A Revelation Principle for (Boundedly) Bayesian Rationalizable Strategies

Peter J. Hammond

Department of Economics, Stanford University, CA 94305–6072, U.S.A.

First version: February 1989; latest revision September 1993; to appear in R.P. Gilles and P.H.M. Ruys (eds.), *Imperfections and Behavior in Economic Organizations* (Boston: Kluwer Academic Publishers).

Abstract

The revelation principle is reconsidered in the light of recent work questioning its general applicability, as well as other work on the Bayesian foundations of game theory. Implementation in rationalizable strategies is considered. A generalized version of the revelation principle is proposed recognizing that, unless agents all have dominant strategies, the outcome of any allocation mechanism depends not only upon agents' "intrinsic" types, but also upon their beliefs about other agents and their strategic behaviour. This generalization applies even if agents are "boundedly rational" in the sense of being Bayesian rational only with respect to bounded models of the game form.

ACKNOWLEDGEMENTS

My special thanks to Fernando Salas, one of whose comments on a preliminary version of Hammond (1990) largely prompted this work. Earlier versions formed the basis of presentations to the conferences on "Rational Behaviour in Games" at the Centre International de Rencontres Mathématiques, Marseille-Luminy, in June 1989, and on "Social Choice and Welfare" at the Universidad Internacional Menéndez Pelayo, Valencia, July 1989. In August 1989 it was also presented at the summer workshop of the Institute for Mathematical Studies in the Social Sciences, Stanford University, and in November 1990, to the Economic Theory seminar at the London School of Economics. I am grateful to Donald Brown, Louis-André Gérard-Varet, Jean-François Mertens, John Moore, and Kevin Roberts for helpful discussion which has mitigated the consequences of my own bounded rationality. And to the Center for Economic Studies of the University of Munich for providing excellent working conditions in which to produce this final version during September 1993.

A REVELATION PRINCIPLE

1. Background

The foundation of recent work on economies with private information is the revelation principle which a number of us discovered more or less independently during the 1970s.[1] But this principle is often misunderstood as giving a fully sufficient rather than merely a necessary condition for implementability of an allocation mechanism. Also, others who understand it very well have recently subjected it to several interesting criticisms.

The main problem with the revelation principle seems to be that, when the equivalent direct revelation mechanism is constructed as a function of what individuals know about the economic environment, truthful revelation of that knowledge is often only one among several equilibrium strategies. Nor is it always the most plausible equilibrium. Green (1984) discussed the difficulties associated with trying to elicit truthful revelation of summary private information. More disturbingly, perhaps, Demski and Sappington (1984) show how, when a principal is confronted with two agents who know about each other, some incentive compatible mechanisms are vulnerable to manipulation by the two agents combining together in order to move to a new "untruthful" equilibrium which makes them both better off. Similar ideas underlie the more recent work of Ma, Moore, and Turnbull (1988). This has led to a revival of the concept of *full* implementation, whereby *every* equilibrium has to produce an outcome which is acceptable according to the social choice rule or performance correspondence being implemented.[2] Other authors have sought implementations using refinements of Bayesian-Nash equilibrium, such as implementation in subgame perfect equilibrium or in undominated strategies.[3]

Yet multiple Nash (or Bayesian) equilibria present their own problems of co-ordination. That is precisely why "Battle of the Sexes" (Luce and Raiffa, 1957) is such an interesting

[1] See Gibbard (1973), Green and Laffont (1977), Myerson (1979, 1982), Dasgupta, Hammond and Maskin (1979), Townsend (1979), Harris and Townsend (1981), Laffont and Maskin (1982), Kumar (1985), Townsend (1988), and Hammond (1992) for various versions of the revelation principle.

[2] Past work on full implementation includes Maskin (1977, 1985), Hurwicz (1979), Dasgupta, Hammond and Maskin (1979), Mookherjee (1984), Williams (1984, 1986), Postlewaite and Schmeidler (1986), Palfrey and Srivastava (1987), Strnad (1987), Ma (1988), Saijo (1988), McKelvey (1989), Mookherjee and Reichelstein (1990), Moore and Repullo (1990), and Jackson (1991).

[3] Examples of these approaches include Moore and Repullo (1988), Howard (1988), Abreu and Sen (1990), and Palfrey and Srivastava (1991).

game. Its outcome clearly depends on the two players' expectations about each other, and may not even be a Nash equilibrium at all. After all, Bernheim (1986) uses the notion of rationalizability due to Bernheim (1984) and Pearce (1984) to argue that, even if a game has a unique Nash equilibrium in pure strategies, that equilibrium is not always the only possible outcome.

In fact recent game theoretical work emphasizes the fundamental rôle of players' expectations. Prominent examples include Aumann (1987) on correlated equilibria, as well as Tan and Werlang (1988) and Rubinstein (1988) on rationalizable strategies, etc. This work makes clear that the outcome of a game is generally very sensitive to what each player believes about other players and their behaviour. Standard Nash or Bayesian equilibrium theory is really a very special case in which almost everything about the game, including the equilibrium strategies played by the players, is supposed to be, if not quite common knowledge in the sense of Lewis (1969) and Aumann (1976), then at least "mutual knowledge" in the sense that all players know it (see Tan and Werlang, 1988). The most interesting exceptions for which much less knowledge suffices occur when each player has a dominant strategy, or when the game is at least "dominance solvable" in the sense of Moulin (1979).

These considerations suggest the need for a generalized version of the revelation principle. The generalization is ultimately intended to allow participants in the economy to have diverse prior beliefs, and very little if any common knowledge or ability to coordinate in reaching a Bayesian or Nash equilibrium. This forces us to consider what can be implemented when it is known only that agents are using rationalizable strategies in the allocation game form. It must also be recognized that implementable allocation mechanisms may well produce outcomes which are sensitive to players' beliefs about each other. And also to their beliefs about each others' beliefs about each others' beliefs ..., and so on *ad infinitum*. The principal exception is the special case considered in Section 7 below, when everybody has a dominant strategy — or at least a "type-dominant" strategy which is optimal no matter what types other agents may be, provided only that they are also using their type-dominant strategies.

In addition, section 8 below is a preliminary exploration of implementation in *boundedly* rationalizable strategies. A rather special concept of bounded rationality will be considered. It is assumed that each agent constructs a simplified — possibly even a trivial — model of

the game form being played, and then optimizes within that model in the usual Bayesian rational manner. This will be called "bounded Bayesian rationality," for obvious reasons. It seems close in spirit to the procedure that Behn and Vaupel (1982) and Vaupel (1986) have suggested for "busy" decision-makers who only have a limited time in which to reach a decision.[4] This all of us surely do when we are not merely deciding how to model rational choice! I believe that it may also relate to the "framing" phenomena discussed in works such as Kahneman, Slovic, and Tversky (1982), and Tversky and Kahneman (1986). After all, the way in which a decision problem is presented to an agent — the way in which it is "framed" — is very likely to influence the (very simplified) decision tree which that agent uses to model the problem.

At first, however, bounded Bayesian rationality seems quite different from Simon's concept of "satisficing," though much closer to "procedural rationality."[5] Of course, satisficing could take the specific form of stopping the analysis of a series of increasingly complex decision trees once some course of action has been found which seems likely to lead to acceptable consequences. Yet, as Behn and Vaupel (1982) argue, a more relevant stopping criterion would seem to be one that takes into account the likelihood that any further analysis will change the final decision. Bounded Bayesian rationality also seems quite different from the approach of Rubinstein and others, who model agents as having strategies which are simple in the sense of being representable by automata with only a few possible states.[6] After all, the full decision tree generated by the problem of choosing even quite a simple automaton in order to solve a difficult decision problem could actually be far too complex for most agents to analyse properly — indeed, it will often be far more complex than the original decision problem itself.[7]

---

[4] For a similar approach to bounded rationality, see Winston (1989).

[5] See Simon (1972, 1982, 1986, 1987a), Radner (1975), Radner and Rothschild (1975).

[6] See Abreu and Rubinstein (1988), Rubinstein (1986, 1987), Kalai and Stanford (1988). For a related approach see Evans and Ramey (1988). In this connection, note that work on games played by unlimited Turing machines, such as that by Anderlini (1988), Canning (1988), Binmore (1989), is not really in the spirit of the bounded Bayesian rationality to be considered here.

[7] More precisely, it has been shown by Gilboa (1988), Ben-Porath (1989), and Papadimitriou (1989) that the problem of choosing an optimal automaton with a bounded number of states to play a game is "NP-complete" — that is, equivalent to a problem like the travelling salesman problem which is sufficiently hard that it is unknown whether it grows faster than any polynomial function of the size of the problem, as the problem becomes large. The general presumption is that such problems cannot in fact be solved in a number of steps which is a polynomial function of its size.

Anyway, Section 8 does not actually consider how the agent chooses which simplified game model to analyse, since that would seem to be a subject which it is better considered by psychologists rather than economists or game theorists. Instead, Section 8 treats each player's final choice of a model in which to analyse the game as essentially exogenous, just as economists usually treat tastes. Using this different notion of "bounded" Bayesian rationality, the conclusion of Section 8 is that the revelation principle still applies, although now agents are characterized by their own models, including the supports of their (exogenous) probabilistic beliefs about other agents' models. Of course, there is no longer any presumption that different agents' models of the game or of each other have anything much in common.

2. Commonly Modelled Game Forms

For the case of games in normal form, the framework I shall use here begins by defining an *intrinsic game form*

$$G = (N, A^N, \Theta^N, X, v^N, \phi) \tag{1}$$

in the way that game forms are usually defined. That is, there is a finite set N of players i who each have specified (action) strategy spaces A_i, and A^N is used to denote the Cartesian product space $\prod_{i \in N} A_i$ of action strategy profiles. Each player i also has an *intrinsic type* space Θ_i. This includes "characteristics" such as endowments and preferences regarding lotteries over outcomes. In other words, a player's intrinsic type consists of those features which would determine behaviour in single person decision models — i.e., in game models which have that one person as their only player. Then Θ^N is used to denote the Cartesian product space $\prod_{i \in N} \Theta_i$ whose members are profiles of intrinsic types. There is also a set X of possible outcomes — economic allocations, or social states, or payoff vectors, depending on the context. Next, each player $i \in N$ has a von Neumann–Morgenstern utility function $v_i : X \times \Theta_i \to \Re$ determining i's utility $v_i(x; \theta_i)$ as a function of the outcome x and of i's intrinsic type $\theta_i \in \Theta_i$. Finally, there is an outcome function $\phi : A^N \to \Delta(X)$ determining the (generally random) outcome $\phi(\cdot; a^N) \in \Delta(X)$ as a function of the pure strategy profile $a^N = \langle a_i \rangle_{i \in N} \in A^N$ chosen by the players $i \in N$. Here, of course, $\Delta(X)$ is used to denote

By contrast, the problem of calculating an unrestricted optimal automaton is a "simple" problem which can be solved in a number of steps which is a polynomial function of its size.

a space of probability distributions over the set X of possible outcomes. Note that at this stage no player has any specified prior probability beliefs about other players' types or about their choices of action. Such beliefs will be specified next.

Indeed, a *commonly modelled game form*

$$\Gamma = (N, A^N, \Theta^N, M^N, T^N, X, v^N, \phi, \mu^N) \tag{2}$$

is defined as an expanded intrinsic game form in which each player i's type space has become a subset T_i of the Cartesian product $\Theta_i \times A_i \times M_i$ of three spaces of different subtypes. Of these three subtypes, the first is just player i's intrinsic type $\theta_i \in \Theta_i$ in the original game form, which has already been discussed.

The second subtype is player i's *behaviour* type. This is just an action strategy $a_i \in A_i$. The idea here is that a type for player i should include everything about which other players may be uncertain, including even i's choice of strategy if there can be any doubt about what this will be. If such behaviour types are not included, the problem of multiple equilibria will remain unresolved. Making explicit players' beliefs about one anothers' strategy choices is, of course, entirely in the spirit of Bernheim (1984, 1986) and Pearce's (1984) work on rationalizable strategies, as well as that of Aumann (1987) and others on correlated equilibrium.

The third subtype is player i's *modelling* type (or just "model") $m_i \in M_i$. The space M_i can be constructed along the lines described in Mertens and Zamir (1985), Tan and Werlang (1988, pp. 373–5), or Brandenburger and Dekel (1993), using ideas pioneered earlier by Böge and his associates.[8] As an implication of this method of construction, an important theorem on projective limits establishes that, provided both the strategy and intrinsic type spaces are compact, complete and separable metric spaces, each player $i \in N$ has a well defined homeomorphism

$$\mu_i : M_i \to \Delta(T_{-i}). \tag{3}$$

This homeomorphism establishes an equivalence between the set of models $m_i \in M_i$ and the set of probability distributions $\mu_i(\cdot; m_i)$ over the product set

$$T_{-i} = \prod\nolimits_{j \in N \setminus \{i\}} T_j \tag{4}$$

[8] See Armbruster and Böge (1979), Böge and Eisele (1979), and the earlier unpublished work cited therein.

whose members are profiles

$$t_{-i} = (\theta_{-i}, a_{-i}, m_{-i}) = \langle\, (\theta_j, a_j, m_j)\, \rangle_{j \in N \setminus \{i\}} \tag{5}$$

of the other players' intrinsic, behaviour, and modelling types. It is precisely this theorem which shows how the infinite recursion of beliefs concerning beliefs concerning beliefs concerning ... converges to something which can be described by a suitable "modelling type" space for each player. It also justifies the above definition of a commonly modelled game form, which has now been made complete by specifying that each component μ_i of μ^N must be the homeomorphism which has just been described.

The game form is called "commonly modelled" because the same spaces M_i $(i \in N)$ both represent each player i's space of possible models and also are the subject of each other player's model of i's model. In fact it has been assumed that all the spaces M_i have been made large and complicated enough to ensure that it is common knowledge among all the players in the game form that each individual player $i \in N$ has some model which belongs to the space M_i. Realistically, spaces large enough to ensure this are likely to be complicated indeed, and so make enormous demands on anybody who is trying to construct such a commonly modelled game form. Accordingly, this important assumption of common modelling will be relaxed in Section 8 below.

3. Bayesian Rationalizable Game Forms

So far nothing has been said about the rationality of the behaviour which players' beliefs ascribe to each other. This will now be remedied. Each player i's type space $T_i \subset \Theta_i \times A_i \times M_i$ is assumed to satisfy Bayesian rationality, and to be the space of all possible "Bayesian rationalizable types," in the following natural sense. First, let player i's expected utility from choosing strategy a_i when his intrinsic and modelling types are (θ_i, m_i) be denoted by

$$\begin{aligned} U_i(a_i; \theta_i, m_i) &:= \mathbb{E}[v_i(x; \theta_i) | a_i, \mu_i(\cdot; m_i)] \\ &= \int_X \int_{A_{-i}} v_i(x; \theta_i)\ \phi(dx; a_i, a_{-i})\ \mathrm{marg}_{A_{-i}}\, \mu_i(da_{-i}; m_i). \end{aligned} \tag{6}$$

Here dx is used to indicate that the outcome x is one variable of integration, and da_{-i} to indicate that a_{-i}, the profile of all the other players' behaviour types, is another. The

integration is with respect to the convolution of the probability distribution $\phi(dx; a_i, a_{-i})$ over outcomes $x \in X$, conditional on a_i and a_{-i}, together with the marginal probability distribution $\text{marg}_{A_{-i}} \mu_i(da_{-i}; m_i)$ over other players' behaviour types $a_{-i} \in A_{-i}$ which is induced by the distribution $\mu_i(dt_{-i}; m_i)$ over other players' entire types $t_{-i} \in T_{-i} := \prod_{j \in N \setminus \{i\}} T_j$, conditional on i's own modelling type m_i.

Now, for all players $i \in N$ and all pairs of intrinsic and modelling types $\theta_i \in \Theta_i$ and $m_i \in M_i$, define the value $B_i(\theta_i, m_i)$ of i's *best response correspondence* as

$$B_i(\theta_i, m_i) = \arg\max_{a_i} \{ U_i(a_i; \theta_i, m_i) \mid a_i \in A_i \}. \tag{7}$$

Thus $B_i(\theta_i, m_i)$ consists of those a_i which maximize i's expected utility conditional upon i's prior probability beliefs about the other players' action strategies or behaviour types a_{-i}, as determined by i's beliefs $\mu_i(\cdot; m_i)$ about other players' entire types t_{-i}. Then, for all players $i \in N$ and all pairs of intrinsic and modelling types $\theta_i \in \Theta_i$ and $m_i \in M_i$, the entire type $t_i = (\theta_i, a_i, m_i)$ is a *Bayesian rationalizable type* in T_i if and only if the strategy a_i satisfies the *Bayesian rationality condition* that

$$a_i \in B_i(\theta_i, m_i). \tag{8}$$

Thus the set T_i of player i's Bayesian rationalizable types is equivalent to the graph

$$T_i = \{ (\theta_i, a_i, m_i) \mid a_i \in B_i(\theta_i, m_i) \} \tag{9}$$

of i's best response correspondence. Note how each player i must therefore have beliefs attaching probability one to the event that all other players $j \in N \setminus \{i\}$ have Bayesian rationalizable types $t_j \in T_j$.

In fact, given any specific intrinsic game form as in (1), the construction of the type spaces along the lines described in Section 2 can be done in a unique way which makes each player i's type set T_i become the largest possible set of Bayesian rational types satisfying (9), for the particular homeomorphism (3) which is also uniquely determined. Any commonly modelled game form (2) which results from this unique construction will be called a *Bayesian rationalizable game form.* Note that, unlike in the traditional Bayesian *equilibrium* game theory, as discussed by Harsanyi (1967–8) and many successors, here there is no presumption that different players' modelling types or prior beliefs are consistent with each other in any

way, except through the requirement that types must be Bayesian rationalizable and that all players must attach probability one to this being so.

4. Implementation

Next we ask what kind of allocation mechanisms or social choice rules can be implemented with such Bayesian rationalizable game forms — in other words, how the outcome of the game form depends upon what aspects of players' types are treated as exogenous variables. Generally it has been assumed that intrinsic types are exogenous, and that both modelling and behaviour types are determined endogenously in equilibrium. For this concept of implementation, the correspondence from intrinsic type profiles $\theta^N \in \Theta^N$ to random outcomes which can be achieved through rationalizable strategies is

$$\Xi(\theta^N) := \{\, \xi \in \Delta(X) \mid \exists (a^N, m^N) : (\theta^N, a^N, m^N) \in T^N \;\&\; \xi = \phi(a^N) \,\}. \tag{10}$$

Thus $\Xi(\theta^N)$ consists of those random outcomes which could result when players' strategies correspond to behaviour types that, in combination with some beliefs about other players, complete the rationalizable Bayesian game model. Similar concepts of implementation, including the standard concept of implementation in Bayesian strategies, would recognize the dependence of the outcome upon just one particular aspect of each player's modelling type — notably, their beliefs about other players' intrinsic types. Yet such concepts of implementation are not really very satisfactory. In the end, only one pure strategy profile $a^N \in A^N$ can be selected — assuming, as I do, that if a player can achieve a "mixed" strategy through some randomization device, the choice of this device should be modelled as part of a pure strategy. Only one profile of modelling types $m^N \in M^N$ describes the actual players in the game. It is just that we do not know which is the right pair $(a^N, m^N) \in A^N \times M^N$, and so which probability distribution of outcomes $\xi \in \Xi(\theta^N)$ will emerge.

Indeed, consider the decision problem which basically underlies all the implementation literature, which is that of selecting a game form whose outcome is satisfactory, or even optimal, relative to some performance criterion. This is a decision problem under uncertainty, including uncertainty about which rationalizable actions $a^N \in A^N$ and which modelling types $m^N \in M^N$ will occur. Like all other uncertainty, it should be described by a subjective probability distribution. This distribution will be essentially exogenous to the game

form, since it could describe the probabilistic beliefs of an external observer, or those of one of the players $i \in N$. There is no reason either to exclude correlated beliefs concerning the behaviour types — i.e., about the strategies chosen by different players. Different subjective probabilities about players' types — especially about their behaviour types — will then give rise to different beliefs about the allocation mechanism which is implemented by the game form.

So it will be assumed that uncertainty about the game form can be represented by the combination of:

(i) a joint probability distribution $\tau \in \Delta(\Theta^N \times M^N)$ describing external beliefs about the pair (θ^N, m^N) of intrinsic and modelling type profiles, and with the property that, for each player $i \in N$, and each type pair $(\theta_i, m_i) \in \Theta_i \times M_i$ of player i, there exists some conditional distribution $\tau_{-i}(d\theta_{-i} \times dm_{-i}|\theta_i, m_i) \in \Delta(\Theta_{-i} \times M_{-i})$ over the other player's intrinsic and modelling types;

(ii) for each $(\theta^N, m^N) \in \Theta^N \times M^N$, a corresponding conditional joint probability distribution $\alpha(\cdot \mid \theta^N, m^N) \in \Delta(B^N(\theta^N, m^N))$ describing possibly correlated external beliefs about the players' selections from their respective sets of optimal strategy profiles, where $B^N(\theta^N, m^N)$ denotes the product set $\prod_{i \in N} B_i(\theta_i, m_i)$.

Game theorists may choose to regard $\alpha(\cdot|\theta^N, m^N)$ as a solution concept, or as a single-valued selection from a "solution correspondence." A very special case is that of a Harsanyi equilibrium, with prior beliefs $\pi_i(\cdot; \theta_i) \in \Delta(\Theta_{-i})$ (all $i \in N$ and all $\theta_i \in \Theta_i$), and special consistency conditions imposed on the model spaces M_i, on the "belief" functions μ_i, as well as on the external probability distributions τ and α described above.

Formally then, a *completed Bayesian rationalizable game model* is defined as

$$(N, A^N, \Theta^N, M^N, X, v^N, \phi, \mu^N, \tau, \alpha) \tag{11}$$

— i.e., as a rationalizable Bayesian game form which has been made into a complete model by the addition of the external probability distributions τ and α whose form has just been described.

Corresponding to each such completion of the original Bayesian rationalizable game form is a unique *equivalent direct mechanism* $\xi^\alpha : \Theta^N \times M^N \to \Delta(X)$ given by the convo-

lution of $\phi(\cdot; a^N) \in \Delta(X)$ with $\alpha(\cdot|\theta^N, m^N) \in \Delta(A^N)$. Thus

$$\xi^\alpha(K|\theta^N, m^N) := \int_{B^N(\theta^N, m^N)} \phi(K; a^N)\ \alpha(da^N|\theta^N, m^N) \tag{12}$$

for every Borel set $K \subset X$. Given the conditional beliefs $\alpha(da^N|\theta^N, m^N)$ regarding the strategy profile a^N, this equivalent direct mechanism specifies the implied probability distribution $\xi^\alpha(dx|\theta^N, m^N)$ over outcomes, as a function of the pair (θ^N, m^N) of intrinsic and modelling type profiles. This is the direct mechanism which will be implemented by the Bayesian rationalizable game form, according to the belief system described by $\alpha(da^N|\theta^N, m^N)$.

It remains to be seen what "incentive constraints" must be satisfied by an equivalent direct mechanism which can be implemented by some such game form. These constraints are most easily expressed in terms of the marginal external beliefs regarding the strategy choice of each player $i \in N$ conditional on knowing only i's type $(\theta_i, m_i) \in \Theta_i \times M_i$. In fact these marginal conditional beliefs are equivalent to a mixed strategy for player i in the game of incomplete information where (θ_i, m_i) describes i's type. The relevant mixed strategy $\alpha_i(da_i|\theta_i, m_i)$ over A_i is given by the marginal distribution on A_i that is derived from the convolution of $\alpha(da^N|\theta_i, \theta_{-i}, m_i, m_{-i})$ with $\tau_{-i}(d\theta_{-i} \times dm_{-i}|\theta_i, m_i)$. Thus

$$\begin{aligned}\alpha_i(K_i|\theta_i, m_i) = \int_{\Theta_{-i} \times M_{-i}} & \tau_{-i}(d\theta_{-i} \times dm_{-i}|\theta_i, m_i) \\ & \int_{K_i} \int_{B_{-i}(\theta_{-i}, m_{-i})} \alpha(da_i \times da_{-i}|\theta_i, \theta_{-i}, m_i, m_{-i})\end{aligned} \tag{13}$$

for every measurable set $K_i \subset A_i$. Note in particular that $\alpha_i(B_i(\theta_i, m_i)|\theta_i, m_i) = 1$. Thus $\alpha_i(da_i|\theta_i, m_i)$ is a probability mixture over the set of i's optimal pure strategies. It can be regarded therefore as an optimal mixed strategy for player i, given i's type (θ_i, m_i). This last property will be crucially important in the following section.

5. A Generalized Revelation Principle

The revelation principle actually applies to any such completed Bayesian rationalizable game model. For there is also an equivalent *completed rationalizable Bayesian game model of direct revelation*

$$(N, A^{DN}, \Theta^N, M^N, X, v^N, \phi^D, \mu^{DN}, \tau^D, \alpha^D). \tag{14}$$

This is a special kind of game form in which each player i's strategy set A_i^D, which is also the set of possible behaviour types, has become equal to the direct revelation strategy set $\Theta_i \times M_i$ of i's possible intrinsic and modelling type pairs. So the outcome function $\phi^D : A^{DN} \to \Delta(X)$ mapping profiles of action strategies into (possibly random) outcomes is effectively defined on the domain $\Theta^N \times M^N$, and is exactly the equivalent direct mechanism $\xi^\alpha : \Theta^N \times M^N \to \Delta(X)$ which has just been defined. Also in this direct revelation game model, each player i's probabilistic beliefs

$$\mu_i^D(\cdot; m_i) \in \Delta(A_{-i}^D \times \Theta_{-i} \times M_{-i}) = \Delta(\Theta_{-i} \times M_{-i} \times \Theta_{-i} \times M_{-i}) \tag{15}$$

about each others' intrinsic and modelling types, together with truthful announcements of those types, are assumed to correspond exactly to those for the original Bayesian rationalizable game form. That is, for every player $i \in N$, modelling strategy $m_i \in M_i$, and measurable subset $K \subset \Theta_{-i} \times M_{-i} \times \Theta_{-i} \times M_{-i}$, it should be true that

$$\mu_i^D(K; m_i) = \mu_i\left(\{\, t_{-i} = (\theta_{-i}, a_{-i}, m_{-i}) \in T_{-i} | (\theta_{-i}, m_{-i}, \theta_{-i}, m_{-i}) \in K \,\}; m_i\right). \tag{16}$$

The direct revelation Bayesian rationalizable game form is assumed to be completed by special external beliefs $\tau^D = \tau \in \Delta(\Theta^N \times M^N)$ and also, for each pair $(\theta^N, m^N) \in \Theta^N \times M^N$, by $\alpha^D(d\theta^N \times dm^N | \theta^N, m^N)$. According to these latter external beliefs, with probability one, each player i's action strategy rule should be the identity map on $\Theta_i \times M_i$. That is, for every pair $(\theta^N, m^N) \in \Theta^N \times M^N$ and every measurable subset $K \subset \Theta^N \times M^N$, the conditional distribution $\alpha^D(d\theta^N \times dm^N | \theta^N, m^N)$ should satisfy

$$\alpha^D(K | \theta^N, m^N) = \begin{cases} 1 & \text{if } (\theta^N, m^N) \in K; \\ 0 & \text{otherwise.} \end{cases} \tag{17}$$

With this construction, truthful revelation happens to be a Bayesian rationalizable strategy in the direct revelation game model. Showing this involves verifying a new version

of the Bayesian rationality condition (8). Note first that any action strategy in the direct revelation game model is a reported pair of types (θ_i', m_i'). According to the external beliefs τ and α, player i's type pair (θ_i', m_i') corresponds to the mixed strategy $\alpha_i(\cdot|\theta_i', m_i')$ defined by (13) in the completion of the original Bayesian rationalizable game form. So in the direct revelation game form we have just constructed, the appropriate new version of player i's expected utility function (6) is

$$U_i^D(\theta_i', m_i'; \theta_i, m_i; \tau, \alpha) := \int_{A_i} U_i(a_i; \theta_i, m_i)\ \alpha_i(da_i|\theta_i', m_i'). \tag{18}$$

Written out in full, $U_i^D(\theta_i', m_i'; \theta_i, m_i; \tau, \alpha)$ is the multiple integral

$$\begin{aligned} \int_{A_i} \int_X \int_{a_{-i} \in A_{-i}} & v_i(x; \theta_i)\ \phi(dx; a_i, a_{-i})\ \mathrm{marg}_{A_{-i}}\, \mu_i(da_{-i}; m_i) \\ & \times \int_{\Theta_{-i}} \int_{M_{-i}} \tau_{-i}(d\theta_{-i} \times dm_{-i} \mid \theta_i', m_i') \\ & \times \int_{b_{-i} \in B_{-i}(\theta_{-i}, m_{-i})} \alpha(da_i \times db_{-i} \mid \theta_i', \theta_{-i}, m_i', m_{-i}). \end{aligned} \tag{19}$$

Note how this is the external expectation, according to the pair of distributions τ and $\alpha(da^N|\theta^N, m^N)$, of player i's own expected utility, according to the intrinsic type θ_i and expectations determined by the model m_i. Alternatively, (19) can be written much more simply as

$$U_i^D(\theta_i', m_i'; \theta_i, m_i; \tau, \alpha) = \int_X v_i(x; \theta_i)\ \xi_i^\alpha(dx|\theta_i', m_i', m_i) \tag{20}$$

where $\xi_i^\alpha : \Theta_i \times M_i \times M_i \to \Delta(X)$ is defined by

$$\xi_i^\alpha(dx|\theta_i', m_i', m_i) := \int_{A_i} \int_{A_{-i}} \phi(dx; a_i, a_{-i})\ \mathrm{marg}_{A_{-i}}\, \mu_i(da_{-i}; m_i)\ \alpha_i(da_i|\theta_i', m_i') \tag{21}$$

for all $i \in N$ and all $\theta_i, \theta_i' \in \Theta_i$, all $m_i, m_i' \in M_i$. The distribution $\xi_i^\alpha(dx|\theta_i', m_i', m_i)$ therefore represents i's beliefs about the outcome of the mechanism, were i to have the true modelling type m_i but then choose the mixed strategy $\alpha_i(da_i|\theta_i', m_i')$ defined by (13) for the intrinsic type θ_i' and the (generally different) modelling type m_i'.

Truthful revelation in the direct revelation game model corresponds to player i's mixed strategy $\alpha_i(\cdot|\theta_i, m_i)$, given that player's true type. At the end of Section 4 it was seen that $\alpha_i(\cdot|\theta_i, m_i)$ attaches probability one to those Bayesian rationalizable strategies in $B_i(\theta_i, m_i)$ which maximize expected utility with respect to all pure strategies $a_i \in A_i$. So, as remarked

at the end of Section 4, $\alpha_i(\cdot|\theta_i, m_i)$ is itself an expected utility maximizing mixed strategy — i.e., it satisfies

$$\alpha_i(\cdot|\theta_i, m_i) \in \arg\max_{\alpha_i'} \left\{ \int_{A_i} U_i(a_i; \theta_i, m_i)\ \alpha_i'(da_i) \mid \alpha_i' \in \Delta(A_i) \right\}. \tag{22}$$

It follows that no "deceptive" mixed strategy $\alpha_i(\cdot|\theta_i', m_i')$ for a different type pair (θ_i', m_i') will increase expected utility, and therefore neither will any corresponding deception in the direct revelation game form. In other words, (22) implies that the mixed strategy $\alpha_i(\cdot|\theta_i, m_i) \in \Delta(A_i)$ is a member of the set

$$\arg\max_{\alpha_i' \in \Delta(A_i)} \left\{ \int_{A_i} U_i(a_i; \theta_i, m_i)\ \alpha_i'(da_i) \mid \exists (\theta_i', m_i') \in \Theta_i \times M_i : \alpha_i' = \alpha_i(\cdot|\theta_i', m_i') \right\}. \tag{23}$$

This implies that

$$\begin{aligned} (\theta_i, m_i) &\in \arg\max_{(\theta_i', m_i')} \left\{ \int_{A_i} U_i(a_i; \theta_i, m_i)\ \alpha_i(da_i|\theta_i', m_i') \mid (\theta_i', m_i') \in \Theta_i \times M_i \right\} \\ &= \arg\max_{(\theta_i', m_i')} \left\{ U_i^D(\theta_i', m_i'; \theta_i, m_i; \tau, \alpha) \mid (\theta_i', m_i') \in \Theta_i \times M_i \right\}, \end{aligned} \tag{24}$$

where the last line follows from (18). This is the appropriate new version of (8), and proves that there is indeed an equivalent direct revelation completed Bayesian rationalizable game model, in which each player's strategy rule is simply truthful revelation of his type.

So the following **revelation principle** applies: the only allocation rules from profiles of players' types to random outcomes which can be implemented by a completed Bayesian rationalizable game model (in the sense of having that allocation rule as their equivalent direct mechanism) are those which are *incentive compatible*, in the usual sense that truthful revelation completes the direct revelation Bayesian game mechanism.

Incentive compatibility requires that (24) be satisfied. This imposes restrictions on the allocation rule which are called *incentive constraints*. From (18) and (6) it follows that these constraints can be expressed as

$$U_i^D(\theta_i', m_i'; \theta_i, m_i; \tau, \alpha) \leq U_i^D(\theta_i, m_i; \theta_i, m_i; \tau, \alpha).$$

Equivalently, because of (20), they can also be expressed as

$$\int_X v_i(x; \theta_i)\ \xi_i^\alpha(dx|\theta_i', m_i', m_i) \leq \int_X v_i(x; \theta_i)\ \xi_i^\alpha(dx|\theta_i, m_i, m_i). \tag{25}$$

Obviously these incentive constraints depend in general upon the external belief system specified by $\tau(d\theta^N \times dm^N)$ and $\alpha(da^N|\theta^N, m^N)$. Indeed, the incentive constraints cannot even be expressed simply in terms of an equivalent direct mechanism. Instead, for each $i \in N$ they involve a separate and artificial mechanism $\xi_i^\alpha(\cdot|\theta_i', m_i', m_i)$, as defined by (21).

Yet no matter what the external belief system may be, truthful revelation is always believed to be Bayesian rational for each player in the corresponding direct revelation game form. This is because the external belief system affects only the allocation mechanism which it is believed that the completed game form implements. It does not affect the set of Bayesian rationalizable types or any players' beliefs about other players. Therefore the set of expected utility maximizing actions remains unchanged.

It has thus been shown that incentive compatibility is a necessary condition for implementability. The same condition is also sufficient, to the following extent. Suppose that, because the incentive constraints are satisfied, the equivalent direct mechanism itself can be set up as a completed Bayesian rationalizable game model, as in (14), with truthful direct revelation as its action strategy rule. Then it is easily checked that the equivalent direct mechanism implements itself. Often, however, the choice of an economic system is subject to additional restrictions which are not modelled within the framework considered here.

6. Concentrating upon Intrinsic Types

If modelling types are not being modelled explicitly, one is led naturally to consider a "reduced form" model. This has modelling types removed by considering the appropriate concentrated marginal distributions. The result will be a marginal conditional probability distribution $\hat{\xi}^{\tau,\alpha}(dx|\theta^N)$ over outcomes, conditional upon the profile of intrinsic types being θ^N. The revelation principle still applies for this new "concentrated" equivalent direct mechanism, as will now be shown.

Formally, assume that the external distribution $\tau(d\theta^N \times dm^N)$ generates, conditional upon each intrinsic type profile $\theta^N \in \Theta^N$, some distribution $\pi^\tau(dm^N|\theta^N)$ over the space M^N of modelling type profiles. Suppose too that, for each individual $i \in N$, the same distribution $\tau(d\theta^N \times dm^N)$ also generates a marginal distribution $\tau_i(d\theta_i \times dm_i)$ over $\Theta_i \times M_i$ having the property that, for each intrinsic type $\theta_i \in \Theta_i$, there is some conditional distribution $\pi_i^\tau(dm_i|\theta_i)$ over the space M_i of player i's modelling types.

With this notation, the equivalent direct mechanism can be expressed as the convolution $\hat{\xi}^{\tau,\alpha} : \Theta^N \to \Delta(X)$ of the distributions $\xi^\alpha(\cdot|\theta^N, m^N)$ and π^τ that is given by

$$\begin{aligned}\hat{\xi}^{\tau,\alpha}(K \mid \theta^N) &:= \int_{M^N} \xi^\alpha(K|\theta^N, m^N)\, \pi^\tau(dm^N|\theta^N) \\ &:= \int_{M^N} \int_{B^N(\theta^N, m^N)} \phi(K; a^N)\ \alpha(da^N|\theta^N, m^N)\ \pi^\tau(dm^N|\theta^N)\end{aligned} \tag{26}$$

for every Borel set $K \subset X$. Now, however, (24) implies that

$$\theta_i \in \arg\max_{\theta_i'} \left\{\, U_i^D(\theta_i', m_i; \theta_i, m_i; \tau, \alpha) \mid \theta_i' \in \Theta_i \,\right\} \tag{27}$$

for every $m_i \in M_i$. From this it follows that

$$\theta_i \in \arg\max_{\theta_i'} \left\{\, \hat{U}_i^D(\theta_i'; \theta_i; \tau, \alpha) \mid \theta_i' \in \Theta_i \,\right\} \tag{28}$$

where

$$\begin{aligned}\hat{U}_i^D(\theta_i'; \theta_i; \tau, \alpha) &:= \int_{M_i} U_i^D(\theta_i', m_i; \theta_i, m_i; \tau, \alpha)\ \pi_i^\tau(dm_i|\theta_i) \\ &= \int_{M_i} \int_{A_i} U_i(a_i; \theta_i, m_i)\ \alpha_i(da_i|\theta_i', m_i)\ \pi_i^\tau(dm_i|\theta_i) \\ &= \int_{M_i} \int_{A_i} \int_X \int_{A_{-i}} v_i(x; \theta_i)\ \phi(dx; a_i, a_{-i})\ \mathrm{marg}_{A_{-i}}\, \mu_i(da_{-i}; m_i) \\ &\qquad\qquad \times \alpha_i(da_i|\theta_i', m_i)\ \pi_i^\tau(dm_i|\theta_i) \\ &= \int_X v_i(x; \theta_i)\ \hat{\xi}_i^{\tau,\alpha}(dx|\theta_i', \theta_i)\end{aligned} \tag{29}$$

This is what i's expected utility would be if his true intrinsic type were θ_i, but then he acted in the game form with the mixed strategy $\alpha_i(da_i|\theta_i', m_i)$ which matches external expectations regarding the behaviour of an agent of intrinsic type θ_i'. The last line of (29) involves the mapping $\hat{\xi}_i^{\tau,\alpha} : \Theta_i \times \Theta_i \to \Delta(X)$ whose value $\hat{\xi}_i^{\tau,\alpha}(dx|\theta_i', \theta_i)$ for any pair (θ_i', θ_i) is defined to be the probability distribution

$$\int_{M_i} \int_{A_i} \int_{A_{-i}} \phi(dx; a_i, a_{-i})\ \mathrm{marg}_{A_{-i}}\, \mu_i(da_{-i}; m_i)\ \alpha_i(da_i|\theta_i', m_i)\ \pi_i^\tau(dm_i|\theta_i). \tag{30}$$

This describes what i's beliefs concerning the outcome $x \in X$ would be if i's true intrinsic characteristic were θ_i, but i chose the mixed strategy $\alpha_i(da_i|\theta_i', m_i)$ which the external beliefs implicitly ascribe to an agent with intrinsic characteristic θ_i'.

As in Section 5, a routine argument then shows that $\theta'_i = \theta_i$ is a Bayesian rationalizable strategy for each player, so the revelation principle remains valid even after eliminating the modelling types. The incentive constraints, however, take the new form

$$\hat{U}_i^D(\theta'_i; \theta_i; \tau, \alpha) \leq \hat{U}_i^D(\theta_i; \theta_i; \tau, \alpha)$$

or equivalently

$$\int_X v_i(x; \theta_i)\ \hat{\xi}_i^{\tau,\alpha}(dx|\theta'_i, \theta_i) \leq \int_X v_i(x; \theta_i)\ \hat{\xi}_i^{\tau,\alpha}(dx|\theta_i, \theta_i) \tag{31}$$

for all $i \in N$ and all $\theta_i, \theta'_i \in \Theta_i$. So obviously these incentive constraints depend in general upon the external belief system $\tau(d\theta^N \times dm^N)$ and $\alpha(da^N|\theta^N, m^N)$ through the induced distribution of modelling types $\pi_i^\tau(dm_i|\theta_i)$. In the end, therefore, external beliefs about modelling types cannot be neglected altogether.

7. Dominant Strategy Incentive Constraints

There is, however, one very important special case in which the outcome is virtually independent of external beliefs about modelling types, and can be treated as independent of behaviour types as well. This is when each player $i \in N$ has a *type dominant* strategy rule $a_i^* : \Theta_i \to A_i$ which is optimal against the other players' strategy rules $a_j^* : \Theta_j \to A_j$ (all $j \in N \setminus \{i\}$), no matter what their type profile $\theta_{-i} \in \Theta_{-i}$ may be. Thus all the strategy rules $a_i^* : \Theta_i \to A_i$ (all $i \in N$) must together have the property that

$$\int_X v_i(x; \theta_i)\ \phi(dx; a_i, a_{-i}^*(\bar{\theta}_{-i})) \leq \int_X v_i(x; \theta_i)\ \phi(dx; a_i^*(\theta_i), a_{-i}^*(\bar{\theta}_{-i})) \tag{32}$$

for each intrinsic type $\theta_i \in \Theta_i$, each action $a_i \in A_i$, and for all $\bar{\theta}_{-i} \in \Theta_{-i}$. Here, of course, $a_{-i}^*(\bar{\theta}_{-i})$ denotes the profile $\langle a_j^*(\bar{\theta}_j)\rangle_{j \in N \setminus \{i\}}$. Note that in this case the best response rule $a_i^*(\theta_i)$ must be entirely independent of i's modelling type.

Now let $a^{*N}(\theta^N)$ denote the strategy profile $\langle a_i^*(\theta_i)\rangle_{i \in N}$. And suppose that the external expectations $\alpha(da^N|\theta^N, m^N)$ concentrate on the point $a^{*N}(\theta^N)$ for all θ^N and all m^N — i.e., suppose they satisfy the restriction that each individual $i \in N$ is believed to select the particular type dominant strategy $a_i^*(\theta_i)$ with probability one. This means that, for every Borel set $K \subset A^N$, one has

$$\alpha(K|\theta^N, m^N) = \begin{cases} 1 & \text{if } a^{*N}(\theta^N) \in K; \\ 0 & \text{otherwise.} \end{cases} \tag{33}$$

Then we have a *dominant strategy completed game form* for which there is an equivalent direct mechanism

$$\xi^*(dx|\theta^N) := \phi(dx; a^{*N}(\theta^N)). \tag{34}$$

The random outcome of this mechanism therefore depends only upon the intrinsic type profile $\theta^N \in \Theta^N$.

Now (32) clearly implies that

$$\int_X v_i(x;\theta_i)\ \phi(dx; a_i^*(\theta_i'), a_{-i}^*(\bar{\theta}_{-i})) \leq \int_X v_i(x;\theta_i)\ \phi(dx; a_i^*(\theta_i), a_{-i}^*(\bar{\theta}_{-i})) \tag{35}$$

for all $i \in N$, all $\theta_i, \theta_i' \in \Theta_i$, and all $\bar{\theta}_{-i} \in \Theta_{-i}$. Therefore, in the equivalent direct revelation Bayesian game model, it must be true that

$$\int_X v_i(x;\theta_i)\ \xi^*(dx|\theta_i', \bar{\theta}_{-i}) \leq \int_X v_i(x;\theta_i)\ \xi^*(dx|\theta_i, \bar{\theta}_{-i}). \tag{36}$$

So truthful direct revelation is always a dominant strategy. Players' models no longer matter, provided each player i believes with probability 1 that all the other players will choose actions in the range space of those a_{-i} for which there exists some $\theta_{-i} \in \Theta_{-i}$ with $a_{-i} = a_{-i}^*(\theta_{-i})$.

In this case both the equivalent direct mechanism and the incentive constraints have become independent of modelling and behaviour types. However, if there are multiple dominant strategies for some intrinsic type profiles $\theta^N \in \Theta^N$, both do depend on the selection rule used to construct $a^{*N}(\theta^N)$. These properties of dominant strategy mechanisms, of course, are just familiar results (as in Dasgupta, Hammond and Maskin, 1979) slightly adapted to suit the new setting.

8. Bounded Modelling

In Section 2, a commonly modelled game form was defined so that the players all had models of the other players in some large common product set M^N. This implies that players must not only have models of each other's preferences and strategy choices, but also models of each other's models, and of each other's models of each other's models, ... etc., without end. It has been common in Bayesian game theory, following the pioneering work of Harsanyi (1967–8), to assume that each player can only have a finite number of possible types, so that the modelling sets M_i are finite. Mertens and Zamir (1985) showed that this could be an acceptable approximation. Yet still the number of different models needed for such an approximation might have to be immensely large. So one could well argue that expecting players to have such rich models of one another imposes excessive demands upon their modelling and reasoning faculties. This, of course, merely adds to all the usual and well known reasons for wanting to model players of games as being merely boundedly rational.

Indeed, one should really think of each player i as having a possibly very limited model in which the modelled set of other players' possible models is also possibly very limited — perhaps even trivial. Players could even simplify their respective models of the game by leaving out some of the other players altogether. Actually, in some games described by Markov processes, as in Shefrin (1981), this could even be fully rational, because no player needs to know who the other players are, but only what Markov process their equilibrium behaviour generates for those state variables that the player can observe. Of course, a player could also restrict greatly the modelled strategy and type sets of those other players whom he does choose to include in his model. He may even simplify his own strategy set. All of these simplifications are things that real players of real game forms do, as we know full well from both introspection and more careful psychological studies. Some such simplifications are clearly necessary for real life game forms which have to be played out in "real time."

As game theorists or social scientists, however, we are free to allow ourselves the conceit that our game models can be much richer. But they are still game models in the sense of (2). Actually, most game theorists' game models have not been rich enough precisely because they have ignored psychological reality and modelled players as if they were unboundedly rational — or, at least, as if they were no less rational than the modeller who is able to draw

upon the as yet imperfect and incomplete conclusions of many collective highly intelligent human-years of game-theoretic studies. This is inevitable if we retain the common modelling assumption. So it is time to abandon that assumption and allow different players to have different models. We should also allow ourselves as game-theorists to have our own models which may well differ from all the players' models. This is precisely what appropriate generalizations of the game models described by (2) allow, as long as we stay well clear of the strait-jacket of common modelling.

Accordingly, a *boundedly modelled intrinsic game form*

$$G = (N, A^N, \Theta^N, X^N, v^N, \phi^N) \tag{37}$$

will now be defined in a way which resembles (1), but with some important differences. Of these, the first is that each player i is allowed to have a modelled action strategy set $A_i(\theta_i)$ which depends upon his intrinsic type $\theta_i \in \Theta_i$. This reflects the possibility that player i will model his own set of possible strategies as just a small subset of the true set A_i. Also, each player i has a modelled outcome set $X_i(\theta_i)$ which depends upon θ_i, reflecting the possibility that a player could even model the range of possible outcomes as some proper subset of the true set X. In a similar way, player i's von Neumann–Morgenstern utility function $v_i(x; \theta_i)$, which is now only defined on the domain of pairs $(x; \theta_i)$ satisfying $x \in X_i(\theta_i)$, could well be a simplification of the true function, though all this is supposed to be reflected already in the description of the intrinsic type θ_i. Finally, each player i also has a boundedly modelled outcome function $\phi_i(dx; a^N, \theta_i) \in \Delta(X_i(\theta_i))$ specifying only modelled outcomes as possible. Later, too, its values will only be used for those strategy profiles a^N which i models as possible.

Next, a *boundedly modelled game form*

$$\Gamma = (N, A^N, \Theta^N, M^N, T^N, X^N, v^N, \phi^N, \mu^N) \tag{38}$$

is defined as a boundedly modelled intrinsic game form in which, as in (2) above, each player i's type space has become a subset T_i of the Cartesian product $\Theta_i \times A_i \times M_i$ of three spaces of different subtypes. As before, player i's *behaviour* type is any action strategy $a_i \in A_i$; there is no need for any other player to impose the same limitations on his model of i's strategy space as those which i places on his own model.

Even with the additional complications which arise because players' models can be bounded, the space M_i of player i's *modelling* types can still be constructed along the lines mentioned earlier in Section 2, provided that both the strategy and intrinsic type spaces remain as compact, complete and separable metric spaces, for each player $i \in N$. Nevertheless, as has been the tradition in discussing Bayesian games of incomplete information, it will be assumed here that, for each player $i \in N$, there are finite model spaces M_i and well defined mappings as in (3) above which together determine, for each $m_i \in M_i$, a probability distribution $\mu_i(\cdot; m_i)$ over the product set $\Theta_{-i} \times A_{-i} \times M_{-i}$ defined in (4). In fact, when players use only bounded models, then both the set of possible models M_i and the function μ_i which is given by (3) have to be regarded as additional parts of the exogenous description of the game model. This is quite unlike the unique construction of the players' Bayesian type spaces and their expectations which yielded a Bayesian rationalizable game form in Section 3 above. This need to specify exogenously the sets M_i and the functions μ_i (all $i \in N$) is really the main change from the previous analysis of unbounded models.

After this slight change of notation but radical change of interpretation, much of the previous analysis of Sections 3, 4 and 5 remains valid. Nevertheless there are a few differences as follows. The first is that player i's expected utility function becomes

$$U_i(a_i; \theta_i, m_i) := \int_{X_i(\theta_i)} \int_{A_{-i}} v_i(x; \theta_i)\ \phi_i(dx; a_i, a_{-i}, \theta_i)\ \mathrm{marg}_{A_{-i}}\, \mu_i(da_{-i}; m_i) \tag{39}$$

instead of (6). This is because of the limitations of i's model of the game form. Also, i's best response correspondence (7) takes the "bounded" form

$$B_i(\theta_i, m_i) = \arg\max_{a_i} \{\, U_i(a_i; \theta_i, m_i) \mid a_i \in A_i(\theta_i) \,\} \tag{40}$$

because player i only considers strategies in the modelled set $A_i(\theta_i)$.

No doubt one should consider more general models of boundedly rationality than those which presume such bounded Bayesian rationality. Here, however, I shall display some bounded rationality of my own by limiting the models which I myself consider to those which presume some degree of Bayesian rationality on the part of all players. In fact I am assuming that the only bounds on a player's rationality are limits on the model in which beliefs are formulated and in which an expected utility maximizing action strategy is chosen, rather than limits on the player's ability to maximize expected utility *per se*. In

other words, players are assumed to use models no more complicated than those in which they can solve the appropriate expected utility maximization problem.

In addition, (11) is changed so that a *completed boundedly rationalizable game model* is defined as

$$(N, A^N, \Theta^N, M^N, X^N, v^N, \phi^N, \phi, \mu^N, \tau, \alpha) \tag{41}$$

— i.e., a boundedly rationalizable Bayesian game form which has been made into a complete model by the addition of an externally assessed outcome function ϕ, as well as the external probability distributions τ, α as in Section 4.

Next (14) is modified to become a *completed boundedly rationalizable Bayesian game model of direct revelation*

$$(N, A^{DN}, \Theta^N, M^N, X^N, v^N, \phi^N, \phi^D, \mu^{DN}, \tau^D, \alpha^D) \tag{42}$$

which is equivalent to (41). Of course, the appropriate new version of player i's expected utility function (39) is

$$U_i^D(\theta_i', m_i'; \theta_i, m_i; \tau, \alpha) := \int_{A_i(\theta_i)} U_i(a_i; \theta_i, m_i)\ \alpha_i(da_i|\theta_i', m_i') \tag{43}$$

rather than (18). Similar changes have to be made to (22), (23), and (24) in turn. This implies that the incentive constraints (25) take the slightly new form

$$\int_X v_i(x; \theta_i)\ \tilde{\xi}_i^\alpha(dx|\theta_i', m_i', m_i) \leq \int_X v_i(x; \theta_i)\ \tilde{\xi}_i^\alpha(dx|\theta_i, m_i, m_i). \tag{44}$$

The only difference is that $\xi_i^\alpha(dx|\theta_i', m_i', m_i)$ has been replaced by $\tilde{\xi}_i^\alpha(dx|\theta_i', m_i', m_i)$, which is a function whose value is defined everywhere as

$$\int_{A_i(\theta_i)} \int_{A_{-i}} \phi(dx; a_i, a_{-i})\ \mathrm{marg}_{A_{-i}}\, \mu_i(da_{-i}; m_i)\ \alpha_i(da_i|\theta_i', m_i'). \tag{45}$$

In this very slightly revised form, the revelation principle of Section 5 still remains valid.

9. Desirable Extensions

The above discussion was conducted throughout for game models in normal form, in which each player is modelled as having a single modelling type m_i in the game form, and as making a single choice of action strategy A_i to last for the entire duration of the modelled game. Of course, extensive form game models could also be discussed in this way, by introducing the standard fiction due to Selten (1975) that each potential player lives only at a single information set before turning into somebody else. But it would obviously be desirable to use an appropriate concept of (bounded) sequential Bayesian rationalizability (generalizing Kreps' and Wilson's (1982) concept of sequential equilibrium). Then, as in Kumar's (1985) "incomplete" revelation principle, the timing of information revelation becomes an important issue.[9]

All this suggests to me that the revelation principle is very much more broadly applicable than has generally been realized, but there is a high price. Those allocation mechanisms or social choice rules which can be implemented by completed rationalizable Bayesian game models generally have outcomes which depend on player's modelling and behaviour types as well as on intrinsic types such as preferences and endowments. Multiple outcomes are indeed possible for any given profile of intrinsic types. It is true that adding modelling and behaviour types makes the outcome unique for each profile of entire types, but then external beliefs about agents' behaviour types affect both the mechanism which a game form is thought to implement and the incentive constraints which that mechanism must satisfy. Perhaps economists' models of the whole economy or of particular parts of it really do need to treat agents' modelling and behaviour types as being at least in part exogenous, rather than wholly endogenous as in standard models of "rational expectations."

This conclusion seems to be reinforced by a considerable body of relevant recent work. For example, McAllister (1988) and Rahi (1993) in particular demonstrate the general multiplicity of rational expectations equilibria. Then Kirman (1983) and other papers in Frydman and Phelps (1983), as well as those of Marcet and Sargent (1988, 1989) and Kurz

[9] See also the recent work on "renegotiation-proof equilibria" by Laffont and Tirole (1987, 1988a, 1988b), Dewatripont (1988), Hart and Tirole (1988), as well as related work by Freixas, Guesnerie and Tirole (1985), Malcomson and Spinnewyn (1988), etc. Townsend (1988) specifically discusses problems with the revelation principle in a dynamic setting, but appears to be unaware of Kumar's work.

(1988), all consider how even fully rational agents are likely to face difficulties in learning to acquire rational expectations. Also, Plott and Sunder (1988), together with Smith, Suchanek and Williams (1988), show how difficult it can be for real people to acquire rational expectations even in laboratory experiments, which are surely much less complex than real economies. Most recently, the published papers by Fudenberg and Kreps (1993) and by Jordan (1993) point out some problems which players face in learning to play mixed strategy Nash equilibria in particular. See also the series of discussion papers by Nyarko (1992a, b). These conclusions may at first seem to contradict those of Kalai and Lehrer (1993), but do not really do so because the latter authors rely on rather special assumptions.

In the end it seems that the only satisfactory alternative to explicit consideration of modelling and behaviour types in the theory of mechanism design is to construct allocation mechanisms which not only ensure that all (sequential, or subgame perfect) Nash equilibrium outcomes are acceptable, but so too are all those which can emerge from rationalizable strategies, etc. Such are the type dominant strategy mechanisms considered in Section 7.

10. Conclusion

The revelation principle can be regarded as saying that, by the time agents have manipulated the economic system or any other game form as much as they please, the resulting equivalent direct mechanism cannot possibly be manipulated any further. With this simple interpretation in mind, it should not be at all surprising how robust this principle turns out to be. The main difficulty, in fact, comes in constructing the equivalent direct mechanism. The natural construction presented in this paper depends upon external prior beliefs over the set of agents' (boundedly) Bayesian rationalizable strategies in the game form. For someone trying to construct an optimal game form, such prior beliefs are an entirely natural Bayesian description of uncertainty regarding agents' strategic choices in that game form.

This paper has shown that there is a sense in which the revelation principle survives, even when agents do not share the same prior distribution over one anothers' intrinsic types, strategy choices, and models, and even when they use models sufficiently simple for them to be able and willing to maximize expected utility within them. Nevertheless, the principle becomes significantly weakened. The equivalent direct mechanism to which the revelation principle applies is generally sensitive to agents' models. Moreover, this mechanism, as well

as the incentive constraints which it has to satisfy, are both sensitive to the specification of external beliefs concerning which Bayesian rationalizable strategies the agents will choose.

This conclusion leaves us with just two alternatives. Either players' action strategies must be modelled as functions of their modelling and behaviour types, and the possible dependence of the economic allocation upon such types duly recognized. For the case of Dutch auctions, this dependence was illustrated in Hammond (1990). Alternatively, attention must be restricted to allocation mechanisms which can be implemented with type dominant strategy game forms.

In fact it seems to me that dominant strategy mechanisms have generally been neglected for too long. This may be due to the (highly deserved) attention paid to such early negative results as those of Hurwicz (1972, 1973), Gibbard (1973, 1977), Satterthwaite (1975), and Barberà (1979) for general economic or social environments. It may be due to the difficulties of getting Groves transfer schemes to balance and even to avoid bankruptcy in some cases, even for the very restricted domain of quasi-linear preferences (Green and Laffont, 1979). Much of this work, however, looked only for mechanisms which achieve first best Pareto efficient outcomes despite incomplete information. Or else failed to consider random mechanisms which make use of cardinal information regarding individuals' von Neumann–Morgenstern utility functions. Often both these restrictions were imposed together. More recently, however, Grossman and Hart (1983), Prescott and Townsend (1984a, b), Townsend (1987), and others have all shown, for environments with finite decision and type spaces, how to set up and solve linear programs which determine dominant strategy mechanisms that are Pareto efficient subject to incentive constraints, or *incentive constrained Pareto efficient*. Moreover, Page's (1987, 1988) work in particular suggests how extensions to larger decision and type spaces may be possible.

References

D. Abreu and A. Rubinstein (1986), "The Structure of Nash Equilibria in Repeated Games with Finite Automata," *Econometrica*, **56**: 1259–1282.

D. Abreu and A. Sen (1990), "Subgame Perfect Implementation: A Necessary and Almost Sufficient Condition," *Journal of Economic Theory*, **50**: 285–299.

L. Anderlini (1988), "Some Notes on Church's Thesis and the Theory of Games," University of Cambridge, Economic Theory Discussion Paper No. 126.

W. Armbruster and W. Böge (1979), "Bayesian Game Theory," in *Game Theory and Related Topics* edited by O. Moeschlin and D. Pallaschke (Amsterdam: North-Holland), pp. 17–28.

R.J. Aumann (1976), "Agreeing to Disagree," *Annals of Statistics*, **4**: 1236–1239.

R.J. Aumann (1987), "Correlated Equilibrium as an Expression of Bayesian Rationality," *Econometrica*, **55**: 1–18.

S. Barberà (1979), "Majority and Positional Voting in a Probabilistic Framework," *Review of Economic Studies*, **46**: 379–389.

R.D. Behn and J. W. Vaupel (1982), *Quick Analysis for Busy Decision Makers*. New York: Basic Books.

E. Ben-Porath (1989), "The Complexity of Computing a Best Response Automaton in Repeated Games with Mixed Strategies," preprint, Stanford University, Graduate School of Business.

B.D. Bernheim (1984), "Rationalizable Strategic Behavior," *Econometrica*, **52**: 1007–1028.

B.D. Bernheim (1986), "Axiomatic Characterizations of Rational Choice in Strategic Environments," *Scandinavian Journal of Economics*, **88**: 473–488.

W. Böge and T. Eisele (1979), "On Solutions of Bayesian Games," *International Journal of Game Theory*, **8**: 193–215.

K. Binmore (1988, 1989), "Modelling Rational Players, I and II," *Economics and Philosophy*, **3**: 179–214, **4**: 9–55.

A. Brandenburger and E. Dekel (1993), "Hierarchies of Beliefs and Common Knowledge," *Journal of Economic Theory*, **59**: 189–198.

R. Campbell and L. Sowden (eds.) (1985), *Paradoxes of Rationality and Cooperation*. Vancouver: University of British Columbia Press.

D. Canning (1988), "Rationality and Game Theory when Players are Turing Machines," London School of Economics, STICERD Discussion Paper No. TE/88/183.

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

J. Demski and D. Sappington (1984), "Optimal Incentive Contracts with Multiple Agents," *Journal of Economic Theory*, **33**: 152–171.

M. Dewatripont (1988), "Commitment through Renegotiation-Proof Contracts with Third Parties," *Review of Economic Studies*, **55**: 377–390.

J. Eatwell, M. Milgate and P. Newman (eds.) (1987), *The New Palgrave: A Dictionary of Economics*. London: Macmillan.

G.W. Evans and G. Ramey (1988), "Calculation Equilibria," Stanford University, Institute of Mathematical Studies in the Social Sciences, Economics Technical Report No. 530.

X. Freixas, R. Guesnerie and J. Tirole (1985), "Planning under Incomplete Information and the Ratchet Effect," *Review of Economic Studies*, **52**: 173–191.

R. Frydman and E.S. Phelps (eds.) (1983), *Individual Forecasting and Aggregate Outcomes: 'Rational Expectations' Examined*. Cambridge: Cambridge University Press.

D. Fudenberg and D.M. Kreps (1993), "Learning Mixed Equilibria," *Games and Economic Behavior*, **5**: 320–367.

A. Gibbard (1973), "Manipulation of Voting Schemes: A General Result," *Econometrica*, **41**: 587–601.

A. Gibbard (1977), "Manipulation of Schemes that Mix Voting with Chance," *Econometrica*, **45**: 665–681.

I. Gilboa (1988), "The Complexity of Computing Best Response Automata in Repeated Games," *Journal of Economic Theory*, **45**: 342–352.

E.J. Green (1984), "On the Difficulty of Eliciting Summary Information," *Journal of Economic Theory*, **32**: 228–245.

J.R. Green and J.-J. Laffont (1977), "Characterization of Satisfactory Mechanisms for the Revelation of Preferences for Public Goods," *Econometrica*, **45**: 427–438.

J.R. Green and J.-J. Laffont (1979), *Incentives for Public Decision Making*. Amsterdam: North-Holland.

S.J. Grossman and O.D. Hart (1983), "An Analysis of the Principal-Agent Problem," *Econometrica*, **51**: 7–45.

R. Guesnerie and J.-J. Laffont (1982), "On the Robustness of Strategy Proof Mechanisms," *Journal of Mathematical Economics*, **19**: 5–15.

F.H. Hahn (1974), *On the Notion of Equilibrium in Economics*. Cambridge: Cambridge University Press; reprinted as ch. 2, pp. 43–71 of Hahn (1984).

F.H. Hahn (1984), *Equilibrium and Macroeconomics*. Cambridge, Mass.: M.I.T. Press.

P.J. Hammond (1990), "Incentives and Allocation Mechanisms," in *Advanced Lectures in Quantitative Economics* edited by R. van der Ploeg (New York: Academic Press), ch. 6, pp. 213–248.

P.J. Hammond (1992), "On the Impossibility of Perfect Capital Markets," in *Economic Analysis of Markets and Games: Essays in Honor of Frank Hahn* edited by P. Dasgupta, D. Gale, O. Hart, and E. Maskin (Cambridge, Mass.: M.I.T. Press), pp. 527–560.

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

J.C. Harsanyi (1967–8), "Games with Incomplete Information Played by 'Bayesian' Players, I–III," *Management Science*, **14**: 159–182, 320–334, 486–502.

O.D. Hart and J. Tirole (1988), "Contract Renegotiation and Coasian Dynamics," *Review of Economic Studies*, **55**: 509–540.

R.M. Hogarth and M.W. Reder (eds.) (1987), *Rational Choice: The Contrast between Economics and Psychology*. Chicago: University of Chicago Press.

J. Howard (1988), "A Social Choice Rule and its Implementation in Perfect Equilibrium," London School of Economics, STICERD Discussion Paper No. TE/88/172.

L. Hurwicz (1972), "On Informationally Decentralized Systems," in C.B. McGuire and R. Radner (eds.), ch. 14, pp. 297–336.

L. Hurwicz (1979), "Outcome Functions Yielding Walrasian and Lindahl Allocations at Nash Equilibrium Points," *Review of Economic Studies*, **46**: 217–225.

L. Hurwicz (1986), "On the Implementation of Social Choice Rules in Irrational Societies," in *Social Choice and Public Decision Making: Essays in Honor of Kenneth J. Arrow, Vol. I* edited by W.P. Heller, R.M. Starr and D.A. Starrett (Cambridge: Cambridge University Press), ch. 4, pp. 75–96.

M.O. Jackson (1991), "Bayesian Implementation," *Econometrica*, **59**: 461–477.

J.S. Jordan (1993), "Three Problems in Learning Mixed-Strategy Nash Equilibria," *Games and Economic Behavior*, **5**: 368–386.

D. Kahneman, P. Slovic and A. Tversky (eds.) (1982), *Judgment under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

E. Kalai and E. Lehrer (1993), "Rational Learning Leads to Nash Equilibrium," *Econometrica*, **61**: (in press).

E. Kalai and W. Stanford (1988), "Finite Rationality and Interpersonal Complexity in Repeated Games," *Econometrica*, **56**: 397–410.

A. Kirman (1983), "On Mistaken Beliefs and Resultant Equilibria," in R. Frydman and E.S. Phelps (eds.) (1983), ch. 8, pp. 147–166.

D. Kreps and R. Wilson (1982), "Sequential Equilibrium," *Econometrica*, **50**: 863–894.

P. Kumar (1985), "Essays on Intertemporal Incentives and Signalling," Ph.D. dissertation, Stanford University, Department of Economics.

M. Kurz (1988), "Fundamental Difficulties in Learning the Equilibrium Price Process of a Complex Economy," preprint, Stanford University, Department of Economics.

J.-J. Laffont and E. Maskin (1982), "The Theory of Incentives: An Overview," in *Advances in Economic Theory* edited by W. Hildenbrand (Cambridge: Cambridge University Press), ch. 2, pp. 31–94.

J.-J. Laffont and J. Tirole (1987), "Comparative Statics of the Optimal Dynamic Incentive Contract," *European Economic Review*, **31**: 901–926.

J.-J. Laffont and J. Tirole (1988a), "The Dynamics of Incentive Contracts," *Econometrica*, **56**: 1153–1175.

J.-J. Laffont and J. Tirole (1988b), "Adverse Selection and Renegotiation in Procurement," preprint, Massachusetts Institute of Technology.

D.K. Lewis (1969), *Convention: A Philosophical Study*. Cambridge, Mass.: Harvard University Press.

R.D. Luce and H. Raiffa (1957), *Games and Decisions: Introduction and Critical Survey*. New York: John Wiley.

C.-T. Ma (1988), "Unique Implementation of Incentive Contracts with Many Agents," *Review of Economic Studies*, **55**: 555–572.

C.-T. Ma, J. Moore and S. Turnbull (1988), "Stopping Agents from 'Cheating'," *Journal of Economic Theory*, **46**: 355–372.

P.H. McAllister (1988), "Rational Behavior and Rational Expectations," Ph.D. dissertation, Stanford University, Department of Economics.

C.B. McGuire and R. Radner (eds.) (1972, 1986), *Decision and Organization (2nd edn.)*. Minnesota: University of Minnesota Press.

R.D. McKelvey (1989), "Game Forms for Nash Implementation of General Social Choice Correspondences," *Social Choice and Welfare*, **6**: 139–156.

J.M. Malcomson and F. Spinnewyn (1988), "The Multi-Period Principal-Agent Problem," *Review of Economic Studies*, **55**: 391–408.

A. Marcet and T.J. Sargent (1988), "The Fate of Systems with 'Adaptive' Expectations," *American Economic Review, Papers and Proceedings*, **78**: 168–172.

A. Marcet and T.J. Sargent (1989), "Convergence of Least Squares Learning Mechanisms in Self-Referential Linear Stochastic Models," *Journal of Economic Theory*, **48**: 337–368.

E.S. Maskin (1977), "Nash Equilibrium and Welfare Optimality," preprint, Massachusetts Institute of Technology.

E.S. Maskin (1985), "The Theory of Implementation in Nash Equilibrium: A Survey," in *Social Goals and Social Organization: Volume in Memory of Elisha Pazner* edited by L. Hurwicz, D. Schmeidler and H. Sonnenschein (Cambridge: Cambridge University Press), pp. 173–204.

J.-F. Mertens and S. Zamir (1985), "Formalization of Bayesian Analysis of Games with Incomplete Information," *International Journal of Game Theory*, **14**: 1–29.

D. Mookherjee (1984), "Optimal Incentive Schemes with Many Agents," *Review of Economic Studies*, **51**: 433–446.

D. Mookherjee and S. Reichelstein (1990), "Implementation via Augmented Revelation Mechanisms," *Review of Economic Studies*, **57**: 453–475.

J. Moore and R. Repullo (1988), "Subgame Perfect Implementation," *Econometrica*, **56**: 1191–1220.

J. Moore and R. Repullo (1990), "Nash Implementation: A Full Characterization," *Econometrica*, **58**: 1083–1099.

H. Moulin (1979), "Dominance Solvable Voting Schemes," *Econometrica*, **47**: 1337–1351.

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

R.B. Myerson (1982), "Optimal Coordination Mechanisms in Principal-Agent Problems," *Journal of Mathematical Economics*, **11**: 67–81.

Y. Nyarko (1992a), "Bayesian Learning in Repeated Games Leads to Correlated Equilibria," preprint, New York University.

Y. Nyarko (1992b), "Bayesian Learning without Common Priors and Convergence to Nash Equilibrium," preprint, New York University.

F. Page (1987), "The Existence of Optimal Contracts in the Principal-Agent Model," *Journal of Mathematical Economics*, **16**: 157–167.

F. Page (1988), "Optimal Contract Mechanisms for Generalized Principal-Agent Problems," preprint, Indiana University, Department of Finance.

T. Palfrey and S. Srivastava (1987), "On Bayesian Implementable Allocations," *Review of Economic Studies*, **54**: 193–208.

T. Palfrey and S. Srivastava (1991), "Nash Implementation Using Undominated Strategies," *Econometrica*, **59**: 479–501.

C. Papadimitriou (1989), "On Players with a Bounded Number of States," preprint, University of California at San Diego.

D. Pearce (1984), "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica*, **52**: 1029–1050.

C.R. Plott and S. Sunder (1988), "Rational Expectations and the Aggregation of Diverse Information in Laboratory Security Markets," *Econometrica*, **56**: 1085–1118.

A. Postlewaite and D. Schmeidler (1986), "Implementation in Differential Information Economies," *Journal of Economic Theory*, **39**: 14–33.

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

E.C. Prescott and R.M. Townsend (1984b), "General Competitive Analysis in an Economy with Private Information," *International Economic Review*, **25**: 1–20.

R. Radner (1975), "Satisficing," *Journal of Mathematical Economics*, **2**: 253–262.

R. Radner and M. Rothschild (1975), "On the Allocation of Effort," *Journal of Economic Theory*, **10**: 358–376.

R. Rahi (1993), "Information Revelation in Financial Markets," Ph.D. dissertation, Stanford University, Department of Economics.

A. Rubinstein (1986), "Finite Automata Play the Repeated Prisoner's Dilemma," *Journal of Economic Theory*, **38**: 83–96.

A. Rubinstein (1987), "The Complexity of Strategies and the Resolution of Conflict: An Introduction," London School of Economics, STICERD Discussion Paper No. TE/87/150.

A. Rubinstein (1988), "Comments on the Interpretation of Game Theory," London School of Economics, STICERD Discussion Paper No. TE/88/181 (Walras-Bowley Lecture to the Econometric Society).

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

D. Samet (1988), "Ignoring Ignorance and Agreeing to Disagree," Northewestern University, Kellogg Graduate School of Management, MEDS Discussion Paper No. 749.

M.A. Satterthwaite (1975), "Strategy-Proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions," *Journal of Economic Theory*, **10**: 187–217.

R. Selten (1975), "Re-examination of the Perfectness Concept for Equilibrium Points of Extensive Games," *International Journal of Game Theory*, **4**: 25–55.

H.M. Shefrin (1981), "Games with Self-Generating Distributions," *Review of Economic Studies*, **48**: 511–519.

H.A. Simon (1972), "Theories of Bounded Rationality," in C. B. McGuire and R. Radner (1972), ch. 8, pp. 161–176.

H.A. Simon (1982), *Models of Bounded Rationality (2 vols.)*. Cambridge, Mass.: M.I.T. Press.

H.A. Simon (1986), "Rationality in Psychology and Economics," *Journal of Business*, **59 (Supplement)**: 25–40, reprinted in Hogarth and Reder (1987).

H.A. Simon (1987a), "Bounded Rationality," in J. Eatwell, M. Milgate and P. Newman (eds.) (1987).

H.A. Simon (1987b), "Satisficing," in J. Eatwell, M. Milgate and P. Newman (eds.) (1987).

V.L. Smith, G. Suchanek and A.W. Williams (1988), "Bubbles, Crashes, and Endogenous Expectations in Experimental Spot Asset Markets," *Econometrica*, **56**: 1119–1151.

J. Strnad (1987), "Full Nash Implementation of Neutral Social Choice Functions," *Journal of Mathematical Economics*, **16**: 17–37.

T.C.-C. Tan and S.R.da C. Werlang (1988), "The Bayesian Foundations of Solution Concepts of Games," *Journal of Economic Theory*, **45**: 370–391.

R.M. Townsend (1979), "Optimal Contracts and Competitive Markets with Costly State Verification," *Journal of Economic Theory*, **21**: 265–293.

R.M. Townsend (1987), "Economic Organization with Limited Communication," *American Economic Review*, **77**: 954–971.

R.M. Townsend (1988), "Information Constrained Insurance: The Revelation Principle Extended," *Journal of Monetary Economics*, **21**: 411–450.

A. Tversky and D. Kahneman (1986), "Rational Choice and the Framing of Decisions," *Journal of Business*, **59 (Supplement)**: 67–94, reprinted in Hogarth and Reder (1987).

J.W. Vaupel (1986), "Un pensiero analitico per decisori indaffarati," in *La decisione: Razionalità collettiva e strategia nell'amministrazione e nelle organizzazioni* edited by L. Sacconi (Milano: Franco Angeli), ch. 13, pp. 226–238.

F. Vega-Redondo (1988), "A Model of Bounded Rationality in a General Decision Framework," paper presented to the Econometric Society European Meeting, Bologna.

S. Williams (1984), "Sufficient Conditions for Nash Implementation," University of Minnesota, Institute for Mathematics and its Applications, Preprint Series No. 70.

S. Williams (1986), "Realization and Nash Implementation: Two Aspects of Mechanism Design," *Econometrica*, **54**: 139–151.

G.C. Winston (1989), "Imperfectly Rational Choice: Rationality as a Result of a Costly Activity," *Journal of Economic Behavior and Organization*, **12**: 67–86.