_i is tractably small (finite, or parameterized by a single variable in $\mathbb{R}$, for example) the result will be a manageable model which can give useful insights.

For it to be common knowledge that the actual type of each player i is in T_i, the set $T_1 \times \ldots \times T_n$ must be a <u>belief-closed</u> subset of $Q_1^\infty \times \ldots \times Q_n^\infty$, in the sense that

$$p_i(T_{-i} \mid t_i) = 1, \qquad \forall i, \quad \forall t_i \in T_i, \tag{2.9}$$

where T_{-i} is as in (1.3). That is, (2.9) asserts that every type in T_i puts probability one on the event that every other player j has beliefs corresponding to some type in T_j.

Mertens and Zamir have shown that finite belief-closed subsets are dense among the belief-closed subsets of $Q_1^\infty \times \ldots \times Q_n^\infty$, in a topology that seems natural (the Hausdorff topology for closed sets). This result suggests that there may be "almost" no loss of generality in assuming that the players' beliefs are in such a finite belief-closed subset.

Thus, let us assume that there is such a finite belief-closed set $T = T_1 \times \ldots \times T_n$ such that it is common knowledge that every player i has beliefs that correspond to some point in T_i. Then we can refer to T_i as the set of possible types for player i; and by restricting the functions p_i and u_i

to the domain $T \subseteq Q_1^\infty \times \ldots \times Q_n^\infty$, we get a finite Bayesian game Γ as in (1.4).

In general, of course, the type-sets $(T_1,\ldots,T_n)$ in a Bayesian game Γ do not actually need to be specified as subsets of universal belief space. For example, as remarked above, if a player's only private information is his reservation wage rate, then we can simply let T_i be a set of the real numbers, where each t_i in T_i is a possible value of i's reservation wage. Given any Bayesian game Γ as in (1.4), for every type t_i in T_i, the corresponding infinite hierarchy of beliefs (i's beliefs about the other players' types, his beliefs about their beliefs, etc.) can be computed from the probability functions $(p_1,\ldots,p_n)$; so T_i is isomorphic to a subset of a universal belief space, even if it is not identified as such. The purpose here of developing the concept of universal belief space was only to verify that any game situation with incomplete information can in principle be modelled as a Bayesian game, by letting each T_i equal Q_i^∞ if no smaller sets will do. On the other hand, we must recognize that the complexity of universal belief space implies that the Bayesian-game model will in practice be applicable only to those game situations where there is enough common-knowledge structure so that each player's private information can be described within a small and tractable set of types.

3. Consistent beliefs, and equivalent Bayesian games

Harsanyi defined the beliefs $(p_1, \ldots, p_n)$ to be consistent iff there exists a probability distribution p^* on the set T such that each players' conditional distribution, given his own type, is identical to that which would have been computed from p^* by Bayes theorem; that is,

$$(3.1) \qquad p_i(t_{-i}|t_i) = p^*(t)/p_i^*(t_i), \quad \forall t_i \in T_i, \quad \forall t_{-i} \in T_{-i},$$

where

$$p_i^*(t_i) = \sum_{t_{-i} \in T_{-i}} p^*(t), \qquad \forall t_i \in T_i. \tag{3.2}$$

(We use here the convention that, whenever t, t_{-i}, and t_i appear in the same formula, then t is the vector of types with i^{th} component t_i and all other components as in t_{-i}.) Harsanyi has argued that we might expect that most Bayesian games which describe real situations ought to be consistent, because the players' types may have been jointly determined before the game by some chance event governed by the distribution p^*.

We have been careful not to speak of "i's subjective probability distribution over T_i" at any point in this discussion. This is because player i already knows his type when the game begins. Even if there had been a time before the game when he did not know his type (and there might not have been any such time, for example if the type is his or her gender), the subjective probability distribution that he would have assessed for his own type cannot have any decision-theoretic significance in the play of the game. However, if there had been a time before the game when no player knew his type and if all players had the same prior beliefs p^*, then the type-conditional beliefs $(p_1,\ldots,p_n)$ should be consistent with p^*.

Interpersonal comparisons of utility cannot be given decision-theoretic significance. That is, there is no decision-theoretic meaning for a statement such as "a movie gives me more utility than an opera gives you", because neither of us could ever be forced to choose between being me at a movie or being you at an opera. Now, for games with incomplete information, we assume that each player already knows his own type before he makes any decisions relevant to the game. Thus, when the game is played, intertype comparisons of utility are also decision-theoretically meaningless. When a player already

knows his type, he cannot be asked to choose it. We cannot ask a player "would you prefer to be an opera fan at the opera or be a non-opera-fan at the movies", when he already knows whether he is an opera fan or not.

Thus, the utility scales of different types can be specified separately. From basic decision theory, it is well known that vonNeumann-Morgenstern utility scales can only be defined up to increasing linear transformations. Thus we say that two Bayesian games with the same decision sets and type sets

$$\Gamma = (D_1, \ldots, D_n, T_1, \ldots, T_n, p_1, \ldots, p_n, u_1, \ldots, u_n)$$

and

$$\hat{\Gamma} = (D_1, \ldots, D_n, T_1, \ldots, T_n, \hat{p}_1, \ldots, \hat{p}_n, \hat{u}_1, \ldots, \hat{u}_n)$$

are utility-equivalent iff they have the same conditional probability distributions (so $\hat{p}_i = p_i$ for all i) and there exist numbers $a_i(t_i)$ and $b_i(t_i)$, for each i and each t_i in T_i, such that

$$a_i(t_i) > 0 \quad \text{and} \quad \hat{u}_i(d,t) = a_i(t_i)\, u_i(d,t) + b_i(t_i), \quad \forall d \in D, \quad \forall t \in T.$$

That is, utility-equivalent Bayesian games differ only in that the utility functions of some types of some players may be linearly rescaled. The Bayesian equilibria and incentive-compatible mechanisms (to be defined later) of two utility-equivalent games will be the same.

Whenever a player chooses an action or decision in a Bayesian game, his criterion for the best decision is that it should give him the highest conditionally expected utility, given his actual type. Expected utility is computed by multiplying utilities times probabilities and then summing over all possible values of the unknowns. For example, if some

function $\sigma:T \to D$ determined how the players' decisions depend on their types, then the conditionally expected utility for type t_i of player i would be

$$\sum_{t_{-i} \in T_{-i}} p_i(t_{-i}|t_i)\, u_i(\sigma(t),t).$$

We define $z_i:D\times T \to \mathbb{R}$ by

$$z_i(d,t) = p_i(t_{-i}|t_i)\, u_i(d,t), \quad \forall d\in D, \quad \forall t\in T,$$

and we call z_i the evaluation function for player i. (See Wilson [1968] and Myerson [1979b] for the origins of this term.) Because only this product of probability-times-utility matters in computing expected utilities, we say that two Bayesian games are probability-equivalent iff they have the same decision sets D_i and type sets T_i and evaluation functions z_i for all players; that is

$$\hat{p}_i(t_{-i}|t_i)\, \hat{u}_i(d,t) = p_i(t_{-i}|t_i)\, u_i(d,t), \quad \forall i, \forall d\in D, \quad \forall t\in T.$$

Probability-equivalence is important because it assures us that consistency of beliefs is not an issue of basic importance in studying general Bayesian games. In particular, if the type sets are all finite then any Bayesian game is probability-equivalent to another Bayesian game with consistent beliefs, and even with stochastically independent types for the n players. (Simply let

$$\hat{p}_i(t_{-i}|t_i) = \frac{1}{|T_{-i}|} \quad \text{and} \quad \hat{u}_i(d,t) = |T_{-i}|\, p_i(t_{-i}|t_i)\, u_i(d,t).)$$

Consistency of beliefs can be important only when we also want to make some restrictions on the form of the utility functions, such as when we assume that there is transferable utility, or that one player's utility depends only on his own type, or that utility functions are continuous in strategies and types. (This last condition would only be relevant when infinite type sets

are considered. See Milgrom and Weber [1981] for a comprehensive analysis of this issue.)

In fact, a more general equivalence relation can be defined among Bayesian games. We say that two Bayesian games Γ and $\hat{\Gamma}$ with the same decision sets and type sets are evaluation-equivalent iff, for every player i, there exist functions $a_i: T_i \to \mathbb{R}$ and $b_i: T \to \mathbb{R}$ such that

$a_i(t_i) > 0$ for every t_i and

$$\hat{p}_i(t_{-i}|t_i)\,\hat{u}_i(d,t) = a_i(t_i)\,p_i(t_{-i}|t_i)\,u_i(d,t) + b_i(t), \quad \forall d\in D, \forall t\in T.$$

Notice that the additive constant can depend on all players' types, while the multiplicative constant can only depend on i's type. All our solution concepts (Bayesian equilibrium and Bayesian incentive-compatibility) will be invariant under any evaluation-equivalent transformation of the game. It can be shown that evaluation-equivalence is the most general equivalence relation that preserves each type's preference ordering over coordination mechanisms (which will be defined in section 5).

4. Bayesian equilibrium

The decision or action chosen by a player in a Bayesian game will generally depend on his type. However, other players do not know player i's actual type, so in choosing their actions they must be concerned with what actions would be chosen by each of player i's possible types. An equilibrium of a Bayesian game is a set of conjectures about how each player would choose his action as a function of his type, such that each type of each player is maximizing his conditionally expected utility given his own type and the functional conjectures about the other players. Formally, $(\sigma_1,\ldots,\sigma_n)$ is a

Bayesian equilibrium of the Bayesian game Γ iff, for every player i, σ_i is a function from T_i to D_i such that, for every t_i in T_i,

$$(4.1) \qquad \sum_{t_{-i} \in T_{-i}} p_i(t_{-i}|t_i)\, u_i(\sigma(t),t)$$

$$= \max_{d_i \in D_i} \sum_{t_{-i} \in T_{-i}} p_i(t_{-i}|t_i)\, u_i((\sigma_{-i}(t_{-i}),d_i),\, t).$$

(Here $\sigma(t) = (\sigma_1(t_1),\ldots,\sigma_n(t_n))$,
and $(\sigma_{-i}(t_{-i}),d_i) = (\sigma_1(t_1),\ldots,\sigma_{i-1}(t_{i-1}),d_i,\sigma_{i+1}(t_{i+1}),\ldots,\sigma_n(t_n)).$)
Equation (4.1) asserts if player i were of type t_i and he expected the other players to select their actions according to their $\sigma_j(\cdot)$ rules, then the action $\sigma_i(t_i)$ would be optimal for him, in that it maximizes his conditionally expected utility.

Bayesian equilibrium is the fundamental solution concept for Bayesian games with incomplete information. Our goal, as theorists analyzing a Bayesian game, must be to predict how each player will choose his decision as a function of his type. Without knowing his type, we cannot hope to predict his actual decision; we can only predict how his decision functionally depends on his type in T_i. If the players themselves also understand these predictions then, unless the predictions constitute a Bayesian equilibrium, at least one type of one player would expect to do better by using some unpredicted decision. Thus, a prediction of the players' behavior can be rationally self-fulfilling if and only if it is a Bayesian equilibrium.

For a simple two-player example, suppose that $D_1 = \{x_1,y_1\}$, $D_2 = \{x_2,y_2\}$, $T_1 = \{1\}$ (so player 1 has only one possible type and no private information), $T_2 = \{2a,\ 2b\}$, $p_1(2a|1) = 0.6$, $p_1(2b|1) = 0.4$, and the payoffs (u_1,u_2) depend on the actions and player 2's type through the following two

bimatrices.

$t_2 = 2a$	x_2	y_2
x_1	1,2	0,1
y_1	0,4	1,3

$t_2 = 2b$	x_2	y_2
x_1	1,3	0,4
y_1	0,1	1,2

TABLE 1

In this game, x_2 is a dominant strategy for type 2a, and y_2 is a dominant strategy for type 2b. Player 1 wants to get either (x_1,x_2) or (y_1,y_2), and he thinks that type 2a is more likely than 2b. Thus, the unique Bayesian equilibrium of this game is

$$\sigma_1(1) = x_1, \quad \sigma_2(2a) = x_2, \quad \sigma_2(2b) = y_2.$$

This example is of interest because it illustrates the danger of analyzing each bimatrix separately, as if it were a game with complete information, when the game is really one of incomplete information. If it were common knowledge that 2's type was 2a, then the players would be in the left bimatrix, where the unique equilibrium is (x_1,x_2). If it were common knowledge that 2's type was 2b, then the players would be in the right bimatrix, where the unique equilibrium is (y_1,y_2). Thus, if we looked only at the full-information Nash equilibria of the two bimatrices, then we might make the prediction "the outcome of this game will be (x_1,x_2) if player 2's type is 2a and will be (y_1,y_2) if player 2's type is 2b."

This prediction would be absurd, however, for the actual game with incomplete information, in which player 1 does not initially know player 2's type. Notice first that this prediction ascribes two different actions to

player 1, depending on 2's type (x_1 if 2a, and y_1 if 2b). So player 1 could not behave as predicted unless he got some information from player 2. But player 2 prefers (y_1,y_2) over (x_1,x_2) if he is 2a, and he prefers (x_1,x_2) over (y_1,y_2) if he is 2b. Thus, even if we revised the game to allow communication between the players before player 1 chooses among x_1 and x_2, player 2 would never communicate the information needed to fulfill this prediction, because it always gives him his less-preferred outcome. Instead, he would rather manipulate his communications to get the outcomes (y_1,y_2) if 2a, and (x_1,x_2) if 2b.

5. Bayesian games with communication

When we defined Bayesian equilibrium as the solution concept for Bayesian games, we assumed that each player's decision in a Bayesian game could depend only on his own type. Let us now consider what can happen if the players are allowed to communicate in a given Bayesian game Γ, as in (1.4). To simplify our analysis, we will henceforth assume that the decision sets D_i, as well as the type sets T_i, are all finite sets.

Let us suppose first that the players communicate with the help of a mediator, who first asks each player to report his type, and who then recommends a strategic action to each player. The mediator's recommendations may depend on the players' reports in a deterministic or random fashion. We let $\mu(d_1,\ldots,d_n|t_1,\ldots,t_n)$ denote the conditional probability that the mediator would recommend to each player i that he should use action d_i, if each player j reported his type to be t_j. Obviously, these numbers $\mu(d|t)$ must satisfy the following probability constraints:

$$(5.1) \qquad \sum_{c\in D} \mu(c|t) = 1 \quad \text{and} \quad \mu(d|t) \geq 0, \ \forall d\in D, \ \forall t\in T.$$

In general, any function $\mu: D \times T \to \mathbb{R}$ that satisfies (5.1) will be called a mechanism (or coordination mechanism) for the Bayesian game Γ.

If every player reports his type honestly and obeys the recommendations of the mediator, then the expected utility for type t_i of player i from mechanism μ would be:

$$(5.2) \qquad U_i(\mu \mid t_i) = \sum_{t_{-i} \in T_{-i}} \sum_{d \in D} p_i(t_{-i} \mid t_i)\, \mu(d \mid t)\, u_i(d,t).$$

We must allow, however, that each player could choose to lie about his type or disobey the mediator's recommendation. That is, we assume that the players' types cannot be verified by the mediator, and each selection of an action d_i in D_i can ultimately be controlled only by player i. Thus, the coordination mechanism μ induces a communication game $\hat{\Gamma}_\mu$ in which each player must select his type report and his plan for choosing an action in D_i as a function of the mediator's recommendation. Formally, $\hat{\Gamma}_\mu$ is itself a Bayesian game, of the form

$$\hat{\Gamma}_\mu = (\hat{D}_1, \ldots, \hat{D}_n, T_1, \ldots, T_n, p_1, \ldots, p_n, \hat{u}_1, \ldots, \hat{u}_n)$$

where

$$\hat{D}_i = \{(s_i, \delta_i) \mid s_i \in T_i \text{ and } \delta_i : D_i \to D_i\}, \quad \text{and}$$

$$\hat{u}_i(((s_1,\delta_1), \ldots, (s_n,\delta_n)), t) =$$

$$= \sum_{t_{-i} \in T_{-i}} \sum_{d \in D} p_i(t_{-i} \mid t_i)\, \mu(d \mid s_1, \ldots, s_n)\, u_i((\delta_1(d_1), \ldots, \delta_n(d_n)), t).$$

A strategy (s_i, δ_i) in $\hat{D}_i$ represents a plan by player i to report s_i to the mediator, and to then choose his action in D_i as a function of the mediator's

recommendation according to δ_i, so that he would do $\delta_i(d_i)$ if the mediator recommended d_i. We assume that each player communicates with the mediator separately and confidentially, so that player i's action cannot depend on the recommendations to the other players.

Suppose, for example, that the true type of player i were t_i, but that he chose to use the strategy (s_i,δ_i) in the communication game $\hat{\Gamma}_\mu$. If all other players were expected to report their types honestly and choose their actions obediently to the mediator, then i's expected utility would be

$$(5.3)\qquad U_i^*(\mu,\delta_i,s_i|t_i) = \sum_{t_{-i}\in T_{-i}} \sum_{d\in D} p_i(t_{-i}|t_i)\,\mu(d|t_{-i},s_i)\,u_i((d_{-i},\delta_i(d_i)),t).$$

(Here $(d_{-i},\delta_i(d_i)) = (d_1,\ldots,d_{i-1},\delta_i(d_i),d_{i+1},\ldots,d_n)$ and $(t_{-i},s_i) = (t_1,\ldots,t_{i-1},s_i,t_{i+1},\ldots,t_n)$.)

Bayesian equilibrium is still the appropriate solution concept for a Bayesian game with communication, except that we must now consider the Bayesian equilibria of the induced communication game $\hat{\Gamma}_\mu$, rather than just the Bayesian equilibria of Γ. We say that a mechanism μ is <u>(Bayesian) incentive compatible</u> iff it is a Bayesian equilibrium for all players to report their types honestly and to obey the mediator's recommendations when he uses the mechanism μ. (Hurwicz [1972] introduced the phrase <u>incentive compatible</u> in a non-Bayesian context, with a somewhat different meaning. Bayesian incentive compatibility was first defined by D'Aspremont and Gerard-Varet [1979]. In this paper, we will always use this term in the Bayesian sense.) Thus, μ is incentive compatible iff

$$(5.4)\qquad U_i(\mu|t_i) \geq U_i^*(\mu,\delta_i,s_i|t_i), \quad \forall i,\ \forall t_i\in T_i,\ \forall s_i\in T_i,\ \forall \delta_i: D_i\to D_i. \quad —$$

If the mediator uses an incentive-compatible mechanism and each player communicates independently and confidentially with the mediator, then no

player could ever gain by being the first one to lie to the mediator or disobey his recommendations. Conversely, we cannot expect all the players to participate honestly and obediently in a coordination mechanism unless it is incentive compatible.

In general, there may be many different Bayesian equilibria of a communication game $\hat{\Gamma}_{\mu}$, even if μ is incentive compatible. Furthermore, we could consider more general classes of coordination mechanisms, in which the messages sent and received by each player i are not necessarily in the sets T_i and D_i. However, for any given coordination mechanism and for any given Bayesian equilibrium of the induced communication game, there exists an equivalent incentive-compatible mechanism, in which every type of every player gets the same expected utility (when all players are honest and obedient) as in the given Bayesian equilibrium of the given mechanism. In this sense, there is no loss of generality in assuming that the players communicate with each other through a mediator who first asks each player to reveal all of his private information (his "type"), and who then gives each player only the minimal information needed to guide his action, in such a way that no player has any incentive to lie or cheat. This result has been observed by many writers independently and it is known as the revelation principle. (See Dasgupta, Hammond, and Maskin [1979], Harris and Townsend [1981], Holmström [1977], Myerson [1979a] and [1982a], Rosenthal [1978], Forges [1982] and, in a non-Bayesian context, Gibbard [1973].)

For any given equilibrium of any given mechanism, the mediator can construct such an equivalent incentive-compatible mechanism as follows. First, he asks each player to (simultaneously and confidentially) reveal his type. Next, the mediator computes what reports would have been sent by the players, with these revealed types, in the given equilibrium. Then, he

computes what recommendations or messages would have been received by the players, as a function of these reports, in the given mechanism. Then, he computes what actions would have been carried out by the players, as a function of these recommendations (and the revealed types) in the given equilibrium. Finally, the mediator tells each player to do the action computed for him in this last step. Thus, the constructed mechanism simulates the given equilibrium of the given mechanism. To check that this constructed mechanism is incentive compatible, notice that any player who could gain by disobeying the mediator in the constructed mechanism could also gain by similarly disobeying his equilibrium strategy in the given mechanism, which is impossible (by definition of equilibrium).

The set of all incentive-compatible mechanisms is a closed convex set, characterized by a system of inequalities (5.1) and (5.4), which are linear in μ. On the other hand, it is generally a difficult problem to characterize the set of all Bayesian equilibria of any given Bayesian game. Thus, by the revelation principle, it may be easier to characterize the set of all Bayesian equilibria of all communication games induced from Γ, than it is to compute the set of Bayesian equilibria of Γ, or of any one communication game $\hat{\Gamma}_{\mu}$. This observation explains why the revelation principle can be so useful.

For example, let us reconsider the game shown in Table 1, in the preceding section. Suppose now that the players can communicate, either directly or through a mediator or through some tatonnement process, before they choose their actions in D_1 and D_2. In the induced communication game, could there ever be a Bayesian equilibrium giving the outcomes (x_1,x_2) if player 2 is type 2a, and (y_1,y_2) if player 2 is type 2b, as naive analysis of the two bimatrices separately might suggest? The answer is No, by the revelation principle. If there were such a communication game, then there

would be an incentive-compatible mechanism achieving the same results. But this would be the mechanism satisfying

$$\mu(x_1,x_2|1,2a) = 1, \quad \mu(y_1,y_2|1,2b) = 1;$$

and it is not incentive compatible, since player 2 could gain by lying about his type. In fact, there is only one incentive-compatible mechanism for this example and this mechanism is μ^*, defined by

$$\mu^*(x_1,x_2|1,2a) = 1, \quad \mu^*(x_1,y_2|1,2b) = 1.$$

Of course, μ^* is equivalent to the unique Bayesian equilibrium of this game without communication.

In general, it maybe possible for all players to increase their expected utility with effective communication. Suppose that there is some given social welfare function which we want to maximize. By the revelation principle, the maximum value that can be achieved by an incentive-compatible mechanism is also the maximum that can be achieved among all Bayesian equilibria of all communication games that can be induced from Γ.

We say that a mechanim μ is <u>incentive-efficient</u> iff μ is incentive compatible and there does not exist any other incentive-compatible mechanism $\hat{\mu}$ such that

$$U_i(\hat{\mu}|t_i) \geq U_i(\mu|t_i), \quad \forall i \in \{1,\ldots,n\}, \ \forall t_i \in T_i, \tag{5.5}$$

with at least one strict inequality. That is, if μ is incentive-efficient, then there is no Bayesian equilibrium of any communication game that gives higher expected utility to some types of some players without giving lower expected utility to at least one type of some player. Conversely, if μ is not incentive-efficient then it is common knowledge that all players would prefer to use some incentive-compatible coordination mechanism. Incentive-efficiency is thus the basic normative concept for welfare analysis of coordination

mechanisms. See Holmström and Myerson [1983] for a more detailed discussion of this concept.

The main goal of the rest of this paper will be to develop more useful conditions for characterizing the incentive-efficient mechanisms. In the next two sections, we will do this for two special cases. First, we will consider Bayesian collective-choice problems, which are situations in which the incentive constraints are purely informational (pure adverse selection problems). Then we will consider games with complete information, in which the incentive constraints are purely strategic (pure moral hazard problems).

6. Bayesian collective-choice problems

A Bayesian collective-choice problem differs from a Bayesian game in that we are given a set of possible outcomes that are jointly feasible for all the players together, rather than a set of strategic decisions or actions for each player separately. That is, a Bayesian collective-choice problem is any Γ^o such that

$$\Gamma^o = (C, T_1,\ldots,T_n, p_1,\ldots,p_n, u_1\ldots,u_n) \tag{6.1}$$

where C is the set of possible outcomes or social choices; each T_i is the set of possible types for player i; each p_i is a function specifying i's conditional probability distribution over T_{-i} for each t_i in T_i; and each u_i is a function specifying i's utility payoff $u_i(c,t)$ for every c in C and every t in $T = T_1\times\ldots\times T_n$.

When we discussed Bayesian games with communication in the preceding section, we assumed that the choice of an action in D_i was inalienably controlled by player i. That is, we assumed that player i could not commit himself to choosing some particular d_i when some other $\hat{d}_i$ in D_i would give him

higher expected utility. (For example, this assumption would be appropriate if D_i were a set of unobservable effort levels that i must choose among when he performs some task, as in a principal-agent problem.) Now, if we assume instead that the players can cooperatively determine their actions in $D_1 \times \ldots \times D_n$ with jointly binding agreements, then the Bayesian game Γ becomes a Bayesian collective-choice problem Γ^o with $C = D_1 \times \ldots \times D_n$.

For another example, to model an exchange economy as a Bayesian collective-choice problem, we could let C be the set of all possible net trades among the players.

In any Bayesian collective-choice problem as in (6.1), the problem is to find efficient or optimal mechanisms for determining the chosen outcome in C as a function of the players' types. We shall assume that C is a finite set, but that random mechanisms are allowed. Thus a <u>mechanism</u> for Γ^o can be defined as any function $\mu : C \times T \to \mathbb{R}$ such that

$$(6.2) \qquad \sum_{e \in C} \mu(e|t) = 1, \ \mu(c|t) \geq 0, \ \forall c \in C, \ \forall t \in T,$$

where $\mu(c|t)$ is interpreted as the probability that c will be the chosen outcome if $t = (t_1, \ldots, t_n)$ is a vector of types reported by n players. As in (5.2) and (5.3), the expected utility for type t_i from mechanism μ if all players report their types honestly is

$$(6.3) \qquad U_i(\mu|t_i) = \sum_{t_{-i} \in T_{-i}} \sum_{c \in C} p_i(t_{-i}|t_i) \, \mu(c|t) \, u_i(c,t);$$

and the expected utility for t_i if he reports s_i while the other players are

honest is

$$(6.4) \qquad U_i^*(\mu, s_i | t_i) = \sum_{t_{-i} \in T_{-i}} \sum_{c \in C} p_i(t_{-i} | t_i)\, \mu(c | t_{-i}, s_i)\, u_i(c,t).$$

The mechanism μ is <u>incentive compatible</u> iff honest reporting by all players is a Bayesian equilibrium of the game induced by μ, that is

$$(6.5) \qquad U_i(\mu | t_i) \geq U_i^*(\mu, s_i | t_i), \ \forall i \in \{1,..,n\}, \ \forall t_i \in T_i, \ \forall s_i \in T_i.$$

These definitions (6.3)-(6.5) are the same as (5.2)-(5.4) except that there is no longer any question of players disobeying recommended actions. In fact, one can easily construct a Bayesian game with n+1 players that is equivalent to the collective choice problem Γ^o, in the sense of generating the same set of incentive-compatible mechanisms. (Let $D_i = \{0\}$ for every i in $\{1,\ldots,n\}$, $D_{n+1} = C$, $T_{n+1} = \{0\}$, and $u_{n+1}(d,t) = 0$ for every d and t.) The revelation principle holds for Bayesian collective-choice problems, just as for Bayesian games with communication. We say that μ is an <u>incentive-efficient</u> mechanism for Γ^o iff μ is incentive compatible and is not dominated by any other incentive-compatible mechanism, in the sense of (5.5).

To simplify our formulas, we will henceforth assume that the players' beliefs are consistent with a common prior p^*, as in (3.1). Furthermore, we will assume that the players' types are independent random variables in the common prior; that is

$$p^*(t) = \prod_{i=1}^{n} p_i^*(t_i), \qquad \forall t \in T,$$

where $p_i^*(t_i)$ is the marginal probability that player i is type t_i. (As was remarked in Section 3, any Bayesian collective-choice problem is probability-

equivalent to another Bayesian collective-choice problem in which beliefs are consistent with such an independent common prior.)

Suppose now that λ and α are vectors of the form

$$\lambda = \left(\left(\lambda_i(t_i)\right)_{t_i \in T_i}\right)_{i=1}^{n}, \qquad \alpha = \left(\left(\alpha_i(s_i|t_i)\right)_{s_i \in T_i,\ t_i \in T_i}\right)_{i=1}^{n}$$

(read "$|$" here as "given") such that

$$\lambda_i(t_i) > 0 \quad \text{and} \quad \alpha_i(s_i|t_i) \geq 0, \quad \forall i \in \{1,\ldots,n\},\ \forall s_i \in T_i,\ \forall t_i \in T_i. \tag{6.6}$$

Then let us define $v_i(c,t,\lambda,\alpha)$ by the following formula:

$$v_i(c,t,\lambda,\alpha) = \Big(\Big(\lambda_i(t_i) + \sum_{s_i \in T_i} \alpha_i(s_i|t_i)\Big)\, u_i(c,t) - \sum_{s_i \in T_i} \alpha_i(t_i|s_i)\, u_i(c,(t_{-i},s_i))\Big)/p_i^*(t_i). \tag{6.7}$$

We shall refer to $v_i(c,t,\lambda,\alpha)$ as player i's <u>virtual utility</u> for outcome c in state t, with respect to the parameters λ and α. This definition (6.7) may seem obscure at first, but it is important because it permits us to state the following characterization of incentive-efficient mechanisms.

<u>Theorem 1</u>: Suppose that μ is an incentive-compatible mechanism for Γ^o. Then μ is incentive-efficient if and only if there exist vectors λ and α satisfying (6.6), such that

$$\alpha_i(s_i|t_i)\big(U_i(\mu|t_i) - U_i^*(\mu,s_i|t_i)\big) = 0, \quad \forall i, \quad \forall s_i \in T_i, \quad \forall t_i \in T_i, \tag{6.8}$$

and

$$\sum_{c \in C} \mu(c|t) \sum_{i=1}^{n} v_i(c,t,\lambda,\alpha) = \max_{c \in C} \sum_{i=1}^{n} v_i(c,t,\lambda,\alpha), \quad \forall t \in T. \tag{6.9}$$

To prove this theorem, observe first that the set of all incentive-compatible mechanisms satisfying (6.2) and (6.5) is a compact convex polyhedron. Thus, by the supporting hyperplane theorem of convex analysis, a mechanism μ is incentive-efficient if and only if there exists a positive vector λ such that μ is an optimal solution to the following problem

$$(6.10) \qquad \underset{\mu}{\text{maximize}} \sum_{i=1}^{n} \sum_{t_i \in T_i} \lambda_i(t_i)\, U_i(\mu|t_i)$$

subject to (6.2) and (6.5).

We interpret $\alpha_i(s_i|t_i)$ as the dual variable or Lagrange multiplier for the incentive constraint (6.5) that asserts that player i should not expect to gain by reporting type s_i if his true type is t_i. Then the Lagrangian function for problem (6.10) is

$$(6.11) \qquad \sum_{i=1}^{n} \sum_{t_i \in T_i} \Big(\lambda_i(t_i) U_i(\mu|t_i) + \sum_{s_i \in T_i} \alpha_i(s_i|t_i)(U_i(\mu|t_i) - U_i^*(\mu, s_i|t_i))\Big)$$

$$= \sum_{t \in T} p^*(t) \sum_{c \in C} \mu(c|t) \sum_{i=1}^{n} v_i(c,t,\lambda,\alpha).$$

That is, the Lagrangian for (6.10) equals the expected sum of the players' virtual utilities. The virtual utility functions were defined in (6.7) precisely so that this equation (6.11) would hold, as may be verified by straightforward algebra. (See Myerson [1982b] and [1982c] for an introduction to the role of virtual utility in the theory of bargaining under incomplete information.)

Condition (6.8) in Theorem 1 asserts that, if the dual variable $\alpha_i(s_i|t_i)$ is positive then the associated incentive constraint must be binding.

Condition (6.9) asserts that $\mu(\cdot|t)$ maximizes the sum of the players' virtual utilities in each state t, subject only to the probability constraint (6.2). Thus, the conditions in Theorem 1 assert that μ and α form a saddlepoint of the Lagrangian function, and so μ must be an optimal solution to (6.10). This completes the proof of Theorem 1.

An incentive-efficient mechanism may be inefficient ex post (after the players have revealed their types) because of the cost of satisfying incentive constraints. However, an incentive-efficient mechanism must maximize ex post the sum of the players' virtual utilities (with respect to some λ and α), so the mechanism will be ex post efficient in terms of these virtual utility functions. Thus, the key to understanding ex post inefficiency in incentive-efficient mechanisms is to understand how virtual utility differs from real utility.

If λ and α satisfy the conditions of Theorem 1 for an incentive-efficient mechanism μ and if $\alpha_i(s_i|t_i) > 0$, then we say that type t_i jeopardizes type s_i in the mechanism μ. That is, t_i jeopardizes s_i iff the constraint that t_i should not be tempted to claim to be s_i is binding and has a positive Lagrange multiplier. Notice that, in (6.7), a player's virtual utility is a positive multiple of his real utility minus a positive linear combination of the utilities of the types that jeopardize his actual type. That is, the virtual utility of a type t_i differs from the real utility in that it exaggerates the difference from the other types that jeopardize t_i. So to understand how the costs of incentive-compatibility should be borne in an incentive-efficient mechanism, we need to recognize which types are jeopardized by which.

There are many situations in which a player's types can be naturally ranked in some order, say from "best" to "worst." In such situations, we can often guess that the better types are jeopardized by the worse types, but not

visa versa, so that the worst type is not jeopardized by any other. In fact, it often happens that each type is jeopardized only by the next-worse type. Optimal auctions in Harris and Raviv [1981] have this structure, where the unjeopardized type of bidder is the one with the highest reservation price.

To illustrate these ideas, suppose that a firm is negotiating with a potential employee, who may either be a "good" type of worker or a "bad" type. We may expect that the bad type jeopardizes the good type; that is, the firm may have difficulty preventing a bad worker from claiming to be good. So the virtual utility of the good type will exaggerate the difference from the bad type. If there is some useless educational process which would be slightly unpleasant for a good worker, but would be intolerably painful for a bad worker, then the good worker may get positive virtual utility from this education, so as to exaggerate his difference from the bad type. As in Spence's [1973] labor-market equilibria, an incentive-efficient mechanism may force a good worker to go through this costly and unproductive educational process (as if he enjoyed it), before he is hired. On the other hand, it seems unlikely that a good worker would be tempted to claim that he is bad in such negotiations. So the bad type of worker is not jeopardized, and the bad type's virtual utility is just a positive multiple of his real utility. Thus, if the worker is bad, the incentive-efficient mechanism should be ex post efficient (in terms of both real and virtual utility), and the bad worker should not suffer through any unproductive educational process.

For another simple example, consider a bargaining problem between one seller of a commodity (player 1) and one buyer (player 2) in a bilateral monopoly situation. The seller has one unit available, and he knows whether it is good quality (type "1a") or bad quality (type "1b"). If it is good quality, then it is worth $40 per unit to the seller and $50 per unit to the

buyer. If it is bad quality, then it is worth \$20 per unit to the seller and \$30 per unit to the buyer. The buyer thinks that the probability of good quality is 0.2.

To formulate this example, we let $T_1 = \{1a,1b\}$, $T_2 = \{2\}$ (so that the variable t_2 can be ignored, since it has only one possible value), and

$$C = \{(x,y) \mid 0 \leq y \leq 1, \quad x \in \mathbb{R}\}$$

Here, for each (x,y) in C, we interpret x as the amount of money paid by the buyer to the seller, and y as the amount of the commodity delivered by the seller to the buyer. The probability and utility functions are:

$$p_1^*(1a) = .2, \quad p_1^*(1b) = .8,$$

$$u_1((x,y),1a) = x - 40y, \quad u_2((x,y),1a) = 50y - x,$$

$$u_1((x,y),1b) = x - 20y, \quad u_2((x,y),1b) = 30y - x.$$

(C is an infinite set in this example, but all of our results will still apply.)

In this example, ex post efficiency would require that the seller should always sell his unit of the commodity, no matter what his type is, since the commodity is always worth \$10 more to the buyer. However, it can be easily shown that there is no incentive-compatible mechanism that is ex post efficient and gives nonnegative expected utility to the buyer and to both types of the seller (i.e., such that $U_2(\mu) \geq 0$, $U_1(\mu|1a) \geq 0$, $U_1(\mu|1b) \geq 0$).

For this example, let λ and α be

$$\lambda_1(1a) = .3, \quad \lambda_1(1b) = .7, \quad \lambda_2 = 1, \quad \alpha_1(1a|1b) = .1, \quad \alpha_1(1b|1a) = 0.$$

By (6.7), the virtual utility functions for these parameters are

$$v_1((x,y),1a,\lambda,\alpha) = \big(.3u_1((x,y),1a) - .1\ u_1((x,y),1b)\big)/.2 = x - 50y,$$

$$v_2((x,y),1a,\lambda,\alpha) = u_2((x,y),1a) = 50y - x,$$

$$v_1((x,y),1b,\lambda,\alpha) = (.7+.1)\ u_1((x,y),1b)/.8 \ = x - 20y,$$

$$v_2((x,y),1b,\lambda,\alpha) = u_2((x,y),1b) = 30y - x.$$

That is, the bad type of seller (1b) jeopardizes the good type (1a), so the good type's virtual utility exaggerates the difference from the bad type and has a virtual reservation price of \$50 (instead of \$40) for the commodity. With this λ and α, virtual ex post efficiency would require only that all of the commodity must be sold to the buyer if it is of bad quality; there are no virtual gains from trade between the buyer and a good-type seller. Thus, this λ and α will satisfy the conditions of Theorem 1 for any mechanism μ such that all the commodity is sold to the buyer if the seller's type is 1b, and the constraint that the 1b-type seller should not claim to be "1a" is binding. For example, consider μ such that

$$\mu(30,1 \mid 1b) = 1, \quad \mu(50/3,\ 1/3 \mid 1a) = 1$$

(that is, the bad type sells all of his commodity for \$30, and the good type sells 1/3 of his supply at a price of \$50 per unit). This mechanism satisfies both of the above conditions (check that $U_1(\mu \mid 1b) = U_1^*(\mu,1a \mid 1b) = 10$), and so it is incentive-efficient, even though the seller cannot sell two-thirds of his commodity if it is good.

7. Correlated equilibria of games with complete information

If the players have no private information (so that each has only one possible type) then the Bayesian game reduces to a game in strategic form (or

normal form) with complete information. That is, we get

$$\Gamma = (D_1,\ldots,D_n,\ u_1,\ldots,u_n)\ , \tag{7.1}$$

where each $u_i(\cdot)$ is a function from $D = D_1\times\ldots\times D_n$ into $\mathbb{R}$. For such games, we can derive a characterization of incentive-efficient mechanisms closely analogous to Theorem 1.

For a game with complete information, a coordination mechanism μ is just a probability distribution over D, satisfying

$$\sum_{e\in D} \mu(e) = 1 \quad \text{and} \quad \mu(d) \geq 0, \quad \forall d\in D. \tag{7.2}$$

(There are no longer any alternative types for the mechanism to depend on.) The condition of incentive-compatibility (5.4) reduces to

$$\sum_{d_{-i}\in D_{-i}} \mu(d)\, u_i(d) \geq \sum_{d_{-i}\in D_{-i}} \mu(d)\, u_i(d_{-i},e_i), \quad \forall i, \quad \forall d_i\in D_i, \quad \forall e_i\in D_i. \tag{7.3}$$

(Here $D_{-i} = D_1\times\ldots\times D_{i-1}\times D_{i+1}\times\ldots\times D_n$ and $d = (d_{-i},d_i)$.)
To interpret this condition, suppose that a mediator randomly selected a joint action in D, selecting d with probability $\mu(d)$, and each player i was then informed only as to which action d_i was his component of the mediator's selection. Then (7.3) asserts that each player's optimal action is to do what the mediator has told him, if all other players are also expected to obey the mediator's recommendation. (To see this, first divide both sides of (7.3) by the marginal probability of d_i being selected, that is,

$$\sum_{d_{-i}\in D_{-i}} \mu(d).$$

Then the left-hand side and right-hand side are player i's conditionally expected utility from using d_i and e_i respectively, given that the mediator

mended d_i.)

Conditions (7.2) and (7.3) are also the definition of a correlated ibrium, due to Aumann [1974]. Thus, the concept of an incentive- itible mechanism is just a generalization of Aumann's concept of lated equilibrium, and the two concepts coincide for games with complete mation.

In this context, a mechanism μ is incentive-efficient if and only if exists a vector $\lambda = (\lambda_1,\ldots,\lambda_n$ such that every $\lambda_i > 0$ and μ is an al solution to the following problem

$$\underset{\mu}{\text{maximize}} \sum_{i=1}^{n} \sum_{d\in D} \lambda_i \mu(d) \quad \text{i}(d)$$

subject to (7.2) and (7.3 .

ollowing theorem, analogous to Theorem 1, is derived by a standard ngian analysis of (7.4), lettin $\beta_i(e_i|d_i)$ denote the Lagrange multiplier he constraint (7.3) that says t at player i should not be tempted to do en told to do d_i.

Theorem 2: Suppose that μ is a correlated equilibrium. Then μ is tive-efficient if and only if t ere exist vectors λ and β such that

$$\lambda_i > 0 \quad \text{and} \quad \beta_i(e_i|d_i) \geq , \quad \forall i, \forall d_i\in D_i, \forall e_i\in D_i;$$

$$\beta_i(e_i|d_i)\Big(\sum_{d_{-i}\in D_{-i}} \mu(d)(u_i \quad d) - u_i(d_{-i},e_i))\Big) = 0, \quad \forall i, \forall d_i\in D_i, \forall e_i\in D_i;$$

$$\sum_{d\in D} \mu(d) \sum_{i=1}^{n} v_i(d,\lambda,\beta) = \max_{d\in} \sum_{i=1}^{n} v_i(d,\lambda,\beta);$$

we define the virtual utility functions $v_i(\cdot)$ by

$$(7.8) \qquad v_i(d,\lambda,\beta) = \lambda_i u_i(d) + \sum_{e_i \in D_i} \beta_i(e_i|d_i)(u_i(d) - u_i(d_{-i},e_i)).$$

Here condition (7.6) asserts that if $\beta_i(e_i|d_i) > 0$ then the constra that "i should not gain by doing e_i when told to do d_i" is binding. Cor (7.7) asserts that μ puts all probability-weight on the joint actions th maximize the sum of the players' virtual utilities.

If $\beta_i(e_i|d_i) > 0$ then we may say that action e_i <u>jeopardizes</u> action player i. Then i's virtual utility $v_i(d,\lambda,\beta)$ is a positive multiple of real utility $u_i(d)$ minus a positive linear combination of what he would he changed to some other action that jeopardizes d_i. Thus, player i's utility when he does d_i differs from his real utility in that it exagge the difference from what he would get from other actions that jeopardiz

To understand these results, let us consider an example based on o Aumann [1974]. There are two players, $D_1 = \{x_1,y_1\}$, $D_2 = \{x_2,y_2\}$, and utility payoffs (u_1,u_2) are as in the following table:

	x_2	y_2
x_1	5,1	0,0
y_1	4,4	1,5

TABLE 2

There are three Nash equilibria of this game: (x_1,x_2), (y_1,y_2), and a randomized Nash equilibrium in which each player gives equal probabilit his two strategies. In the randomized equilibrium, all four outcomes h

equal probability, and each player gets expected utility 2.5.

The best symmetric payoff in this example is (4,4), but the players cannot achieve this because (y_1,x_2) is not an equilibrium. Player 1 would choose x_1 if he expected player 2 to choose x_2. However, with communication, the players can make self-enforcing plans of action that give them both higher expected utility than 2.5. For example, they could agree to toss a coin and then choose (x_1,x_2) if it is heads and (y_1,y_2) if it is tails. This plan of action is self enforcing, even though the coin-toss has no binding impact on the players. (Player 1 could not gain by choosing x_1 after tails, since player 2 is then expected to choose y_2.) Thus, this plan is a correlated equilibrium, and it gives each player an expected utility of 3.

With the help of a mediator, the players can achieve even higher expected utility in a correlated equilibrium. Suppose that the mediator randomizes among outcomes according to μ, where

$$\mu(x_1,x_2) = \mu(y_1,x_2) = \mu(y_1,y_2) = \frac{1}{3}, \quad \mu(x_1,y_2) = 0.$$

When the mediator tells each player separately which of his actions was in the randomly selected pair, then it is self-enforcing for both players to use the action designated by the mediator. For example, if player 1 is told "y_1", then he thinks that it is equally likely that player 2 has been told "x_2" or "y_2"; so y_1 would be as good as x_1 for player 1 (both give expected utility 2.5) if he expects that player 2 will also do as he is told. Thus, μ is a correlated equilibrium, and it gives each player an overall expected utility of 3.33.

In fact, this mechanism μ is incentive-efficient, so that (3.33, 3.33) is the highest symmetric expected-utility allocation that the players can achieve in any correlated equilibrium. To check that μ is incentive-efficient, let

$$\lambda_1 = \lambda_2 = 1, \quad \beta_1(x_1|y_1) = \beta_2(y_2|x_2) = \frac{2}{3}, \quad \beta_1(y_1|x_1) = \beta_2(x_2|y_2) = 0.$$

Then the virtual utility functions (v_1, v_2) are

	x_2	y_2
x_1	5.00, 1.66	0, 0
y_1	3.33, 3.33	1.66, 5.00

TABLE 3

and μ puts all weight on the outcomes that maximize $v_1 + v_2$. Furthermore, as required by (7.6), μ satisfies without slack the two incentive constraints that have positive Lagrange multipliers.

8. General conditions for incentive-efficiency

In Myerson [1982a], a class of Bayesian incentive problems were defined in a way which includes strategic-form games, Bayesian collective-choice problems, and Bayesian games, all as special cases. Formally, a Bayesian incentive problem is any Γ of the form

$$(8.1) \qquad \Gamma = (D_0, D_1, \ldots, D_n, T_1, \ldots, T_n, p_1, \ldots, p_n, u_1, \ldots, u_n)$$

where D_0 is a set of enforceable or public actions, and, for each i in $\{1, \ldots, n\}$, D_i, T_i, p_i, and u_i are as in a Bayesian game, except that now the domain of the utility function u_i is

$$D \times T = (D_0 \times D_1 \times \ldots \times D_n) \times (T_1 \times \ldots \times T_n).$$

That is, a general Bayesian incentive problem differs from a Bayesian game in that there may be some publicly controllable actions, as well as the privately controlled actions in $D_1, D_2, \ldots, D_n$. For example, suppose that the players

are managers of different divisions in a firm. For each player i, his type in T_i may represent his private information about the production function in his division, and his private action in D_i may be his level of effort in carrying out his management responsibilities. The public actions in D_0 may be specifications of how the firm's capital resources are to be allocated to the divisions, and how each manager is to be paid as a function of output.

In general, all decision variables that the players can control cooperatively, or about which they can make binding promises, should be components of the "public actions" in D_0. Any decision variables that player i controls inalienably, or about which he cannot make any promises that conflict with his own utility-maximizing behavior, must be components of the "private actions" in D_i.

Thus, a Bayesian collective-choice problem is just a Bayesian incentive problem in which each player i has only one possible private action ("doing nothing"), so that $|D_i| = 1$ and the variable d_i can be ignored. On the other hand, a Bayesian game is just a Bayesian incentive problem in which $|D_0| = 1$. (Actually, any Bayesian incentive problem could be reduced to a Bayesian game, by introducing an $n+1^{th}$ player "0" who controls the action in D_0 as his private action, has no private information, and has $u_0(d,t) = 0$ for all d and t.)

The definitions of incentive-compatible and incentive-efficient mechanisms for a Bayesian incentive problem are the same as for a Bayesian game, except that now in equations (5.1)-(5.4) we let

$D = D_0 \times D_1 \times \ldots \times D_n$, $d = (d_0,d_1,\ldots,d_n)$, and

$(d_{-i},\delta_i(d_i)) = (d_0,d_1,\ldots,d_{i-1},\delta_i(d_i),d_{i+1},\ldots,d_n)$.

The following theorem generalizes Theorems 1 and 2 to the general case of the Bayesian incentive problem. We assume that D and T are finite sets, and

that the players' beliefs are consistent with an independent common prior p^*, as in Theorem 1.

Theorem 3: Suppose that μ is an incentive-compatible mechanism for the Bayesian incentive problem Γ, as above. Then μ is incentive-efficient if and only if there exist vectors λ, α, and γ such that

$$(8.2)\qquad \lambda_i(t_i) > 0, \quad \alpha_i(s_i|t_i) \geq 0, \quad \gamma_i(e_i|d_i,s_i,t_i) \geq 0$$

$$\forall i \in \{1,\ldots,n\},\ \forall t_i \in T_i,\ \forall s_i \in T_i,\ \forall d_i \in D_i,\ \forall e_i \in D_i;$$

$$(8.3)\qquad \sum_{e_i \in D_i} \gamma_i(e_i|d_i,s_i,t_i) = 1, \quad \forall i,\ \forall d_i \in D_i,\ \forall s_i \in T_i,\ \forall t_i \in T_i;$$

$$(8.4)\qquad \sum_{t_{-i}} \sum_{d} \sum_{e_i} p_i(t_{-i}|t_i)\, \mu(d|t_{-i},s_i)\, \gamma_i(e_i|d_i,s_i,t_i)\, u_i((d_{-i},e_i),t)$$

$$= \max_{\delta_i : D_i \to D_i} U_i^*(\mu,\delta_i,s_i|t_i), \qquad \forall i,\ \forall s_i \in T_i,\ \forall t_i \in T_i;$$

$$(8.5)\qquad 0 = \alpha_i(s_i|t_i)\,\big(U_i(\mu|t_i) - \max_{\delta_i} U_i^*(\mu,\delta_i,s_i|t_i)\big),$$

$$\forall i,\ \forall s_i \in T_i,\ \forall t_i \in T_i; \qquad \text{and}$$

$$(8.6)\qquad \sum_{d \in D} \mu(d|t) \sum_{i=1}^{n} v_i(d,t,\lambda,\alpha,\gamma) = \max_{d \in D} \sum_{i=1}^{n} v_i(d,t,\lambda,\alpha,\gamma), \qquad \forall t \in T,$$

where

$$(8.7)\qquad v_i(d,t,\lambda,\alpha,\gamma) = \Big(\big(\lambda_i(t_i) + \sum_{s_i} \alpha_i(s_i|t_i)\big)u_i(d,t)$$

$$- \sum_{s_i} \alpha_i(t_i|s_i) \sum_{e_i} \gamma_i(e_i|d_i,t_i,s_i)\, u_i((d_{-i},e_i),(t_{-i},s_i))\Big)/p_i^*(t_i).$$

As before, these conditions are derived from the Lagrangian conditions for an optimization problem, to maximize

$$\sum_i \sum_{t_i \in T_i} \lambda_i(t_i)\, U_i(\mu|t_i)$$

subject to the probability constraints (5.1) and the incentive constraints (5.4). Let $\alpha_i^o(\delta_i, s_i|t_i)$ denote the Lagrange multiplier of the incentive constraint that type t_i of player i should not be tempted to claim that he is type s_i and to then disobey his recommended action according to $\delta_i(\cdot)$. Let us choose α and γ so that they satisfy (8.2), (8.3),

$$(8.8) \qquad \alpha_i(s_i|t_i) = \sum_{\delta_i : D_i \to D_i} \alpha_i^o(\delta_i, s_i|t_i), \quad \forall i, \quad \forall s_i \in T_i, \quad \forall t_i \in T_i,$$

and

$$(8.9) \qquad \gamma_i(e_i|d_i, s_i, t_i)\, \alpha_i(s_i|t_i) = \sum_{\{\delta_i | \delta_i(d_i) = e_i\}} \alpha_i^o(\delta_i, s_i|t_i),$$

$$\forall i, \ \forall d_i \in D_i, \ \forall e_i \in D_i, \ \forall t_i \in T_i, \ \forall s_i \in T_i.$$

(If $\alpha_i(s_i|t_i) = 0$ then we can choose $\gamma(\cdot|\cdot, s_i, t_i)$ so that it also satisfies (8.4).) Then the Lagrangian function of this optimization problem can be written

$$\sum_i \sum_{t_i} \lambda_i(t_i)\, U_i(\mu|t_i) + \sum_i \sum_{t_i} \sum_{s_i} \sum_{\delta_i} \alpha_i^o(\delta_i, s_i|t_i)\big(U_i(\mu|t_i) - U_i^*(\mu, \delta_i, s_i|t_i)\big)$$

$$= \sum_t p^*(t) \sum_d \mu(d|t) \sum_{i=1}^n v_i(d, t, \lambda, \alpha, \gamma).$$

The conditions of Theorem 3 follow directly from this equation and the saddlepoint conditions of Lagrangian analysis.

To interpret the conditions in Theorem 3, think of $\gamma_i(e_i|d_i, s_i, t_i)$ as the

probability that player i would choose action e_i if he were cheating when his type was t_i, he reported s_i, and he was told to do d_i. Condition (8.4) asserts that using $\gamma_i(\cdot|\cdot,s_i,t_i)$ should be an optimal plan for type t_i after reporting s_i. Condition (8.5) asserts that $\alpha_i(s_i|t_i)$ can be positive only if type t_i would be willing to report type s_i. (We may have $\alpha_i(t_i|t_i) > 0$, if type t_i would be willing to disobey the mediator's recommended actions after reporting honestly.) Formula (8.7) extends the virtual utility formula (6.7) for Bayesian collective-choice problems. The virtual utility of type t_i differs from the real utility in that it exaggerates the difference from the types that jeopardize t_i, when they use their optimal disobedience plans γ_i. By (8.6) an incentive-efficient mechanism must maximize the sum of the players' virtual utilities, in every state t. Thus, the conditions of Theorem 3 can give us some intuition as to the qualitative nature of incentive-efficient mechanisms, even though these conditions may be too complex to apply numerically in many problems.

Mertens, J.-F., and S. Zamir [1982], "Formalization of Harsanyi's Notions of 'Type' and 'Consistency' in Games with Incomplete Information," C.O.R.E. discussion paper, Université Catholique de Louvain.

Milgrom, P. R., and R. J. Weber [1981], "Distributional Strategies for Games with Incomplete Information," Center for Math Studies D.P. #427, Norhtwestern University, to appear in Mathematics of Operations Research.

Myerson, R. B. [1979a], "Incentive Compatibility and the Bargaining Problem," Econometrica 47, 61-73.

Myerson, R. B. [1979b], "An Axiomatic Derivation of Subjective Probability, Utility and Evaluation Functions," Theory and Decision 11, 339-352.

Myerson, R. B. [1982a], "Optimal Coordination Mechanisms in Generalized Principal-Agent Problems," Journal of Mathematical Economics 10, 67-81.

Myerson, R. B. [1982b], "Two-Person Bargaining Problems with Incomplete Information," Center for Math Studies D.P. #527, Northwestern University, to appear in Econometrica.

Myerson, R. B. [1982c], "Cooperative Games with Incomplete Information," Center for Math Studies D.P. #528, Northwestern University,to appear in International Journal of Game Theory.

Raiffa, H. [1968], Decision Analysis, Reading, Mass.: Addison-Wesley.

Rosenthal, R. W. [1978], "Arbitration of Two-Party Disputes under Uncertainty," Review of Economic Studies 45, 595-604.

Savage, L. J. [1954], The Foundations of Statistics, New York: Wiley.

Spence, M. [1973], "Job Market Signaling," Quarterly Journal of Economics 87, 355-374.

Von Neumann, J., and O. Morgenstern [1944], Theory of Games and Economic Behavior, Princeton: Princeton University Press.

Wilson, R. [1968], "The Theory of Syndicates", Econometrica 38, 119-132.

References

Armbruster, W., and W. Böge [1979], "Bayesian Game Theory," in Game Theory and Related Topics, ed. by O. Moeschlin and D. Pallaschke, Amsterdam: North-Holland, 17-28.

Arrow, K. J. [1970], Essays in the Theory of Risk-Bearing, Amsterdam: North-Holland.

Aumann, R. J. [1974], "Subjectivity and Correlation in Randomized Strategies," Journal of Mathematical Economics 1, 67-96.

Aumann, R. J. [1976], "Agreeing to Disagree," Annals of Statistics 4, 1236-1239.

Billingsley, P. [1968], Convergence of Probability Measures, New York: Wiley.

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

D'Aspremont, C. and L.-A. Gerard-Varet [1979], "Incentives and Incomplete Information," Journal of Public Economics 11, 25-45.

Forges, F. [1982], "A First Study of Correlated Equilibria in Repeated Games with Incomplete Information," CORE discussion paper #8218, Université Catholique de Louvain.

Gibbard, A. [1973], "Manipulation of Voting Schemes: A General Result," Econometrica 41, 587-602.

Harris, M., and A. Raviv [1981], "Allocation Mechanisms and the Design of Auctions," Econometrica 49, 1477-1499.

Harris, M. and R. M. Townsend [1981], "Resource Allocation under Asymetric Information," Econometrica 49, 231-259.

Harsanyi, J. C. [1967-8], "Games with Incomplete Information Played by 'Bayesian' Players," Management Science 14, 159-189, 320-334, 486-502.

Holmström, B. [1977], "On Incentives and Control in Organizations," Ph.D. dissertation, Stanford University.

Holmström, B., and R. B. Myerson [1983], "Efficient and Durable Decision Rules with Incomplete Information," Econometrica 53, 1799-1819.

Hurwicz, L. [1972], "On Informationally Decentralized Systems," in Decision and Organization, ed. by R. Radner and B. McGuire, Amsterdam: North-Holland, 297-336.