Kalai, Ehud; Samet, Dov

Working Paper

Are Bayesian-Nash Incentives and Implementations Perfect?

Discussion Paper, No. 680

Provided in Cooperation with:
Kellogg School of Management - Center for Mathematical Studies in Economics and Management Science, Northwestern University

Suggested Citation: Kalai, Ehud; Samet, Dov (1986) : Are Bayesian-Nash Incentives and Implementations Perfect?, Discussion Paper, No. 680, Northwestern University, Kellogg School of Management, Center for Mathematical Studies in Economics and Management Science, Evanston, IL

This Version is available at:
https://hdl.handle.net/10419/221039

Discussion Paper No. 680

ARE BAYESIAN-NASH INCENTIVES AND IMPLEMENTATIONS PERFECT?[†]

by

Ehud Kalai[*]
and
Dov Samet[**]

May 1986

ABSTRACT: A revelation (or delegation) game is Nash incentives compatible if and only if it is perfectly incentives compatible whenever the types of the players are personal (independently drawn with private valuations of actions). An outcome in a personal-types Bayesian environment is Nash implementable, if and only if it is perfectly implementable. These observations are shown to be immediate consequences of a general decomposition theorem regarding the perfect equilibria of Bayesian games.

[†]The authors wish to thank Roger Myerson and Morton Kamien for helpful conversations. This research was supported by a grant from the National Science foundation NSF No. SES-8409798.

[*]Department of Managerial Economics and Decision Sciences, Northwestern University, Evanston, Illinois 60201.

[**]Department of Managerial Economics and Decision Sciences, Northwestern University, Evanston, Illinois 60201 and Department of Economics, Bar-Ilan University, Ramat Gan, Israel.

ARE BAYESIAN-NASH INCENTIVES AND IMPLENTATIONS PERFECT?

by Ehud Kalai and Dov Samet

1. Introduction

Economists have come to use game theoretic models and solutions in order to analyze problems involving information and incentives. Gibbard [1973] and Satterthwaite [1975] attempted to use dominated strategy incentive compatible revelation games in order to deal with the information and incentives issues arising in Arrow's [1951] social choice problem. As they and many other authors discovered, this dominant strategy approach was too demanding and failed to bring about meaningful analysis. Given the failure of this approach, Hurwicz [1972], Peleg [1977], Groves-Ledyard [1977], Hurwicz-Schmiedler [1978], Kalai-Rosenthal [1978], and many others (see Maskin [1985], Postelwaite [1985], and Roberts [1986] for general studies and references) have replaced the solution concept of dominant strategy by the weaker notion of Nash equilibrium and applied it to general classes of strategic economic games. Their approach paralleled the game theoretic development of the notion of correlated equilibrium due to Aumann [1974] and has proven to be more successful.

More recently as researchers wanted to model asymmetries of information in a more explicit manner, they recognized that many economic situations can be viewed as the games of incomplete information discussed by Harsanyi [1967-8]. Consequently, we have seen many economic and managerial problems being studied and analyzed as Bayesian games of incomplete information. A partial list of studies of this type includes Wilson [1967], d'Aspermont-Gerard-Varet [1979], Milgrom-Weber [1982], Myerson [1981], Matthews [1979], Milgrom and Roberts [1982], Harris-Raviv [1981], Maskin-Riley [1986], Riley-Samuelson

[1981], Crawford-Sobel [1982], Myerson and Satterthwaite [1976], Postlewaite-Schmiedler [1986], Palfrey-Srivastava [1986], and Ledyard [1986]. For a comprehensive survey and study of Bayesian games, implementations, and incentives, we refer the reader to Myerson [1985].

An active direction of research in game theory deals with the weaknesses of the Nash equilibrium concept. Selten [1975] has argued that this concept may predict outcomes which are not likely to occur and proposed a strengthening of it which he called perfect (Nash) equilibrium. Following Selten [1975], Harsanyi and Selten [1980], Myerson [1978], Kreps and Wilson [1982], Kalai and Samet [1984], Mertens and Kohlberg [1982], Bernheim [1984], Pearce [1984], McLennan [1985], and others have studied a variety of these types of modifications. We refer the reader to Van Damme [1983] for a comprehensive study and a survey of this literature. Recently Forges [1984] and Myerson [1984] have been developing these concpets further in order to adapt them better for incentive problems involving communications and other considerations of this type. However, a major portion of the literature on incentives uses the unmodified notion of Nash (i.e., Bayesian) equilibrium, and thus suffers from the weaknesses discussed by Selten. To illustrate an example of such a weakness and for further discussion we introduce here the following hypothetical game.

The Confession Game. Two players might have jointly committed an illegal act. No one but the two of them knows whether they did it. Society would like to implement the following outcomes. If they did commit the crime then they should each be penalized (say impose an $M fine on each of them), but otherwise no penalty should be imposed. It is suggested that this conditional outcome be implemented by the following game. A representative of the society will ask each one of them, separately, whether he has participated in the

illegal act. If any one of them admits to it then the penalty will be imposed on both of them. However, if they both deny the accusation then they will be let go without penalty.

It is argued that the game suggested above is "Nash incentives compatible" and that the players will therefore reveal the truth. The underlying logic is that every player will reason to himself that if his opponent tells the truth then he cannot gain anything by lying, and thus telling the truth does not contradict his incentives.

Obviously it would be naive to assume that the scheme proposed above is satisfactory. While it is possible that a criminal player would not lose by telling the truth it is very likely that he would. If there is any possibility that his criminal partner would lie then he would also prefer to lie (we assume here that there is no inherent satisfaction from telling the truth). Lying is a dominant strategy for a criminal player in this game. When we apply Selten's perfect equilibrium notion to the analysis of this game it selects for us what we consider to be the more reasonable prediction, that they would deny having committed the crime under such a scheme.

A main purpose of this paper is to report general circumstances under which incentives and implementations are perfect in the sense of Selten. In the process of doing that we introduce a general decomposition property of Bayesian games and exhibit a structural theorem relating perfect equilibrium to this decomposition. The perfect equilibrium notion that we use is sometimes referred to as "trembling hand" perfection as opposed to "subgame perfection." In other words, we use Selten's original terminology. Perfection of an equilibrium is a stronger requirement than subgame perfection. For example, in the confession game discussed earlier, the "bad" equilibrium is still subgame perfect. The same relationship holds with Kreps

and Wilson's notion of sequentiality.

For a strategy to be a Nash equilibrium it is required that simultaneously every player's equilibrium action be an optimal reaction to his opponent's equilibrium actions. For a strategy to be a perfect equilibrium it is further required that every player's equilibrium action be an optimal reaction to some interior strategies of his opponents which are arbitrarily close to their equilibrium actions. As usual, a strategy of a player is called interior if there is a positive (however small) probability that he would take every action available to him at every state of his information. This can be viewed as a minimal continuity or robustness requirement on the proposed equilibrium concept. An important implication of this condition is that optimality is tested not only at the equilibrium point, but also at some interior points in every neighborhood of it. This is important because it rules out knife-edge strategic stability which is based upon players assigning certainty (probability one) to their conjectures on the actions of their opponents. Under such idealized certainty the players can get locked in unbelievable behavior which is really not credible if any small doubt enters their considerations. Consider, for instance, the confession game discussed earlier. True revelation is optimal to a criminal player if he is certain that his opponent reveals truthfully. However, if there is any small positive probability that his criminal opponent does not reveal truthfully then his only optimal reaction is to lie. Thus the true revelation could not be a perfect equilibrium behavior of the criminal player.

In the next section we formally introduce the notions of Bayesian game, Nash equilibrium, and perfect (Nash) equilibrium. We basically follow the notations and conventions as in Myerson [1985] for the Bayesian games and their Nash equilibria. For the prefect equilibrium concept we use the

definition described in Kalai-Samet [1985]. It is easy to verify that this definition is equivalent to the original one given in Selten [1975].

The Bayesian games studied here may be thought of as n person two-stage extensive form games. In the first stage nature draws a type for every one of the n players from a commonly known probability distribution on the finite set consisting of all type combinations (n-tuples of types). After the draw every player is informed about the realization of his type and only his type. In the second stage, simultaneously, every player chooses an action from a finite set of actions available to him. The set of actions available to a player is the same regardless of the particular type realization. At this point the game is over and every player is rewarded a von Neumann-Morgenstern utility level which is determined by the realized type and action combinations.

The third section of the paper is devoted to the study of special families of Bayesian games and the relationship between their Nash and perfect equilibria. The family of revelation games consists of the ones in which the set of actions available to every player consists of his set of types. Thus in such a game we may think of an action of a player as being a declaration (possibly a false one) of the type that he is. For example, the confession game discussed previously is such a game. A revelation game is called incentive compatible if declaring one's true type by every one of the players turns out to be a Nash equilibrium. The revelation games have special importance in the incentives literature because of the "revelation principle." This useful principle states that to every Bayesian game with a choice of a Nash equilibrium outcome, we can associate a derived incentive-compatible revelation game in which the truthful playing yields the same outcome. Thus in classifying the set of outcomes that may result from the Nash equilibria of a given family of games, it suffices to restrict our

attention to the set of outcomes that result from truthful play in the derived incentive-compatible revelation games.

A somewhat broader family of games consists of what we call delegation games. In a delegation game every player is prescribed an action for every one of his type realizations. We refer to these prescriptions as guidelines. We say that a delegation game (and its guidelines) are incentive compatible if the total obedience strategy by all the players turns out to be a Nash equilibrium. Mathematically these games turn out to be very similar to the revelation games. Indeed formally every revelation game can be viewed as a delegation game (with the guidelines being for every player to reveal his true type). Because of this fact the "revelation principle" induces immediately a "delegation principle." This latter principle says that to every Bayesian game with a choice of Nash equilibrium outcome, we can associate a derived incentive-compatible delegation game in which obedient playing yields the same outcome.

Because of the importance of revelation and delegation games we are interested in knowing when such incentive-compatible games are perfectly incentive compatible. By perfect incentive compatibility we mean that the truthful or the obedient strategies are perfect, and not just Nash, equilibria.

As we report in Section 3, for a significant class of revelation and delegation games, being incentive compatible is equivalent to being perfectly incentive compatible. This class is characterized by the property that the types of each player are what we call personal. For a given Bayesian game we say that it has personal types if two conditions hold. The first condition is that the types of the various players are drawn independently (in the probabilistic sense). The second condition is that the individual player

payoffs are private. This means that a player's payoff depends only on his type and the action combination of the group of all players. However, given his own fixed type and given a fixed group action combination his payoff will not be affected if we vary the types of his opponents (provided that their actions are not changed). The confession game discussed earlier violates the independent type assumption (the types there--criminal or not--are completely correlated) but satisfies the independent payoff condition.

As it turns out, many Bayesian games satisfy the two conditions just discussed. For example, most of the Bayesian game papers cited earlier make use of these conditions. The two main results of section 3 state that if an incentive-compatible revelation or delegation game has only personal types then it must be perfectly incentive compatible.

In section 4 we discuss the relationships between Nash implementability and perfect implementability of an outcome function. We define the notion of an n-person Bayesian environment to consist of four components. The first component describes again the finite sets of possible type realizations for every one of the n players. The second component is a commonly known probability distribution on the set consisting of all possible type combinations. The third component is a finite set describing all the possible outcomes in the environment. The fourth component of the Bayesian environment consists of a von Neumann-Morgenstern utility function for every one of the players. We assume that a player's utility depends on the choice of an outcome in the environment and on the type combination of the n players. In such an environment it is assumed that after the initial drawing of a type combination, every player is informed of his own particular type. Based on this information, and his updated probabilistic belief regarding the type combination of his opponents, a player has an ex post preference over the

possible outcomes in the environment. As in the case of Bayesian games we say that the environment consists of personal types if the types are independently drawn and every player's utility depends on the outcome and only on his own type.

An implementor in a Bayesian environment wishes to bring about an outcome for every combination of types. The implementor may be a public agent with an interest in maximizing some social welfare or he may be a private agent pursuing his own private interests. For example, a public agent may want to bring about a Pareto optimal outcome for every combination of type realization. A private implementor may be a seller selling an object to the n players, and his interest may be to bring about the outcome that will maximize the total revenues that he can receive from every configuration of buyer types. Generally, what the implementor wishes to implement, is an outcome function, which is a function assigning a probability distribution over the set of outcomes for every type combination. However, the implementor may face two difficulties. First, becuase of their own strategic considerations, the players may not reveal their true types to the implementor. The second difficulty may come up in situations where in order to bring about a certain otucome the players themselves have to cooperate by taking some necessary actions, and the implementor does not have the authority to control their actions. In order to have a unified model that covers all the different types of implementors we take the following approach.

We define a Bayesian game in a given Bayesian environment as a pair consisting of two components. The first component has a set of actions available to every one of the n players. The second component is a result function which assigns a probability distribution over the set of the environmental outcomes to every action combination of the players. But the

utility functions that the players have over the environment's outcomes induces a utility function over the players' actions. Thus every Bayesian game in the given Bayesian environment induces a regular Bayesian game in one natural way. In this way the notions of Nash and perfect equilibrium get extended to games in the environment.

We assume that an implementor in a given Bayesian environment is described by a family of environmental games from which he has the authority to select one. This family describes the amount of control that he has. For example, if the family of games available to him is a singleton, then he has no real control and the players will play this one given game. On the other extreme, a social implementor, for example a legislator, can have much control in the sense that his family of games may be large and he can choose the "rules of the social game" from many available ones. Interesting implementors are intermediary ones with partial control. For example, a seller wishing to sell an object can consider many different auction methods. Each choice of an auction method would mean one choice of a game from his available set. It would be reasonable to assume that if he is a private seller then his family contains all the legal auction methods provided that they include the action "do not participate" in the feasible action set of every one of the players (buyers).

We say that an outcome function is Nash implementable in a given Bayesian environment by a given implementor if his family of implementing games includes a game with a Nash equilibrium that yields this outcome function. Similarly, an outcome function is perfectly implementable if the implementing Nash equilibrium is perfect.

The main result of section 4 is in illustrating that in a Bayesian environment with personal type, an implementor can Nash implement a given

outcome if and only if he can perfectly implement it. This equivalence is illustrated under the assumptions that the family of implementing games is complete (as defined there), and is a mathematical corollary of the results in the previous section.

In section 5 we discuss briefly correlated equilibria which may be viewed as equilibria in Bayesian games. Our results for general Bayesian games show that those correlated equilibria which correspond to totally mixed strategy equilibria of the game in normal form are perfect. This result is not just a corollary of the trivial fact that totally mixed strategy equilibria of games in normal form are perfect.

In section 6 we deal with Bayesian games having a general type structure (not necessarily personal types). We show that the type combinations in such a game can be decomposed into cells of personal types. Such a decomposition separates the types of every player by two characteristics. One characteristic is public and its realization may affect the other players through the probability of their own type selection and through their payoffs. The other characteristic is personal and its realization does not affect the other players in either of these two ways. We show that every Bayesian game has such a unique coarsest personal type decomposition. It turns out that the perfect equilibrium notion on Bayesian game factors nicely through the type decomposition. For a strategy combination to be a perfect equilibrium it suffices that it be a best reply to an arbitrarily close strategy combination which is interior but only relative to the personal decomposition. By being interior relative to the decomposition we mean that in every cell of personal types, the cell types use all the available strategies among them (rather than every type uses all of the strategies). It is then obvious to see that the results in sections 3 and 4 follow immediately

from this general result when the decomposition involved is the trivial coarse decomposition, i.e., consists of only one such cell.

In section 7 we give two examples of Bayesian games which demonstrate that neither independence of types nor private payoffs is enough to guarantee perfection even when the probability of each type combination is positive. One of the examples shows in particular that totally mixed correlated equilibria are not necessarily perfect.

2. Definitions and Notations

We define an n-person Bayesian game G by a set of players $N = \{1,2,\ldots,n\}$ and by a four tuple

$$G = (T = \underset{i\in N}{\times} T_i,\ A = \underset{i\in N}{\times} A_i,\ p,\ u = (u_1,u_2,\ldots,u_n)).$$

The interpretations of these symbols are as follows.

T_i is a finite set describing the possible types that player i may turn out to be.

A_i is a finite set describing the set of actions that player i may take. He can take any of these actions regardless of his type.

p is a probability distribution on T. Thus, for $t = (t_1,t_2,\ldots,t_n)$, $p(t)$ is the prior probability that player one would be of type t_1, player two of type t_2, etc. We assume throughout this paper that for every player i and every $t_i \in T_i$ the marginal probability of type t_i being chosen is positive ($P(t_i) > 0$).

u_i describes the utility of player i as a function of the collective action that is taken and of the types of all the players. Thus

u_i: $A \times T \to \mathbb{R}$.

We may think of the game G as an extensive form game being played in two stages. First, nature selects a type for each player resulting in a $t \in T$ according to the probability distribution p. Now, simultaneously, every player discovers his own type and proceeds to choose any one action from his set of actions A_i. In his choice of an action he can use randomizations over actions. This process results in a pair (a,t) which yields to the players a vector of utilities $u(a,t) = (u_1(a,t),\ldots,u_n(a,t))$.

Since the actions of this game involve randomizations, we extend the utility functions by the usual expected utility rules. If r is a probability distribution on $A \times T$ then

$$u_i(r) = \sum_{(a,t)\in A\times T} u_i(a,t)r(a,t)$$

For a vector $v = (v_1,v_2,\ldots, v_i,\ldots,v_m)$ and a scalar w_i we use the conventions

$$(v_{-i}) = (v_1,v_2,\ldots,v_{i-1}, v_{i+1},\ldots,v_m), \text{ and}$$
$$(v_{-i}: w_i) = (v_1,v_2,\ldots,v_{i-1}, w_i, v_{i+1},\ldots,v_m).$$

Also, for any finite set M we use the symbol $\Delta(M)$ to denote the set of probability distributions on M. Thus, $g \in \Delta(M)$ means that $g: M \to \mathbb{R}$, $g(m) \geq 0$ for every $m \in M$, and $\sum_{m\in M} g(m) = 1$. The int $\Delta(M)$ denotes the distributions that put positive probabilities on every element of M ($g(m) > 0$ for every $m \in M$).

A <u>strategy of player</u> i is a function

$$s_i: T_i \to \Delta(A_i)$$

We denote the set of all strategies of player i by S_i and we let $S = \underset{i\in N}{\times} S_i$ denote the set of <u>strategy tuples</u> of the n players.

A strategy s_i is called <u>completely mixed for type</u> t_i if $s_i(t_i) \in \text{Int } \Delta(A_i)$. s_i is called <u>completely mixed</u> if it is completely mixed for every $t_i \in T_i$. A strategy tuple s is <u>completely mixed</u> if all its s_i's are completely mixed.

For a given strategy tuple $s \in S$ and a types vector $t \in T$ we use $\overline{s(t)}$ to denote the probability distribution induced on A when each player plays his randomized strategy independently of the others. Thus, for every $a \in A$

$$\overline{s(t)}\,(a) = \prod_{i\in N} (s_i(t_i))(a_i).$$

Now we can extend the utility functions u_i to be defined on $S \times T$, in the natural ways as follows

$$u_i(s,t) = \sum_{a\in A} \overline{s(t)}(a)u_i(a,t)$$

$$u_i(s) = \sum_{t\in T} u_i(s,t)p(t).$$

We also define $p(t\,|\,t_i)$ to be the conditional distribution on t given that player i is of type t_i, $p(t_i)$ is the marginal distribution of p on T_i and

$$u_i(s\,|\,t_i) = \sum_{t\in T} u_i(s,t)p(t\,|\,t_i)$$

his expected utility conditioned on his type. We will also use other standard notations for conditional and marginal probabilities as necessary. Observe

that

$$(*) \qquad u_i(s) = \sum_{t_i \in T_i} u_i(s|t_i)p(t_i)$$

In general if $\{L_1, L_2, \ldots, L_m\}$ is a partition of T then

$$u_i(s|L_k) = \sum_{t \in L_k} u_i(s,t)p(t|L_k)$$

and

$$u_i(s) = \sum_{k=1}^{m} u_i(s|L_k)p(L_k)$$

We define the best reply (b.r.) strategies as follows. For a strategy tuple s and player i^{th}'s strategy $r_i \in S_i$ we say that $r_i \in b.r_i(s)$ (really of s_{-i}) if $u_i(s_{-i}: r_i) = \max\{u_i(s_{-i}: q_i): q_i \in S_i\}$. For $r,s \in S$ we say that $r \in b.r(s)$ if $r_i \in b.r_i(s)$ for every $i \in N$. We say that a strategy tuple f^* is a Nash equilibrium if $f^* \in b.r.(f^*)$.

We say that f^* is a perfect (Nash) equilibrium if there is a sequence of completely mixed strategies $f^r \to f^*$ and $f^* \in b.r.(f^r)$ for $r = 1,2,\ldots$. In other words, f^* is a best reply to a sequence of completely mixed strategies approaching it. It follows immediately that a perfect equilibrium is a Nash equilibrium.

3. Games, Revelation Games and Delegation Games with Personal Types

We say that the types in a Bayesian game are personal if two conditions hold:

1. Statistical Independence, i.e., $p(t) = \prod_{i \in N} p(t_i)$ for every $t \in T$; and
2. Private Payoffs, i.e., if $t, \bar{t} \in T$ with $t_i = \bar{t}_i$ then $u_i(a,t) = u_i(a,\bar{t})$ for every $a \in A$.

Observe that many games are of this type. For examples all games involving exchange of commodities for which players have private valuations and types are drawn independently fall into this category.

A strategy $s_i \in S_i$ is completely mixed relative to T_i if for every $a_i \in A_i$ there is a type $t_i \in T_i$ with $(s_i(t_i))(a_i) > 0$. Thus being completely mixed relative to T_i is much weaker than being completely mixed because it is not required that every type of player i completely mixes his actions but only that the group as a whole completely mix their actions.

Theorem 3.1. Let G be a game with personal types. A strategy tuple s^* is a perfect equilibrium if and only if there is a sequence of strategy tuples $s^r \to s^*$ with $s^* \in b.r.(s^r)$ and each s_i^r being completely mixed relative to T_i.

Proof. This theorem follows immediately from its generalization which is proved in the following sections.

Corollary 3.1. In a game with personal types if a Nash equilibrium is completely mixed relative to the T_i's, then it is perfect.

Proof. Choose $s^r = s^*$ for each r.

A game G is called a revelation game if $A_i = T_i$ for every player i. In other words the strategies of the players are to declare their type (with the option to lie).

A revelation game is called Nash incentive compatible if playing honestly is a Nash equilibrium. In other words, the strategy defined by $(s_i(t_i))(t_i) = 1$ for every $i \in N$ and every $t_i \in T_i$ is a Nash equilibrium.

A revelation game is called perfectly incentive compatible if playing honestly is a perfect Nash equilibrium.

Corollary 3.2. A revelation game with personal types is Nash incentive compatible if and only if it is perfectly incentive compatible.

Proof. It is obvious that the honest strategy is completely mixed relative to T_i for every player i. Thus, if honesty is a Nash equilibrium then it must be perfect equilibrium.

By a delegation game we mean a Bayesian game with a vector of functions $g = (g_1,\ldots,g_n)$ such that each g_i is a function from T_i onto A_i. We refer to the functions in g_i as the individual guidelines. Since each g_i is onto it follows that in a delegation game $|A_i| \leq |T_i|$ for every player i.

A delegation game is (or its guidlines are) Nash incentive compatible if obedience to the guidelines is a Nash equilibrium. In other words the strategy tuple s defined by $(s_i(t_i))(g_i(t_i)) = 1$ is a Nash equilibrium. Similarly, it is perfectly incentive compatible if s is a perfect equilibrium.

Observe that every revelation game is a delegation game in one natural way when we define the guidelines to be honest revelation. Also with these guidelines every Nash incentive compatible (respectively perfectly incentive compatible) revelation game is a Nash incentive compatible (and respectively perfectly incentive compatible) delegation game. However, not every delegation game is a revelation or even strategically equivalent to one. If every guideline function of the delegation game is one-to-one then it is clear that the delegation game is essentially a revelation game (the only difference is that the names of the actions may not coincide with the names of the types). But if the guideline functions are not one-to-one then we have a delegation game with $|A_i| < |T_i|$ for some players and thus it cannot be a revelation game.

Observe, however, that the well-known "revelation principle" induces immediately a "delegation principle" which states that every Nash equilibrium

of a Bayesian game induces a Nash incentive compatible delegation game with the same payoff distribution. This follows immediately from the revelation principle and the fact that every revelation game is a delegation game.

Corollary 3.3. A delegation game with personal types is Nash incentive compatible if and only if it is perfectly incentive compatible.

Proof. This is also an immediately consequence of Corollary 3.1.

4. Nash Implementability and Perfect Implementability

In this section it would be convenient to introduce a universal set of outcomes over which the types of the players have preferences. As the players play games, outcomes in this set will result, yielding utility to the players through their preferences over the outcomes.

We first describe an n-person Bayesian environment by

$$E = (T = \underset{i \in N}{\times} T_i,\ P,\ C,\ u = (u_1, u_2, \ldots, u_n)).$$

Each T_i is a finite set describing the types of player i, p is a probability distribution on T with $p(t_i) > 0$ for every player i and every $t_i \in T_i$, C is a finite set of outcomes, and every u_i is a von-Neumann Morgenstern utility function representing the preferences of player i, u_i: $C \times T \to \mathbb{R}$.

To illustrate such an environment consider the confession game described in the introduction and the following Bayesian environment. $N = \{1,2\}$ are the two players, $T_1 = T_2 = \{H,C\}$ denote whether a player is the honest or the criminal type, and $p(H,H) = 1 - \varepsilon$, $p(C,C) = \varepsilon$, $p(H,C) = p(C,H) = 0$ describes the probabilities of the type combinations. The relevant set of outcomes (penalties) here may be modeled as $C = \mathbb{R}^2_-$ $(= -\mathbb{R}^2_+)$ and the utility of player 1 (and similarly player 2) may be viewed as

$$u_1((-x, -y), t) = -x$$

for every one of the possible four type combinations.

Now we define a Bayesian game in the environment E by a set of actions $A = \underset{i \in N}{\times} A_i$ and a resulting outcome assignment $R: A \to \Delta(C)$. While we have now defined the notion of a Bayesian game twice we observe that a Bayesian game in the environment E described by a pair (A,R) induces a Bayesian game (T, A, P, u) in one natural way. The T, A and p components of the Bayesian game are the same while the utility functions u_i are extended from the environment to the game by

$$u_i(a,t) = u_i(R(a), t)$$

Returning to the confession game environment described above we could describe the scheme of the introduction as the following game G.

$A_1 = A_2 = \{A,D\}$, standing for Admit and Deny,

$R(D,D) = (0,0)$ with probability one and,

$R(D,A) = R(A,D) = R(A,A) = (-M,-M)$

with probability one.

The induced utilities here are then for $i = 1,2$ and every type combination t

$u_i((D,D), t) = 0$, and

$u_i((D,A), t) = u_i((A,D), t) = u_i((A,A), t) = -M$.

We are interested in the question of what outcome functions can be implemented by an implementor. An outcome function is a function $O: T \to \Delta(C)$. The implementator may be a public or private agent and may have a variety of incentives in implementing various outcome functions. There are two problems that the implementor may face. First, he may not fully know the types of the players. Second, he may not be able to fully control their actions. We assume that he has some control over the rules or the choice of the game that will be played. The implementor hopes that by choosing the game cleverly and assuming a type of solution regarding the players' behavior he may be able to bring about the outcome function that he chooses.

We let Γ be a set of Bayesian games in the Bayesian environment E and refer to them as the implementation games. The interpretation is that these are the games from which the implementor can choose.

We say that an outcome function O is Nash implementable by Γ if there exists a game in Γ with Nash equilibrium strategy tuple that induces the outcome function O. Formally, there should be a game (A,R) in Γ with a Nash equilibrium strategy tuple s^* such that for every $t \in T$

$$\sum_{a \in A} R(a)\overline{s^*(t)}(a) = O(t)$$

Namely for every type combination t at the Nash equilibrium s^* the players should take actions that induce the distribution O(t).

Similarly we say that an outcome function O is perfectly implementable by Γ if there is a perfect Nash equilibria strategy s^* which implements O.

What is implementable obviously depends on the solution concept. It is clear that every perfectly implementable outcome is Nash implementable but not the converse. Also what is implementable depends on the set of games Γ

available to the implementor. The larger Γ is the more outcomes it may implement. For example, returning to the confession games, the outcome function that we want to implement there is

$O(C,C) = O(H,C) = O(C,H) = (-M,-M)$ with probability 1,

and $O(H,H) = (0,0)$ with probability 1.

If we let $\Gamma = \{G\}$ consists of the game described above (i.e., the scheme from the introduction), then since true revelation is a Nash equilibrium of G we would say that Γ Nash implements O (Γ also Nash implements the total denial strategy). However, Γ does not perfectly implement O. It is interesting to see that by enlarging Γ society can perfectly implement this desired O. The tradeoff in doing this is that the implementor may punish the innocent players harder than before (only off the equilibrium path). Consider the game G_2 defined as follows:

$A_1 = A_2 = \{A,D\}$ as before but

$R(A,D) = (-M, -1.5M)$ with probability 1,

$R(D,A) = (-1.5M, -M)$ with probability 1,

$R(A,A) = (-M,-M)$ with probability 1,

$R(D,D) = (0,0)$ with probability 1.

In other words, G_2 is designed to punish a single denier harder than a confessor. If we let $\Gamma = \{G,G_2\}$ then Γ perfectly implements O by using G_2 in which true revelation is a perfect equilibrium.

One assumption that is necessary for our implementability equivalence theorem is that Γ is rich enough. It will state that the implementor,

starting from every feasible game in Γ can reduce it by crossing out actions which are not used at a Nash equilibrium of this feasible game. More formally we say that Γ is complete if for every $G = (A,R) \in \Gamma$ and every Nash equilibrium strategy s^* of G the game

$$G^-(s^*) = \left(\times_{i \in N} A_i^-, R \right) \in \Gamma$$

where

$$A_i^- = \{a_i \in A_i : \text{for some } t_i \in T_i \ (s_i^*(t_i))(a_i) > 0\}.$$

(See Remark 4.1 for further discussion.)

We say that a Bayesian environment E consists of only personal types if as before the types of the players are statistically independent, and their preferences are private, i.e., for every $i \in N$ and $c \in C$, $u_i(c,t) = u_i(c,\bar{t})$ whenever $t_i = \bar{t}_i$.

Theorem 4.1: The Equivalence of Nash and Perfect Implementability. In a Bayesian environment consisting only of personal types with a complete set of implementation games the sets of Nash implementable outcome functions and perfectly implementable outcome functions coincide.

Proof. We only have to argue that every Nash implementable outcome is perfectly implementable. But if O is Nash implementable by a game G and a Nash equilibrium s^* then it is Nash implementable by the game $G^-(s^*)$ where the strategy tuple s^* (with a little abuse of notations) is also a Nash equilibrium inducing O. But by Corollary 3.1, s^* is a perfect equilibrium of $G^-(s^*)$. Thus O is perfectly implementable by Γ.

Remark 4.1. The condition that Γ is complete means that the implementor can impose a game where actions that are not used at a given equilibrium are

omitted. This assumption is consistent with the philosophy of the implementation literature that we follow in this paper. In this literature it is implicitly assumed that in the case of multiple Nash equilibrium the implementor would have a way of choosing the final one. Our completeness requirement is a weak version of this assumption because it only assumes that the implementor chooses the family of actions that may be used at this equilibrium strategy (rather than the exact strategy). A stronger assumption would have been to let the implementor choose any subset of the players' actions. This would have been too strong because one of its implications would have been that the implementor (remembering that he may be a private agent) can force the players to actions which would yield less than individually rational payoffs. For example, a seller of an item with such a power can force his buyers to pay for his item much more than their valuations for the item. This of course could not happen under our completeness assumption.

We also note that a similar theorem could have been attained by other notions of completeness. For example, starting with a game $G = (A,R)$ and a Nash equilibrium s^*, we could have required that Γ include the games $\hat{G} = (\hat{A},R)$ where the actions available to player i are the mixed strategies induced by his types, i.e.,

$$\hat{A}_i = \{(s_i^*(t_i): t_i \in T_i\}.$$

Notice that the game $\hat{G}$ is equivalent (up to names of actions) to what is known as the revelation game induced by G and s^*. We then in effect could apply the revelation principle and Corollary 3.2 to obtain the same result.

5. Correlated Equilibria

The implementation problem, described in section 4, is that of finding sets of actions A_i for the players and an outcome assignment R defined for these actions given the types T_i and the probability distribution p over types. The designed actions and outcome assignment with the given structure of types generate a Bayesian game as described above. The mediation problem treated in this section is reversed. The actions of the players A_i and the utilities resulting from these actions are given to the mediator (i.e., the mediator faces a game in normal form). What has to be designed is a message space T_i for each player i and a probability distribution p on $T = \underset{i=1}{\overset{n}{\times}} T_i$. The Bayesian game is then defined as follows. The mediator chooses t in T according to the probability distribution p, each player i is informed of his type t_i and chooses accordingly an action $a_i \in A_i$. One can easily verify that an "instruction principle" analogous to the "revelation principle" holds for the mediation problem. That is, any distribution over outcomes that can be achieved by such games can be achieved also if we confine ourselves to message spaces $T_i = A_i$ and to equilibria of the Bayesian game in which each player chooses the action given to him as a message.

In terms of the definitions of section 3 such Bayesian games are Nash incentive compatible. (Of course, the interpretation of these games is different here; "types" should be replaced by "instructions" and "revelation" by "obedience.")

Now let $g = ((A_i)_{i\in N}, (U_i)_{i\in N})$ be a game in normal form. Let p be a probability distribution over $A = \underset{i\in N}{X} A_i$ and let G(p) be the Bayesian game described above. We say that p is a correlated equilibrium in g if G(p) is Nash incentive compatible, i.e., if "obeying" is a Nash equilibrium in G(p). Similarly, p is a perfectly correlated equilibrium of G(p) if it is perfectly

incentive compatible.

Applying Theorem 3.1 to correlated equilibrium yields the following.

Corollary 5.1: Let p be a correlated equilibrium. If $p > 0$ and instructions are drawn independently for each player then p is a perfect correlated equilibrium.

Proof: In the game G(p) types are personal, since statistical independence is assumed and payoffs are independent of types and therefore are private. Since $p > 0$ the requirement that each type is drawn in some positive probability is also satisfied.

The corollary appears to be vacuous, because when instructions are drawn independently the correlated strategy p corresponds to an equilibrium in the game g. Each player can draw his instruction independently of the others and obey his instruction which seems to be tantamount to playing mixed strategy. But then if $p > 0$ we have a totally mixed strategy equilibrium in g which is perfect by definition.

This argument is flawed. Indeed our correlated equilibrium p corresponds to a totally mixed strategy equilibrium in the sense that both yield the same distribution over outcomes. But these two equilibria are two distinct equilibria in two different games. One is a real mixed strategy equilibrium in g; the other is a pure strategy equilibrium in G(p). This mathematical difference between playing the mixed strategy p in the game g and playing the pure obedient strategy in the game G(p) can be interpreted in terms of commitment. The question whether to obey the result of a random choice of action is left open in G(p) while in g playing the mixed strategy should be interpreted as a commitment made by the player before the choice has been made to obey it. This distinction is crucial, for example, in the minimax theory

as is shown by Aumann and Maschler [1972]; mixed minimax strategies cannot in general be supported if players are not precommited to obey the results of the random choice.

Corollary 5.1 should be understood, then, as asserting that perfection unlike minimax theory, is not affected by precommitment in games in normal form. A totally mixed strategy equilibrium remains perfect even if players can disobey the result of randomization.

In Example 2 of section 8 we show that this perfection is due not only to total mixture but also to the independence of players' randomization. Indeed correlated equilibria which are independent in this sense may fail to be perfect even when $p > 0$.

6. General Type Decomposition

If B is a partition of a set S and $s \in S$ then we let B(s) denote the element of B containing s. Given a Bayesian game we say that $L = (L_1, L_2, \ldots, L_n)$ is a decomposition of the type space $T = \underset{i \in N}{\times} T_i$ if for every i

$$L_i = (L_i^1, L_i^2, \ldots, L_i^{\ell(i)})$$

is a finite partition of T_i. For $t \in T$ we let

$$L(t) = \underset{i \in N}{\times} L_i(t)$$

and refer to it as the cell of L containing t. Obviously L generates only finitely many distinct cells. By a personal type decomposition we mean a decomposition $L = (L_1, L_2, \ldots, L_n)$ satisfying the following two conditions:

1. Types are consistently independent relative to the decomposition L, i.e., $p(t|L(t)) = \prod_{i \in N} p(t_i|L_i(t_i))$ for every $t \in T$.
2. Payoffs are private relative to the decomposition L, i.e., for every $t \in T$ and every $\bar{t} \in L(t)$ if $t_i = \bar{t}_i$ then $u_i(a,t) = u_i(a,\bar{t})$ for every $a \in A$.

Notice that personal type decompositions always exist because the trivial fine decomposition, $L_i = (\{t_i\})_{t_i \in T_i}$, trivially yields a personal decomposition. Also, if L is the coarse trivial decomposition, $L_i = (T_i)$, then we are back in the special cases discussed in the previous sections.

To illustrate a personal decomposition consider an example where two players are about to engage in a bidding game for a bottle of old wine. Each of the two players could be of three possible types, denoted by CH, CM, and D. The type CH is one who plans to consume the wine and has a high valuation for this act. The CM type is also a consumer but his desire to drink the wine is moderate. The D type, on the other hand, is a dealer who thinks that this wine has a high potential future value. Suppose that the prior distribution on type combinations is as follows:

	CH	CM	D
CH	.4 × .7 × . 7	.4 × .7 × .3	.1 × .7
CM	.4 × .3 × .7	.4 × .3 × .3	.1 × .3
D	.1 × .7	.1 × .3	.4

This probability distribution satisfies the consistent independence condition

relative to the partition illustrated in the table. The partition cells have the following probability distribution

	C	D
C	.4	.1
D	.1	.4

and within the cells the types are consistently independent with

$$p(CH|C) = .7 \quad \text{and} \quad p(CM|C) = .3$$

The condition of private payoff for this decomposition would require that a consumer of any type should not care against what type of consumer he won (or lost). A dealer type should not care against what type of consumer he won (or lost) buy may care to know that he won (or lost) against another dealer.

A strategy of player i, $s_i \in S_i$, is <u>completely mixed relative to a partition</u> L_i of player i types, T_i, if all the actions of player i are used by the members of every element of his partition, i.e., for every $L_i^j \in L_i$ and for every $a_i \in A$ there is a type $t_i \in L_i^j$ with $s_i(t_i)(a_i) > 0$.

<u>A strategy</u> $s \in S$ <u>is completely mixed relative to the decomposition</u> L if for every player i s_i is completely mixed relative to the partition L_i.

Consider, for example, the bidding game described above with the choice of bid high and bid low available to all types and the following strategy for every one of the two players. Bid high if you are CH and bid high with probability .50 if you are of type D. Clearly this is not an interior

strategy but it is interior relative to the decomposition.

Theorem 6.1. A strategy tuple $s^* \in S$ is a perfect equilibrium if and only if for some (and equivalently for every) personal type decomposition L there exists a sequence of strategies $s^r \to s^*$ with every s^r being completely mixed relative to L and with s^* being a best reply to each s^r.

Proof. The "only if" direction is obvious by the definition of perfect equilibrium and the fact that the trivial fine decomposition is personal. To prove the "if" direction we assume the existence of a sequence s^r as described in the theorem and will construct a sequence of strategy tuples $g^r \to s^*$, g^r being completely mixed and with the property that for every player j:

$$(*) \qquad u_j(g^r_{-j}: h_j) = u_j(s^r_{-j}: h_j)$$

for every $h_j \in S_j$. Thus, since s^* is a best reply to s^r it would also be best reply to g^r showing that s^* is a perfect equilibrium. For $t_i \in T_i$ we let

$$a_i^r(t_i) = \sum_{c_i \in L_i(t_i)} s_i^r(c_i)/\left|L_i(t_i)\right|$$

Thus $a_i^r(t_i)$ is the average action of types of player i which are personally equivalent to t_i. Since the s_i^r's are completely mixed relative to L_i it follows immediately that the $a_i^r(t_i)$'s are completely mixed for every t_i. Now we define g_i^r by

$$g_i^r(t_i) = [1 - 1/rp(t_i|L_i(t_i))]s_i^r(t_i) + [1/rp(t_i|L_i(t_i))]a_i^r(t_i)$$

It is obvious that for sufficiently large r's g_i^r is a well-defined completely

mixed strategy with $g_i^r \to s_i^*$. Now, since for any strategy s, $u_j(s) = \sum_C u_j(s|C)p(C)$ (where we sum over all the cells generated by L) it suffices to show that (*) holds conditionally on C, for each cell C. Moreover, since the strategies g_i^r are defined on a cell C, only in terms of the restrictions of the strategies s_j^r to types in C, we may analyze the restriction of our game to types in C as a complete Bayesian game. Thus we assume without loss of generality that the game has only one such cell and that L is the coarsest decomposition, i.e., for i = 1,2,...,n,

$$L_i = (T_i),$$

$$L_i(t_i) = T_i,$$

$$|L_i(t_i)| = |T_i|,$$

$$a_i^r(t_i) = a_i^r = \sum_{t_i \in T_i} s_i^r(t_i)/|T_i|, \text{ and}$$

$$g_i^r(t_i) = [1 - 1/(rp(t_i))]s_i^r(t_i) + [1/(rp(t_i))]a_i^r.$$

It would suffice to show (*) for this special circumstance. We show indeed that for each type $t_j \in T_j$ and $h_j \in S$; $u_j((g_{-j}^r, h_j)|t_j) = u_j((s_{-j}^r, h_j)|t_j)$. From this (*) follows because for any strategy s,

$$u_j(s) = \sum_{t_j \in T_j} u_j(s|t_j)p|t_j).$$

We observe now that if a game has only personal types than $u_j(a,t) = u_j(a,t_j)$ and for every strategy tuple f,

$$u_j(f|t_j)) = \sum_{a \in A} u_j(a,t_j)(f_j(t_j))(a_j)[\prod_{i \in N \setminus j} \sum_{t_i \in T_i} (f_i(t_i))(a_i)p_i(t_i)]$$

So, for our g^r strategies and any strategy h_j of player j

$$u_i((g^r_{-j}: h_j)\big|t_j) =$$

$$= \sum_{a\in N} u_j(a,t_j)(h_j(t_j))(a_j)\Big[\prod_{i\in N\setminus j}\sum_{t_i\in T_i}(g^r_i(t_i))(a_i)p_i(t_i)\Big] =$$

$$= \sum_{a\in A} u_j(a,t_j)(h_j(t_j))(a_j)\Big[\prod_{i\in N\setminus j}\sum_{t_i\in T_i}(s^r_i(t_i))(a_i)p_i(t_i)\Big] =$$

$$= u_j((s^r_{-j}: h_j)\big|t_j). \qquad \text{Q.E.D.}$$

Clearly the usefulness of Theorem 5.1 depends crucially upon the availability of "coarse" personal decompositions (for the trivially fine decomposition Theorem 5.1 does no more than repeating the definition of a perfect equilibrium). It is therefore useful to know that a unique coarsest personal decomposition $C = (C_1,C_2,\ldots,C_n)$ exists. Also each one of its individual partitions C_i can be constructed independently without having to construct all the C_i's simultaneously.

For a subset of player i types, $D_i \subseteq T_i$, we say that D_i is a <u>set of personal types</u> if

1. $p(t_{-i}\big|d_i) = p(t_{-i}\big|D_i)$ for every $d_i \in D_i$ and $t_{-i} \in T_{-i}$, and
2. $u_j(a,(t_{-i}: d_i)) = u_j(a,(t_{-i}: d_i'))$ for every $j \neq i$, $t_{-i} \in T_{-i}$, d_i, $d_i' \in D_i$ and $a \in A$.

The following are two obvious facts.

1. If B_i and D_i are sets of personal types and $B_i \cap D_i \neq \emptyset$ then $B_i \cup D_i$ is a set of personal types.

2. $\{t_i\}$ is a set of personal types for every $t_i \in T_i$.

It follows immediately that there is a unique coarsest partition of T_i to sets of personal types. We denote this partition by $C_i = \{C_i^1, C_i^2, \ldots, C_i^{\ell(i)}\}$ and refer to it as the coarsest personal type partition of player i. Observe that it follows immediately that a partition B_i is personal to player i if and only if it is a refinement of C_i. We define $C = (C_1, C_2, \ldots, C_n)$ as the coarsest personal type decomposition. This terminology is justified by the following theorem.

Theorem 6.2. L is a personal decomposition of the type set T if and only if for every player i L_i is a refinement of C_i.

Proof. Let $L = (L_1, L_2, \ldots, L_n)$ be a decomposition of T, we want to show that it is a personal decomposition if and only if for every player i L_i is a personal partition of T_i. We first show that if L is personal then for every player i, for every $L_i^j \in L_i$, for every $a,b \in L_i^j$, and for every $t_{-i} \in T_{-i}$, $P(t_{-i}|a) = p(t_{-i}|b)$ (and hence $= p(t_{-i}|L_i^j)$).

Let $L = L(t_{-i}: a) = L(t_{-i}: b)$. Then from the definition of L being personal we obtain that

$$p((t_{-i}: a)|L)/p(a) = p((t_{-i}: b)|L)p(b)$$

it follows that

$$p((t_{-i}: a))/p(a) = p((t_{-i}: b))/p(b)$$

and hence

$$p((t_{-i}|a)) = p((t_{-i}|b)).$$

Next observe that if L is personal and $c,d \in L_i^j$ for some element of the partition L_i then for every $j \neq i$ for every $t_{-i} \in T_{-i}$, and for every $a \in A$

$$u_j(a,(t_{-i}: c)) = u_j((a,(t_{-i}: d)))$$

because $(t_{-i}: c)$ and $(t_{-i}: d)$ belong to the same cell of L. Thus we conclude that if L is a personal decomposition then every L_i is a personal partition and must be refinement of C_i.

Now, to prove the other direction of the theorem, assume that L is a decomposition with each of its L_i's being a personal partition (hence a refinement of C_i). We want to show that for every cell t, L(t) satisfies conditions (1) and (2) in the definition of a personal decomposition. To see that the second condition holds assume that for player i $\bar{t} \in L(t)$ $t_i = \bar{t}_i$ and $a \in A$. Then from the fact that $L_j(t_j)$ $(=L_j(\bar{t}_j))$ is personal for all j, it follows that

$$\begin{aligned}
u_i(\bar{t}) &= u_i(\bar{t}_1,\bar{t}_2,\ldots,\bar{t}_{i-1},t_i,\bar{t}_{i+1},\ldots,\bar{t}_n) \\
&= u_i(t_1,\bar{t}_2,\ldots,\bar{t}_{i-1},t_i,\bar{t}_{i+1},\ldots,\bar{t}_n) \\
&= u_i(t_1,t_2,\ldots,\bar{t}_{i-1},t_i,\bar{t}_{i+1},\ldots,\bar{t}_n) \\
&\quad \vdots \\
&= u_i(t_1,t_2,\ldots,t_{i-1}, t_i, t_{i+1},\ldots,t_n).
\end{aligned}$$

We can change one coordinate at a time because in each change we only change the coordinate of one player j without leaving his $L_j(t_j)$ and thus not affecting the payoff if i ($i \neq j$ and no change was necessary for i).

Condition 1 in the definition of a personal decomposition follows from the identity:

$$P(\underset{i=1}{\overset{s}{\times}} L_i(t_i) \times \{t_{-\{1,2,\ldots,s\}}\}) =$$

$$p(t_{s+1} | L_{s+1}(t_{s+1}))p(\underset{i=1}{\overset{s+1}{\times}} L_i(t_i) \times \{t_{-\{1,2,\ldots,s+1\}}\}) \text{ for } s = 0,1,2,\ldots,n-1$$

Q.E.D.

7. Examples

The truth revealing equilibrium in the Confession Game described in the introduction fails to be perfect. In terms of the conditions guaranteeing perfection in Corollary 3.2, this is due to the statistical dependence of types. But this game has also the special feature that although each type of player has a positive probability, there are type combinations which have probability zero. One might suspect that when $p > 0$ (i.e., when every type combination is possible) one of the conditions defining personal types is superfluous. the next two examples show that even when $p > 0$ neither statistical independence nor private payoffs can guarantee alone the perfection of truth revealing equilibrium.

Example 7.1: The Insufficiency of Statistical Independence of Types Alone. In the following revelation game there are two players, I and II. The sets of types and actions are: $A_1 = T_1 = \{t_1,t_2\}$, $A_2 = T_2 = \{s_1,s_2\}$, and the probability distribution of types is $p(t_1) = p(t_2) = p(s_1) = p(s_2) = 1/2$.

The payoffs for I are given by:

type s_1

type t_1	s_1	s_2
t_1	1	0
t_2	0	2

type s_2

	s_1	s_2
t_1	1	0
t_2	0	1

type t_2	s_1	s_2
t_1	1	0
t_2	0	1

	s_1	s_2
t_1	2	0
t_2	0	1

We assume for simplicity that player II's payoffs are constant in the game. Truth revealing is an equilibrium; player II is clearly indifferent between playing s_1 and s_2. As for player I, each one of his types is indifferent between playing t_1 and t_2; in either case, his expected payoff is 1/2. For example, as type t_1 when he plays t_1 he is paid 1 with probability 1/2 (the probability that player II is of type s_1 and therefore plays s_1) and he is paid 0 with probability 1/2 (the probability that player II is of type s_2 and therefore plays s_2). So in particular truth revealing for player I is a best response to the truth revealing of player II.

Now suppose that player II deviates from truth revealing and he plays $(1 - \varepsilon_1, \varepsilon_1)$ as type s_1 (i.e., he plays s_1 with probability $1 - \varepsilon_1$ rather

than 1), and $(\varepsilon_2, 1 - \varepsilon_2)$ as type $s_2(\varepsilon_1, \varepsilon_2 > 0)$. As type t_1, player I when playing t_1 is paid: $1/2(1 - \varepsilon_1) + 1/2\varepsilon_2$ and when playing t_2: $1/2(2\varepsilon_1) + 1/2(1 - \varepsilon_2)$. So I will not choose the truth (i.e., playing t_1) if $2\varepsilon_1 < 2\varepsilon_1$. Similar computation for type t_2 shows that player I will not choose the truth (i.e., t_2) if $2\varepsilon_1 < 3\varepsilon_2$. Since one of the inequalities always holds it follows that truth revealing for player I is not best response for any deviation of player II from truth revealing. Truth revealing is therefore not a perfect equilibrium.

Example 7.2: The Insufficiency of Private Payoffs Alone. In the revelation game G_2 there are two players, I and II. Types and action sets are $A_1 = T_1 = \{t_1, t_2, t_3\}$ and $A_2 = T_2 = \{s_1, s_2\}$. The probabilities $p(t_i, s_j)$ are given as entries in the following table:

$p(t_i, s_j)$

	s_1	s_2
t_1	2/9	1/9
t_2	1/9	2/9
t_3	1/6	1/6

The payoffs in the game are independent of any type.

II

		s_1	s_2
	t_1	3,0	0,0
I	t_2	0,0	3,0
	t_3	2,0	2,0

Since the payoffs of each player are independent even of his own type a truth revealing equilibrium can be interpreted as a correlated equilibrium.

Truth revealing (or obedience) is an equilibrium in this game. Player II is always indifferent between his two actions. Player I is indifferent between actions t_1 and t_3 when he is of type t_1 (his expected payoff is 2 for either action) and he strictly prefers both actions to t_3. A similar result holds for player I as type t_2. As t_3, player I strictly prefers action t_3.

When player II trembles he plays $(1 - \varepsilon_1, \varepsilon_1)$ as type s_1, and $(\varepsilon_2, 1 - \varepsilon_2)$ as type s_2. When $\varepsilon_2 < 2\varepsilon$, player I prefers action t_3 to t_1 as type t_1 and when $\varepsilon_1 < 2\varepsilon_2$ player I prefers action t_3 to t_2 as player t_2. Under no tremble will player I reveal his type (or obey his instructions) both when he is t_1 and t_2. It follows that truth revealing is not a perfect equilibrium and as a result the correlated equilibrium p is not perfect.

References

Arrow, Kenneth J. [1951], Social Choice and Individual Values, New York: Wiley.

Aumann, Robert J. and M. Maschler [1972], "Some Thoughts on the Minimax Theorem," Management Science, 18, 54-63.

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

Bernheim, Douglas B. [1984], "Rationalizable Strategic Behavior," Econometrica, 52, 1007-1028.

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

Forges, Francoise [1984], "An Approach to Communication Equilibria," CORE Discussion Paper No. 8435, University Catholique de Louvain.

Gibbard, Allan [1973], "Manipulation of Voting Schemes: A General Result," Econometrica, 41, No. 4, 587-601.

Groves, Theodore and John O. Ledyard [1977], "Optimal Allocation of Public Goods: A Solution to the Free Rider Problem," Econometrica, 45, 783-810.

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

Harsanyi, John C. [1967-8], "Games with Incomplete Information Played by 'Bayesian' Players," Management Science, 14, 154-189, 320-334, 486-502.

Harsanyi, John C. and Reinhard Selten [1980], "A General Theory of Equilibrium Selection in Games" (Chapters 1, 2 and 3), Working Papers Nos. 91, 92, and 105, IME, University of Bielefeld, Bielfeld.

Hurwicz, L. [1972], "On Informationally Decentralized Systems," in Decision and Organization, McGuire and Radner (eds.), Amsterdam: North Holland.

Hurwicz, L. and David Schmeidler [1978], "Outcome Functions Which Guarantee the Existence and Pareto Optimality of Nash Equilibria," Econometrica, 46, 144-174.

Kalai, E. and R. W. Rosenthal [1978], "Arbitration of Two-Party Disputes Under Ignorance," International Journal of Game Theory, 7, 65-72.

Kalai, Ehud and Dov Samet [1984], "Persistent Equilibria in Strategic Games," International Journal of Game Theory, 13, 129-144.

Kalai, Ehud and Dov Samet [1985], "Unanimity Games and Pareto Optimality," International Journal of Game Theory, 14, 41-50.

Kreps, David M. and Robert Wilson [1982], "Sequential Equilibria," Econometrica, 50, 863-894.

Maskin, Eric S. [1985], "The Theory of Implementation in Nash Equilibrium: A Survey," in Social Goals and Social Organization: Essays in Memory of Elisha Pazner, Hurwicz, Schmeidler and Sonnenschein (eds.), Cambridge University Press.

Maskin, Eric and J. Riley [1986], "Optimal Auctions with Risk Averse Buyers," Econometrica, forthcoming.

Matthews, Steve A. [1979], "Risk Aversion and the Efficiency of First and Second Price Auctions," Working Paper No. 586, University of Illinois.

McLennan, Andrew [1985], "Justifiable Beliefs in Sequential Equilibria," Econometrica, 54, 889-904.

Mertens, Jean-Francois and Elon Kohlberg [1982], "On the Strategic Stability of Equilibria," CORE Discussion Paper No. 8248, University Catholique de Louvain.

Milgrom, P. R. and John Roberts [1982], "Limit Pricing and Entry Under Incomplete Information: An Equilibrium Analysis," Econometrica, 50, 443-459.

Milgrom, P. R. and R. J. Weber [1982], "A Theory of Auctions and Competitive Bidding," Econometrica, 50, 1089-1122.

Myerson, Roger B. [1978], "Refinement of the Nash Equilibrium Concept," International Journal of Game Theory, 7, 73-80.

Myerson, R. [1981], "Optimal Auction Design," Mathematics of Operations Research, 6, 58-73.

Myerson, Roger B. [1985], "Bayesian Equilibrium and Incentive Compatibility: An Introduction," in Social Goals and Social Organization: Essays in Memory of Elisha Pazner, Hurwicz, Schmeidler and Sonnenschein (eds.), Cambridge University Press.

Myerson, Roger B. [1984], "Acceptable Correlated Equilibria," Discussion Paper No. 591, CMSEMS, Northwestern University, Evanston, Illinois.

Myerson, Roger B. and Mark A. Satterthwaite [1986], "Efficient Mechanisms for Bilateral Trading," The Journal of Economic Theory, forthcoming.

Palfrey, Thomas R. and Sanjay Srivastava, "Private Information in Large Economies," The Journal of Economic Theory, forthcoming.

Pearce, David G [1984], "Rationalizable Strategic Behavior and the Problem of Perfection," Econometrica, 52, 1029-1050.

Peleg, Bezalel [1977], "Consistent Voting Systems," Econometrica, 46, 153-162.

Postlewaite, Andrew [1985], "Implementation via Nash Equilibria in Economic Environments," in Social Goals and Social Organization: Essays in Memory of Elisha Pazner, Hurwicz, Schmeidler and Sonnenschein (eds.), Cambridge University Press.

Postlewaite, Andrew and David Schmeidler [1986], "Implementation in Differential Information Economics," The Journal of Economic Theory, forthcoming.

Riley, J and W. Samuelson [1981], "Optimal Auctions," American Economic Review, 71, 381-392.

Roberts, D. John [1986], "Incentives, Information and Iterative Planning," in Information, Incentives and Economic Mechanisms, Groves, Radner, and Reiter (eds.), University of Minnesota Press (forthcoming).

Satterthwaite, Mark A. [1975], "Strategy Proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions," Journal of Economic Theory, 10, No. 2, 187-217.

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

Vincent, P. Crawford and Joel Sobel [1982], "Strategic Information Transmission," Econometrica.

Wilson, R. [1967], "Competitive Bidding with Proprietary Information," Management Science, 13, A816-A820.

van Damme, Eric [1985], Refinements of the Nash Equilibrium Concept, Berlin, Heidelberg, New York, Tokyo: Springer-Verlag.