R. de Econometria | Rio de Janeiro | v. VI nº 1 | p. 41 - 78 | abr 1986

BAYESIAN FOUNDATIONS OF NASH EQUILIBRIUM BEHAVIOUR

Sérgio Ribeiro da Costa Werlang*

ABSTRACT

This is part of a larger project to investigate the Bayesian foundations of non-cooperative solution concepts. Elsewhere Bernheim and Pearce prove that common knowledge of Bayesian rationality is not enough to justify the non-cooperative solution concept defined by Nash. Here several alternative behavioural assumptions are considered. In general the coordination required to achieve a Nash equilibrium is very strong. Not only Bayesian rationality, but also the actions taken, have to be common knowledge. For particular kinds of games the coordination required is not as strong.

* IMPA and EPGE/FGV. This is part of the author's Ph.D. dissertation at Princeton University. The author acknowledges financial support from CNPq.

1. INTRODUCTION

Until the mid seventies there was very little disagreement concerning the appropriateness of Nash equilibria as a solution concept for non-cooperative games. To a large extent it was believed that for an arbitrary game all Nash equilibria were reasonable outcomes and that any non-Nash set of actions was unreasonable. Today the situation is very different. A literature has developed that deals with refinements of Nash and different game theorists believe in different refinements. Also, there are those that hold that non-Nash actions may be perfectly reasonable. In order to help to clarify the issues involved in arguing for refinement or relaxation of the Nash equilibrium concept one is lead to an investigation of the logical basis for solution concepts in general. We start with the idea that a solution concept should be based on assumptions regarding Bayesian rationality, what is known, what is common knowledge, and behavioural norms.

This essay is part of a larger project whose aim is to provide foundations for different non-cooperative solution concepts. To illustrate the point that the knowledge and common knowledge of certain characteristics of the players plays a central role in the choice of the solution concept, let us imagine that you are going to play a given bimatrix game against two alternative partners. The payoffs of the game are in dollar terms. The first player is an intelligent acquaintance of yours, whom you know very well. The second player is a stranger. He comes from the Himalayas, and the only relevant information you know about him is that he was taught the meaning of a bimatrix game (the rules of the game) and what a dollar can buy. For the sake of argument, let us say that: (i) there is a unique pure strategy Nash equilibrium, which gives you a thousand dollars; (ii) your security level is nine hundred dollars; and (iii) if the other player does not play his part of the Nash equilibrium you get at most live hundred dollars. How should you play against the two different opponents? It seems clear to me that everyone who is faced with this situation is much more likely to follow the Nash strategy when facing the acquaintance, than when facing the stranger. Thus a well defined game may be played in different ways by the same person. This fact indicates that, in the specification of a game, some additional information about the background of the players is essential for the solution of this game. By explicitly modelling knowledge and common knowledge of different attributes, for example, of Bayesian rationality, one obtains which solution concept is suitable for each situation considered.

In order to get started one has to understand formally the notion of common knowledge. Suppose that there are n players. One says that a statement is common knowledge if everyone knows it, everyone knows that everyone knows it, ..., everyone knows (everyone knows)$^{m-1}$ it, and so on, for all m. The framework that is suitable for study of common knowledge in a game is based on the idea of an infinite hierarchy of beliefs (Armbruster and Böge (1979), Böge and Eisele (1979), Mertens and Zamir (1985)).

Section two contains a mathematical description of this formalism. Tan and Werlang (1985) apply this formalism to common knowledge.

Section three begins with a discussion of games and solution concepts. Bernheim (1984) and Pearce (1984) show that common knowledge of rationality is not enough to justify Nash behaviour. They introduce a non-cooperative solution concept which is derived from the hypothesis that Bayesian rationality is common knowledge. They call their solution concept rationalisable strategic behaviour. The point I wish to emphasise is that Bernheim and Pearce derive their solution concept for games from assumptions about the behaviour of the players. One can quite generally approach the analysis of solution concepts in the same manner. Which are the implicit behavioural assumptions behind a given solution concept? From a Bayesian point of view, the decision of each player in a game is determined by this player's beliefs about the actions of other players. But, if, in their turn, other players' beliefs about other players' actions affect their own actions, then it must be that the beliefs one player has about the beliefs of other players also affect the decision of this player in the game. If we carry this argument further, we see that the action taken by a player is determined by his infinite hierarchy of beliefs about actions of other players. The space of these infinite hierarchies of beliefs is the appropriate space for the study of behavioural assumptions about the players. Section three deals with this matter in detail, and poses formally the relationship between solution concepts and behavioural assumptions implicit in them. As a first illustration of this framework, Tan and Werlang (1984) discuss the solution concept given by rationalisable strategic behaviour.

The fourth section deals with Nash equilibrium behaviour. It starts by formally stating a justification for Nash behaviour which is closely related to the classical one. Not only rationality should be taken as common knowledge, but also the actions to be chosen. This allows one to see how strongly coordinated the players have to be. When one relaxes this hypothesis slightly, everything breaks down. Another behavioural assumption is studied: that each player "knows" the other players. When the game has two players, Armbruster and Böge (1979) proved that this yields Nash equilibrium beliefs. It is shown that this is false in the case of three (and consequently more than two) players. Finally, we provide theorems that generalise the results which justify Nash behaviour.

2. MATHEMATICS OF INFINITE HIERARCHIES OF BELIEFS

This section is aimed at giving the basic results on infinite recursions of beliefs (also called hierarchies of beliefs), the essential mathematical tool needed in the text. Let S be a compact metric space. From now on we will concentrate only on this case: all our spaces are compact and metric. This topology shall, furthermore, reflect the economic situation to be analysed. For example, Milgrom and Weber (1979) derive topologies which are relevant for games with incomplete information. Define the set of probability measures over S endowed with the Borel σ-algebra, as Δ(S). A natural topology over this set is the weak convergence of measures (see Billingsley (1968) and Hildenbrand (1974)). The main result is:

2.1. Theorem

S compact and metric with the Borel σ-algebra. If Δ(S) denotes the set of probability measures on S, and is endowed with the topology of the weak convergence of measures, then Δ(S) is compact and metric.

Proof. This theorem follows from Bilingsley (1968, pp. 238-240, Theorems 5 and 6).

Q.E.D.

The formal framework to be developed appeared before in Armbruster and Böge (1979), Böge and Eisele (1979), Mertens and Zamir (1985) and Myerson (1983).

Let the the set of possible states of nature, as perceived by agent i, be represented by a compact and metric set S_{0i}. In sections 3 and 4, since we are interested in games of complete information, $S_{0i} = A_{-i}$. In other problems the correct specification of these spaces is fundamental.

Given these sets of states of nature, agent i has subjective beliefs about the occurrence of a state in S_{0i}. This subjective belief is the first order belief, $s_{1i} \in \Delta(S_{0i})$. Set $S_{1i} = \Delta(S_{0i})$. The second order beliefs will be beliefs about beliefs of other agents. However, it is also possible to consider the possibility of these being correlated with agent i's beliefs about the states of nature he perceives.

Therefore $s_{2i} \in \Delta(S_{0i} \times \Pi_{k\neq i}S_{1k}) = S_{2i}$. Inductively, we define the m-th order beliefs of agent i as $s_{mi} \in \Delta(S_{0i} \times \Pi_{k\neq i}S_{m-1,k}) = S_{mi}$.

Notice that we could have modelled S_{mi} to include correlation among all the previous layers of beliefs. This is the approach followed by Mertens and Zamir (1985, p. 7, Th. 2.9), but given the consistency requirements they impose (as we do below), their framework is equivalent to the one we use here.

Observe that an arbitrary m-th order belief contains information about all beliefs of order less than m. An obvious requirement that should hold is that the first order belief of agent i should be the marginal of his second-order belief on his basic uncertainty space. Given a probability distribution $Q \in \Delta(C \times D)$, one defines the marginal of Q on C, and writes $\text{marg}_C[Q]$, as the following probability defined over C: given any measurable X contained in C, $\text{marg}_C[Q](X) = Q(X \times D)$. We will construct a way of determining the lower-order beliefs, given a belief of a certain order. This is the approach of Myerson (1983, 1984). We include Myerson's proof for completeness. Let us impose on a agent's beliefs the <u>minimal consistency requirement</u>: that if it is possible to evalue the probability of an event through his m-th order beliefs and through his p-th order beliefs, with $m \neq p$, then both probability assessments agree. Define inductively the functions which will recover the $(m-1)^{-th}$ the order beliefs, given the m-th order beliefs, by:

(i) for $m \geq 2$, $\Psi_{m-1,i}: S_{mi} \to S_{m-1,i}$;

(ii) if $m = 2$, $\Psi_{1i}(s_{2i})(E) = s_{2i}(E \times \Pi_{k\neq i}S_{1k})\ \forall\ E$ contained in S_{0i}. As was said before, the first order belief is simply the marginal of the second;

(iii) if $m \geq 3$, by induction on m we assume $(\Psi_{m-2,k})_{k\in N}$ defined,

and: for all E contained in $S_{0i} \times \Pi_{k\neq i}S_{m-2,k}$, $\Psi_{m-1,i}(s_{mi})(E) =$

$= s_{mi}(\{(s_{0i}, (s_{m-1,k})_{k\neq i}) \in S_{0i} \times \Pi_{k\neq i}S_{m-1,k} \mid (s_{0i}, (\Psi_{m-2,k}(s_{m-1,k}))_{k\neq i}) \in E\})$.

We then have:

2.2 Proposition

Suppose all agents are aware that each of them satisfies the minimum consistency requirement. Then $\forall i$, $\forall m \geq 2$: $\Psi_{m-1,i}(s_{mi}) = s_{m-1,i}$.

Proof The proof goes by induction. For m = 2, let E be contained in S_{0i}. Then the event E (event = measurable set) is evaluated by s_{1i} as $s_{1i}(E)$. However E is the same as $E \times \Pi_{k\neq i} S_{1k}$ evaluated by s_{2i}. Therefore, by the consistency requirement: $s_{1i}(E) = s_{2i}(E \times \Pi_{k\neq i} S_{1k}) = \Psi_{1i}(s_{2i})(E)$. Thus $s_{1i} = \Psi_{1i}(s_{2i})$. Agents also know that $s_{1k} = \Psi_{1k}(s_{2k})$, because other agents are also consistent. The first step of the induction process is proved. Let us assume it is true for $m \geq 2$. We will prove it is true for $m + 1$. If it is true for m, we know that $\forall k \in N (N=\{1,...,n\})$: $\Psi_{m-1,k}(s_{mk}) = s_{m-1,k}$. Given the event E contained in $S_{0i} \times \Pi_{k\neq i} S_{m-1,k}$, define E* contained in $S_{0i} \times \Pi_{k\neq i} S_{mk}$ by:

$$E^* = \{(s_{0i},(s_{mk})_{k\neq i}) \in S_{0i} \times \Pi_{k\neq i} S_{mk} \mid (s_{0i},(\Psi_{m-1,k}(s_{mk}))_{k\neq i}) \in E\}.$$

By the induction hypothesis we have that

$$E^* = \{(s_{0i},(s_{mk})_{k\neq i}) \in S_{0i} \times \Pi_{k\neq i} S_{mk} \mid (s_{0i},(s_{m-1,k})_{k\neq i}) \in E\}.$$

Therefore E* and E are the same events (same in the sense used before: one is true if and only if the other is). Hence by the minimum consistency requirement $s_{mi}(E) = s_{m+1,i}(E^*)$. But $\Psi_{mi}(s_{m+1,i})(E) = s_{m+1,i}(E^*)$, so that the result follows.

Q E D.

Given the proposition above, we will restrict ourselves to beliefs which satisfy the minimum consistency requirement. The set of all possible beliefs becomes, then:

$$S_i = \{(s_{1i}, s_{2i}, \ldots) \in \Pi_{m\geq 1} S_{mi} \mid \forall m: \Psi_{mk}(s_{m+1,k}) = s_{mk}\}.$$

The proposition below is proved in Armbruster and Böge (1979, p.19, Th. 4.2), Böge and Eisele (1979, p. 196, Th. 1), and Mertens and Zamir (1985, p. 7, Th. 2.9).

2.3 Proposition

S_i is compact and metric, in the topology induced by the product topology.

The proposition above just says that the space of characteristics of the agents is tractable. The proof of the proposition is simple. One has only to show that the functions Ψ are continuous.

Notice that one can look at the space of beliefs which are consistent and are of level up to m. By 2.2 the lower-order beliefs are entirely determined by those of the highest order. Then this is the same as the space S_{mi}. Moreover, there is an immediate way of recovering the lowest-order beliefs: just apply successively the functions Ψ. The most important result of this section states this for the case where we consider the whole stream of beliefs. The result is proved in Armbruster and Böge (1979, p. 19, Th. 4.2), Böge and Eisele (1979, p. 196, Th. 1), Mertens and Zamir (1985, p. 7, Th. 2.9) and Brandenburger and Dekel (1985, p. 10, Th. 3.2). The first proof of this result seems to have appeared in Böge (1974).

2.4 Theorem

$\forall i$ there exists $\Phi_i : S_i \to \Delta(S_{0i} \times \Pi_{j\neq i} S_j)$, which is a homeomorphism. Throughout the essay we will be referring to the homeomorphism shown in the proof below.

Sketch of the proof (essentially taken from Brandenburger and Dekel (1985)) Let $Y_i = \Pi_{m\geq 1} S_{mi}$. Then $Y_i \supset S_i = \{(s_{mi})_{m\geq 1} \mid \forall m\geq 1: s_{mi} = \Psi_{mi}(s_{m+1,i})\}$. Suppose $s_i = (s_{1i}, s_{2i}, \ldots) \in S_i$. We will construct $\Phi_i(s_i)$. $S_{0i} \times \Pi_{j\neq i} Y_j = S_{0i} \times \Pi_{j\neq i} (\Pi_{m\geq 1} S_{mj}) = S_{0i} \times \Pi_{m\geq 1} (\Pi_{j\neq i} S_{mj})$, by exchanging the order of the Cartesian product. First we will exhibit $\Phi_i(s_i)$ as a probability measure in $S_{0i} \times \Pi_{j\neq i} Y_j$. Then we will prove that the support of $\Phi_i(s_i)$ is contained in $S_{0i} \times \Pi_{j\neq i} S_j$. To construct this measure one invokes Kolmogorov's extension theorem (see Dellacherie and Meyer (1978, p. 68, III. 51-52)). To construct a probability in a countably infinite product of spaces, it is necessary and sufficient to give all the finite dimensional marginals, provided these marginals are not contradictory. Furthermore, this probability is uniquely determined by these marginals. Given k, let q_k denote a probability defined on $S_{0i} \times \Pi_{1\leq m\leq k} (\Pi_{j\neq i} S_{mj})$. For k=0, let $q_0 = s_{1i}$. For $k + 1 \geq 1$, we define $q_{k+1} \in \Delta(S_{0i} \times \Pi_{1\leq m\leq k+1} (\Pi_{j\neq i} S_{mj}))$ in the sets which are of the form $E_{0i} \times E_{1,-i} \times \ldots \times E_{k,-i} \times E_{k+1,-i}$, where E_{0i} is an event in S_{0i}, $E_{1,-i}$ is an event in $\Pi_{j\neq i} S_{1j}$, and in general, $E_{m,-i}$ is an event in $\Pi_{j\neq i} S_{mj}$. This is enough to define the probability q_{k+1}. The functions Ψ_{mi}, which are given above, will be used.

$q_{k+1}(E_{0i} \times E_{1,-i} \times \ldots \times E_{k,-i} \times E_{k+1,-i}) =$

$s_{k+2,i}(\{(s_{0i}, (s_{k+1,j})_{j\neq i}) \in E_{0i} \times E_{k+1,-i} \mid (s_{0i}, (\Psi_{kj}(s_{k+1,j}))_{j\neq i}) \in E_{0i} \times E_{k,-i},$

$(s_{0i}, (\Psi_{k-1,j} \circ \Psi_{kj}(s_{k+1,j}))_{j\neq i}) \in E_{0i} \times E_{k-1,-i}$, and, successively,

$(s_{0i}, (\Psi_{1j} \circ \ldots \circ \Psi_{k-1,j} \circ \Psi_{kj}(s_{k+1,j}))_{j\neq i}) \in E_{0i} \times E_{1,-i}$. By the minimum consistency requirement q_k is the marginal of q_{k+1} on the first k coordinates. By Kolmogorov's theorem there exists a unique probability, which we call $\Phi_i(s_i)$, on the space $S_{0i} \times \Pi_{m\geq 1}(\Pi_{j\neq i} S_{mj})$ whose marginals are given by the q_k. We now have to show that the support of $\Phi_i(s_i)$ is in $S_{0i} \times \Pi_{j\neq i} S_j$. Let $(s_{kj}, s_{k+1,j})$ be in the support of $\Phi_i(s_i)$ in $S_{kj} \times S_{k+1,j}$. We will show that $s_{kj} = \Psi_{kj}(s_{k+1,j})$. By the construction of $\Phi_i(s_i)$, we have $(s_{kj}, s_{k+1,j})$ in the support of q_{k+1}. Define B_{km} to be the closed ball of radius $1/m$ around s_{kj}. Analogously, let $B_{k+1,m}$ be the closed ball of radius $1/m$ around $s_{k+1,j}$. As $(s_{kj}, s_{k+1,j})$ is in the support of q_{k+1}, this means that for every $m\geq 1$ the joint probability of $(s_{kj}, s_{k+1,j}) \in B_{km} \times B_{k+1,m}$ is greater than zero. Thus, by the definition of q_k, $\Psi_{kj}(B_{k+1,m}) \cap B_{km}$ is nonempty for all $m\geq 1$. These sets are also compact (because Ψ_{kj} is continuous), so that the intersection over all m is nonempty. But by the definition of B_{km} only one point could be in this infinite intersection: s_{kj}. In the same manner the infinite intersection of $\Psi_{kj}(B_{k+1,m})$ for all $m\geq 1$ can consist only of $\Psi_{kj}(s_{k+1,j})$. Thus, $s_{kj} = \Psi_{kj}(s_{k+1,j})$, and the support of $\Phi_i(s_i)$ is in the space $S_{0i} \times \Pi_{j\neq i} S_j$. Therefore the function Φ_i can be viewed as $\Phi_i : S_i \to \Delta(S_{0i} \times \Pi_{j\neq i} S_j)$. To check that Φ_i is a bijection between these two spaces, suppose we are given a probability q on the space $S_{0i} \times \Pi_{j\neq i} S_j$. The point $s_i = (s_{1i}, s_{2i}, \ldots) \in S_i$ such that $q = \Phi_i(s_i)$, can be obtained simply by taking marginals on S_{0i}, $S_{0i} \times \Pi_{j\neq i} S_{1j}$, ..., $S_{0i} \times \Pi_{j\neq i} S_{kj}$, ..., for all $k\geq 1$. The sequence $(s_{1i}, s_{2i}, \ldots)$ thus obtained satisfies the minimum consistency requirement, because the probability q has its support on $S_{0i} \times \Pi_{j\neq i} S_j$. Finally, we have to check that Φ_i is a homeomorphism. However, by 2.3 S_i is compact and metric.

Thus, by 2.1 so is $\Delta(S_{0i} \times \Pi_{j \neq i} S_j)$. Therefore it is enough to prove that either Φ_i or $[\Phi_i]^{-1}$ is continuous. The continuity of $[\Phi_i]^{-1}$ is easily checked, because it is composed of marginals, and these are continuous (in the weak topologies, as we have in the spaces here).

Q E D.

Another way to view this result is by noting that any infinite stream of beliefs can be seen as a belief about the realisation of agent i's uncertainty and the characteristics of the other agents. One may interpret s_i as the agent himself/herself: it is a "psychology", or "type", of agent i. The S_i's are also know as "psychology spaces", or as "type spaces".

3. FOUNDATIONS OF NON-COOPERATIVE SOLUTION CONCEPTS

We begin by repeating the example given in the introduction. Imagine that you are going to play a given bimatrix game with two alternative players. The payoffs of the game are in dollar terms. The first player is an intelligent acquaintance of yours, whom you know very well. The second player is a stranger. He comes from the Himalayas, and the only relevant information you know about him is that he was taught the meaning of a bimatrix game (the rules of the game) and what a dollar can buy. For the sake of argument, let us say that: (i) there is a unique pure strategy Nash equilibrium, which gives you a thousand dollars; (ii) your security level is nine hundred dollars; and (iii) if the other player does not play his/her part of the Nash equilibrium you lose at least a hundred dollars. How should you play the same game against the two different opponents? It seems clear to me that everyone who is faced with this situation is much more likely to follow the Nash strategy when facing the acquaintance than when facing the stranger.

The fact that a well defined game may be played in different ways by the same person indicates that in the specification of a game some additional information about the background of the players is essential for the solution of this game. Therefore, a solution concept which depends only on the payoff matrix, as Nash equilibrium does, needs an interpretation. This points out the need for foundational analysis of solution concepts. In this section we establish a general methodology for the analysis of solution concepts. In particular, we point out which extra information the players need to play a game. In sections four and five we apply this methodology.

A fairly widespread notion among theorists these days is the lack of a non-cooperative solution concept (not to speak of cooperative ones) which will satisfactorily describe the behaviour of players in a game[1]. The main non-cooperative solution concept is that of Nash equilibrium. The literature on non-cooperative solution concepts follows two distinct trends. One states that there are too many Nash equilibria, and suggests alternative ways of refining Nash's notion. Some of the most important refinements of the Nash equilibrium concept are: subgame perfection, Selten (1965); perfection, Selten (1975); properness, Myerson (1978); sequentiality, Kreps and Wilson (1982); tracing procedure, Harsanyi (1975) and Harsanyi and Selten (1980-84); persistent equilibrium, Kalai and Samet (1982); strategic stability, Kohlberg and Mertens (1985); justifiable beliefs, McLennan (1985); forward induction equilibrium, Cho (1985); perfect sequential equilibrium, Grossman and Perry (1985). In the context of signalling games other refinements were proposed: intuitive criterion, Kreps (1985); divinity, Banks and Sobel (1985); neologism-proof, Farrell (1985). For a more complete

[1] Binmore in two delightful recent papers, Binmore (1984, 1985) discusses issues which are closely related to those dealt with here. See also Reny (1985) and Basu (1985).

account of the literature on refinements of Nash equilibrium, see van Damme (1983). A discussion of the recent progress in the area can be found in Kohlberg and Mertens (1985) and in Cho (1985).

The other trend followed by the literature is that of expanding the concept of Nash equilibrium. The important contributions in the area are: correlated equilibrium, Aumann (1974, 1985); refinements of correlated equilibrium, Myerson (1985); and rationalisable strategic behaviour, Bernheim (1984) and Pearce (1984).

In view of the many existing suggestions, as the list above exemplifies, which solution concept should be chosen to solve a specific game? We will not answer this question directly. Alternatively, we propose a methodology which enables us to analyse solution concepts. Bernheim (1984) and Pearce (1984) show that common knowledge of rationality is not enough to justify Nash behaviour. They introduce a non-cooperative solution concept which is derived from the hypothesis that Bayesian rationality is common knowledge. They call their solution concept rationalisable strategic behaviour. The point I wish to emphasise is that Bernheim and Pearce derive their solution concept for games from assumptions about the behaviour of the players. One can generally approach the analysis of solution concepts in the same manner. Which are the implicit behavioural assumptions behind a given solution concept?

In order to answer this question, let us be more precise. We shall be concerned with the following complete information simultaneous game:

3.1 Definition A game with n players, u, is a 2n-tuple

$(A_1, \ldots, A_n, u_1, \ldots, u_n)$, where:

(i) each A_i, the set of strategies or actions available to player i, is a compact metric space;

(ii) let $A = A_1 \times \ldots \times A_n$. Then $u_i : A \to R$, is a function which gives the payoff to player i, for each possible combination of strategies of all players. For each i, u_i is assumed to be continuous.

Let U mean the set of all n-tuples of payoff functions. By an abuse of notation, we say $u \in U$ is a given game, where u represents the n-tuple $(u_1, \ldots, u_n)$.

3.2 Definition A solution concept (also called equilibrium notion) is a correspondence $\Gamma : U \to A$.

A solution concept is a correspondence that assigns to each game a set of prescribed action profiles. The interpretation given to a solution of a game u, is that among the n-tuples of actions in $\Gamma(u)$, there is one which will be chosen when players play the game u. Hence, we are allowing that different players playing the same game could choose different n-tuples of actions, as long as these n-tuples are in $\Gamma(u)$.

From a Bayesian point of view, the decision of each player in a game is determined by this player's beliefs about the actions of other players. But, if, in their turn, other players' beliefs about other players' actions affect their actions, then it must be that the beliefs one player has about the beliefs of other players also affect the decision of this player in the game. If we carry this argument further, we see that the action taken by a player is determined by his infinite hierarchy of beliefs about actions of other players, beliefs about beliefs about other players' actions, and so on. The space of these infinite hierarchies of beliefs is the appropriate space for the study of behavioural assumptions about the players. For every i, let $A_{-i} = A_1 \times \ldots \times A_{i-1} \times A_{i+1} \times \ldots \times A_n$. We follow the mathematical formalism of section 2. The set of states of the world, as perceived by player i, is A_{-i}. A first order belief of player i is a point $s_{1i} \in \Delta(A_{-i}) = S_{1i}$. In general, the m-th order belief of player i is a point $s_{mi} \in \Delta(A_{-i} \times \Pi_{k \neq i} S_{m-1,k}) = S_{mi}$. As in section 2, we impose the minimum consistency requirement. Thus, by Theorem 2.4, the infinite hierarchy of beliefs s_i can be viewed as a joint belief about what other players play, and which hierarchies of beliefs other players have. This is done through the homemorphisms Φ_i. As before, the infinite hierarchy of beliefs s_i is interpreted as the "psychology" (or "type") of player i. It embodies all relevant decision-theoretic variables which are necessary for understanding how player will play the game u. The space of these psychologies, S_i, is the appropriate space for the study of behavioural assumptions about player i.

We have to determine how each different psychology s_i will play the game u. First we define a general Bayesian decision problem.

3.3 Definition A Bayesian decision problem for player i is given by:

(i) T_i a compact metric probability space endowed with the Borel σ-algebra. It represents all the elements of uncertainty for player i;

(ii) A_i a compact set of actions available to player i;

(iii) $U_i : A_i \times T_i \to R$, his subjective utility function;

(iv) $Q_i \in \Delta(T_i)$, his subjective prior on S_i.

Given a decision problem, one can derive the structure above from more basic facts as in Savage (1954): It is important to note that U_i and Q_i characterise player i.

Let $V_i : A_i \times \Delta(T_i) \to R$ be the expected subjective utility for player i, when he takes an action a_i, and has prior Q_i: $V_i(a_i, Q_i) = \int_{S_i} U_i(a_i, t_i)\, dQ_i(t_i)$. To avoid unnecessary notation, we will simply write $V(a_i, Q_i)$ instead of $V_i(a_i, Q_i)$.

<u>3.4 Definition</u> Player i is <u>Bayesian rational</u> when, faced with a Bayesian decision problem, he chooses an action $\tilde{a}_i \in A_i$ such that the expected subjective utility is maximised: $V(\tilde{a}_i, Q_i) \geq V(a_i, Q_i)$, $\forall a_i \in A_i$.

To determine how a given psychology s_i plays the game u, we have:

<u>3.5 Definition</u> Given a game u and psychology s_i, we define

The <u>Bayesian decision problem associated with u and s_i</u> as:

(i) $T_i = A_{-i} \times S_{-i}$, where $S_{-i} = \Pi_{k \neq i} S_k$;

(ii) A_i is the same as A_i for the game u;

(iii) $U_i\ (a_i, t_i) = u_i\ (a_i, \mathrm{Proj}_{A_{-i}}(t_i))$;

(iv) $Q_i \in \Delta(T_i)$ is given by $\Phi_i(s_i)$.

The viewpoint of this approach to game theoretical situations can be summarised by:

<u>3.6 Axiom</u> The decision problem player i faces in the game u when its psychology is s_i, is the same as the Bayesian decision problem associated with u and s_i. In other words: all that we need to know about player i to determine her/his behaviour in the game II, is given by the psychology s_i.

Let us comment a little about the above. One sees that the only relevant probability distribution for player i is the first order belief $s_{1i} = \mathrm{marg}_{A_{-i}}[\Phi_i(s_i)]$, since in the expected value function $V(a_i, t_i)$ only the the projection in the actions of the other players is considered relevant. This seems to tell us that the only important part of s_i is the first order belief. Ultimately that is so. However, we cannot forget that higher order beliefs influence the lower order beliefs. By an abuse in notation, one defines $V(a_i, \Phi_i(s_i)) = V(a_i, s_i)$.

Now we are ready to state our methodology. Let us consider a given game u fixed. That is to say, we will concentrate on the correspondence Γ restricted to a singleton {u} contained in U. We are not interested in
berg and Mertens (1985). In the case of a fixed game u ε U, a solution concept Γ is simply a subset of $A = A_1 x \ldots x A_n$. Associated with Γ, we want to find a subset B(Γ) contained in $S_1 x \ldots x S_n$, a subset of the set of psychologies of all players, such that:

(i) $\forall (a_1, \ldots, a_n) \in \Gamma, \exists (s_1, \ldots, s_n) \in B(\Gamma)$ such that for every i: a_i is an action which maximises the subjective utility for player s_i (according to 3.5);

(ii) $\forall (s_1, \ldots, s_n) \in B(\Gamma), \exists (a_1, \ldots, a_n) \in \Gamma$ such that for every i: a_i is an action which maximises the subjective utility for player s_i (according to 3.5).

In words: the first statement says that any n-tuple of actions in the solution set Γ can be played by some psychology in B(Γ). Conversely, the second statement says that any psychology in B(Γ) can play an action in Γ. This means that for every solution Γ we associate a set of psychologies which corresponds to Γ. The set B(Γ) can be interpreted as a set of behavioural assumptions behind the solution concept Γ. It is important to notice that the set B(Γ) is not uniquely determined. Each different B(Γ) represents a different set of behavioural assumptions under which the solution concept Γ is justified. However, if $B_1(\Gamma)$ and $B_2(\Gamma)$ both satisfy (i) and (ii), so does $B_1(\Gamma) \cup B_2(\Gamma)$. Hence, there is a maximal set BM(Γ) satisfying (i) and (ii). This set is to be interpreted as the set of all behavioural assumptions which justify the solution concept Γ.

The ultimate aim of the methodology described here is to obtain sets $B(\Gamma)$ for every solution concept Γ. This would facilitate the selection of the solution concept: one should see which behavioural assumptions apply to the economic situation being modelled, and choose the solution concept accordingly. In the next section we consider Γ = Nash equilibrium.

4. NASH EQUILIBRIUM BEHAVIOUR

4.1 Coordination and Nash Behaviour

In this section we consider the foundations for Nash equilibrium. We begin by exhibiting the result most theorists have in mind when they try to justify the use of Nash's non-cooperative solution concept. Then, using the framework of Section 3 one we will describe the assumptions that underlie the Nash solution correspondence.

We will be dealing with two alternative manners of interpreting the concept of Nash equilibrium. The first, the classical view, is that the players should choose a Nash action. The second, a subjective interpretation, is that every player can be Bayesian rational and believe that everyone else follow their Nash actions. Some Nash equilibria are such that the Nash actions are not unique best responses against the beliefs that the other players follow their Nash actions. Therefore, the subjective interpretation does not imply the classical interpretation. This point is exemplified in subsection 4.3. In subsections 4.1 and 4.2 we will focus on the classical interpretation of Nash equilibrium. In subsections 4.3 and 4.4 we will focus on the subjective interpretation.

For simplicity, in this subsection we will look at a selection (call it Γ_N) from the Nash solution correspondence. This function associates to every n-tuple of payoff functions a pure strategy Nash equilibrium. Obviously for this purpose we are looking at games where pure strategy Nash equilibria exist. We call a function with these properties a Nash theory.

The usual justification for the Nash equilibrium concept is that no player has any incentive to deviate from the action prescribed by the theory, if this player believes the other players are going to fulfill their rôle. This is expressed in the classical quote below, taken from Luce and Raiffa (1957, page 173):

"Nonetheless, we continue to have one very strong argument for equilibrium points: if our non-cooperative theory is to lead to an n-tuple of strategy choices, and if it is to have the property that knowledge of the theory does not lead one to make a choice different from that dictated by the theory, then the strategies isolated by the theory must be equilibrium points".

As one can see, this justification is a simple restatement of the definition of a Nash equilibrium. In this subsection we give an alternative interpretation to Nash equilibrium points. The Nash equilibria are the only n-tuples of actions which are consistent with common knowledge of the actions taken, as well as of rationality. If one takes a theory to be single-valued, then the Nash equilibria are the only n-tuples of actions which are consistent with common knowledge of the theory and of rationality.

Fix a game $u \in U$. The formalisation of the knowledge of a theory by the players, is simply the fact that the actions this theory predicts are the only actions which are considered possible by the players. The notation is the same as in sections three and four. In particular, if one wants to refer to "knowledge of a theory Γ", where Γ is contained in $A = A_1 \times \dots \times A_n$, we have:

<u>4.1.1 Definition</u> Given Γ contained in A, a theory, we say that <u>player i knows a theory Γ</u> when $s_i \in \Gamma_{1i} = \{s_i \in S_i \mid Proj_{A-i}\, \Gamma \supset supp\ marg_{A-i}[\phi_i(s_i)]\}$. In other words: player i knows a theory when he thinks other players are going to fulfill their rôle in this theory.

<u>4.1.2 Definition</u> <u>A theory Γ is common knowledge in the eyes of player i</u> if:

$s_i \in \bigcap_{m\geq 1}\Gamma_{mi}$, where: Γ_{1i} is given above, and

$\forall m\geq 2$: $\Gamma_{mi} = \{s_i\in\Gamma_{m-1,i} | \forall\ k\neq i$: $s_k \in$ supp $marg_{S_k}[\Phi_i(s_i)] \Rightarrow s_k \in \Gamma_{m-1,k}\}$.

The following theorem express this point in formal terms:

<u>4.1.3 Theorem</u> Assume that $\Gamma = \{(\tilde{a}_1, \ldots, \tilde{a}_n)\}$, that is to say, Γ is a single-valued theory. Suppose there exists i ε N such that rationality and the theory Γ are common knowlege in the eyes of player i. Then Γ is a Nash theory (that is to say: $(\tilde{a}_1,\ldots,\tilde{a}_n)$ is a Nash equilibrium of the game). Moreover, any Nash theory, Γ_N, is compatible with common knowlege of the theory and common knowledge of rationality.

<u>Proof</u> Since player i knows that player k is rational and player k knows the theory, it follows that a_k is a best response to $\tilde{a}_{-k}$, for all $k\neq i$. To check that $\tilde{a}_i$ is a best response to $\tilde{a}_{-i}$, it is enough to carry the same argument above one step further. Observe that it was necessary to use only $s_i \in K_{2i} \cap \Gamma_{3i}$. The second part of the theorem is immediate.

<u>Q E D</u>

The result above gives one set of behavioural assumptions which justifies the Nash equilibrium concept. This set of assumptions is the main thrust os Hash equilibrium. However, we feel that the theorem above also shows the weakness of the concept. In fact, the Nash equilibrium is played when the actions which are going to be taken are common knowledge, <u>before they have been taken</u>. It shows the strong need for coordination in obtaining Nash behaviour. This is the rôle played by several of the "stories" to justify Nash equilibrium behaviour: they are mere coordination mechanisms. Famous examples of these stories are the "book of Nash" and the "gentlemen's club". The former is well known. The latteris simply a revised version of the former: every player should belong to the same gentlemen's club, where the club's statute tells them how to behave in a game-theoretic situation. As they are gentlemen (and very possibly English), they all give their word of honour they will follow this statute (Binmore (1984)).

The main purpose of the rest of this section is to give alternative sets of behavioural assumptions under which Nash equilibrium behaviour is justified.

4.2 Common Knowledge that Players May Play Nash Equilibrium

If one restricts the class of games to be considered, the coordination mechanism required to achieve a Nash equilibrium may be very reasonable. Bernheim (1984, section 5) and Moulin (1984) give examples os classes of games for which the set of rationalisable strategies and Nash equilibrium strategies coincide. For these restricted classes of games, common knowledge of rationality is enough to ensure that a Nash equilibrium is played. An important game that belongs to this class is the Cournot duopoly with linear demand and constant marginal costs. However, this class of games is very restricted. If one considers the oligopoly above with three firms, instead of two, the result is not true any more: there is a continuum of rationalisable actions, while only one Cournot-Nash equilibrium.

4.2.1 Example (Cournot oligopoly with linear demand and constant marginal costs.) Let there be n identical firms, each of them with maximum capacity 10. Suppose marginal costs are constant and equal to 1. The market inverse demand function is given by $P(Q) = \max \{10-Q, 0\}$. The firms play with quantities in the fashion of Cournot. The strategy set of firm i is: $A_i = [0, 10]$, with generic element q_i. The payoffs are given by the profit functions $\Pi_i(q_1,...,q_n) = P(\Sigma q_k) \cdot q_i - q_i$. This game has a unique Cournot-Nash solution: all firms produce the quantity $q_i = 9/(n+1)$. When n=2 the only rationalisable action for a firm is the Cournot-Nash equilibrium $q_1 = 3 =$

= 9/3 (see Bernheim (1984) and Moulin (1984)). For n=3 the set of rationalisable actions for each firm is the interval [0, 9/2], that is to say, any quantity between zero and the monopoly level is rationalisable.

In this subsection we consider a weakening of the assumption that a Nash theory is common knowlege. We will assume that it is common knowledge that the players may play a Nash theory. At the same time, we maintain the assumption of common knowledge of rationality. Therefore, the class of games for which these two assumptions are a sufficient coordination mechanism to achieve Nash equilibrium, is potentially larger than the classes of games considered by Bernheim and Moulin. We will show that this new class of games is indeed larger than theirs.

4.2.2 Definition Given Γ contained in $A = A_1 \times \ldots \times A_n$, a theory, we say that player i knows that other players may play the theory Γ when: $s_i \in \Gamma P_{1i} = \{s_i \in S_i \mid \mathrm{Proj}_{A_{-i}} \Gamma \cap \mathrm{supp\ marg}_{A_{-i}} [\Phi_i(s_i)] \neq \emptyset\}$.

4.2.3 Definition (Knowledge and Common Knowledge that other players may play a Nash theory). Given a game u and $(\bar{a}_1, \ldots, \bar{a}_n)$ a Nash equilibrium (in pure strategies) of this game, we say that player i knows that other players may play it, if: $s_i \in N_{1i} = \{s_i \in S_i \mid \bar{a}_{-i} \in \mathrm{supp\ marg}_{A_{-i}} [\Phi_i(s_i)]\}$. In the same way we say that it is common knowledge in the eyes of player i that the Nash theory $(\bar{a}_1, \ldots, \bar{a}_n)$ may be played by other players, if $s_i \in \bigcap_{m \geq 1} N_{mi}$, where:

$$\forall m \geq 2: N_{mi} = \{s_i \in N_{m-1,i} \mid \forall k \neq i: s_k \in \mathrm{supp\ marg}_{S_k} [\Phi_i(s_i)] \Rightarrow s_k \in N_{m-1,k}\}.$$

The next proposition shows that the class of games for which the common knowledge of rationality and the common knowledge that players may play the Nash equilibrium is a sufficient coordination mechanism to attain Nash behaviour, is strictly larger than those classes of games provided by Bernheim and Moulin. We do this by showing that, in the Cournot oligopoly example seen above, when the number of firms is three, the common knowledge of rationality, as well as the common knowledge of the fact that the players may play their Cournot-Nash actions, yields the Cournot-Nash outcome.

4.2.4 Proposition Let the game be as in example 4.2.1, with $n = 3$. Assume that rationality is common knowledge and that the possibility of playing the Cournot equilibrium is also common knowledge Then, the only possible action taken by a rational firm is the Cournot-Nash equilibrium (which is $q_i = 9/4$).

Proof[2] The requirement above is that $s_i \in (\bigcap_{m \geq 1} K_{mi}) \bigcap (\bigcap_{m \geq 1} N_{mi})$ (*) and that every player is rational (the sets N_{mi} are generated according to definition 4.2.3, taking as $(\bar{a}_1, \ldots, \bar{a}_n)$ the triple - (9/4, 9/4, 9/4)). The first point to notice is that due to the symmetry of the game, it is enough to concentrate the analysis in one particular firm. We are going to show that the only action which is compatible with rational behaviour and condition (*) is 9/4. Rearranging (*): $s_i \in \bigcap_{m \geq 1} (K_{mi} \bigcap N_{m+1,i})$. This allows us to reinterpret the assumption of the theorem. For example, $K_{1i} \bigcap N_{2i}$ means not only that player i thinks k is rational, but also that any action k takes may be rationalised by beliefs which contain the Cournot actions of the other players in the support. One can easily see that (*) is verifyied if and only if: (i) i thinks the others may play (9/4, 9/4); and (ii) all actions i thinks k may take have to be rationalised by beliefs which contain the Cournot actions in the support, and using the symmetry of the game, every actionin this support has to be rationalised by beliefs which contain the Cournot actions in the support, and so on. Let us study what happens in each mental into action described above. Let q_i be an action which is a best response to a belief $\mu \in \Delta([0,10] \times [0,10])$. The first thing to notice is that $q_i \notin [9/2, 10]$. In fact, suppose not. One can check that the action 9/2 will give a higher payoff. Letus go to the second round (notice: we still have not used the fact that (9/4, 9/4) is in the support of μ). Then $[0,9/2)^2 \supset \text{supp}\,[\mu]$, from the analysis above. In this case the response function can be computed and it is:

$q_i(\mu) = (1/2).(9-E_\mu(\Sigma_{k \neq i} q_k))$.

[2] The symbol K_{mi} represents the set of psychologies of player i, for whose rationality is known up to level m. See Tan and Werlang (1984) for more details.

Let q_{inf} and q_{sup} be the infimum and the supremum of the support after the infinite recursion. By the above formula, for every q ε supp[μ]: $q \leq (1/2).(9-2q_{inf})$ (A), and $q \geq (1/2).(9-2q_{sup})$ (B). But the beliefs which support q must contain (9/4, 9/4) in the support. Thus inequality (A) must be strict if $q_{inf} \neq 9/4$, and inequality (B) in case $q_{sup} \neq 9/4$. Suppose one of the strict inequalities above holds, let us say (A). One can take q to be q_{sup} in (A) and q to be q_{inf} in (B), since the support is a closed set. Hence: $q_{sup} < 9/2 - q_{inf}$ and $q_{inf} \geq 9/2 - q_{sup}$. This is a contradiction. Thus $q_{inf} = q_{sup} = 9/4$, and the proposition follows.

Q E D.

The result above does not generalise. For the case of four firms we do not obtain the Cournot-Nash equilibrium as the only possible outcome:

4.2.4 Example (Common knowledge of rationality and of the possibility of a Nash theory being played is not enough to obtain Nash equilibrium). Consider the same game as above, with n=4. In this case the Nash equilibrium is (9/5, 9/5, 9/5, 9/5). We show, for example, that 0 can be an outcome in this game. This follows from the observations in the proof of the proposition above, plus the fact that:

(i) 0 is the best response to a belief which assigns probability 18/23 to (10/3, 10/3, 10/3), and 5/23 to actions (9/5, 9/5, 9/5);

(ii) 10/3 is the best response against a belief which assigns probability 46/81 to (0, 0, 0) and 35/81 to (9/5, 9/5, 9/5).

From this example one sees the need to investigate further the foundations of Nash behaviour: the mere common knowledge of the possibility of a Nash theory being played does not imply Nash behaviour, even in an example with a unique Nash equilibrium (with or without mixtures) whose actions have the property of being unique best responses given the actions of the others. The next subsection will present another set of behavioural assumptions which will yield Nash behaviour for any two-person game. The assumptions and the main result are taken from Armbruster an Böge (1979).

4.3 The Knowledge of the Other Players and Nash Equilibrium

In this subsection we will focus on an alternative justification for the concept of Nash equilibrium. We use a subjective interpretation of mixed strategy Nash equilibria. In this interpretation the belief of every player about other players coincides with the mixed strategy part of the other players in the Nash equilibrium. It is important to note that this does not imply that the players should play his/her part of the Nash equilibrium. The example below, due to Myerson, ilustrates the point. There are two players, with action spaces given by $A_1 = \{u, d\}$ and $A_2 = \{l, r\}$. The payoff functions are given by:

		II	
		l	r
I	u	(1, 1)	(1, 1)
	d	(1, 1)	(0, 0)

In this example the Nash equilibrium (u, l) could be the only possible belief in both players' minds. However, the two Bayesian rational players could actually play (d, r), which is not a Nash equilibrium of this game. This problem arises because the Nash actions are not unique best responses. For this reason the subjective interpretation of Nash equilibrium is not as compelling as the classical interpretation. Nevertheless, this subjective view sheds light on some properties of Nash equilibria, as we can see in this and the next subsections. To differentiate the subjective view from the classical view, we will define the former as being a belief that the Nash equilibrium is played.

4.3.1 Definition Let $(\mu_1, \ldots, \mu_n)$ be a mixed-strategy Nash equilibrium for the game u, where $\mu_i \in \Delta(A_i)$. We say that then n-tuple of psychologies $(s_1, \ldots, s_n)$ believes the Nash equilibrium $(\mu_1, \ldots, u_n)$ if for all i: $\text{marg}_{A_{-i}}[\Phi_i(s_i)] = \otimes_{k \neq i}\mu_k =$
$= \mu_1 \otimes \ldots \otimes \mu_{i-1} \otimes \mu_{i+1} \otimes \ldots \otimes \mu_n$.

The main result of this subsection is due to Armbruster and Böge (1979). It says that for two players, if rationality is common knowledge, and if each player knows the other player, then they play a mixed-strategy Nash equilibrium.

4.3.2 Definition Given an n-tuple $(s_1, \ldots, s_n)$ of psychologies, we say that player i knows the other players if:
$\text{supp marg}_{S_{-i}}[\Phi_i(s_i)] = \{s_{-i}\}$.

This definition simply says that player i thinks that the only possible (n-1)-tuple of psychologies of other players is the actual one: $s_{-i} = (s_1, \ldots, s_{i-1}, s_{i+1}, \ldots, s_n)$.

The following theorem is a characterisation of Nash equilibria in two-person games. The first part of the theorem below is in Armbruster and Böge (1979).

4.3.3 Theorem Let u be a two-person game. Suppose rationality is common knowledge, and that player 1 knows player 2 and player 2 knows player 1. Then they believe a mixed-strategy Nash equilibrium of the game u. Conversely, if (μ_1, μ_2) is a mixed-strategy

Nash equilibrium of u, there are psychologies (s_1, s_2) such that rationality is common knowledge, and each player knows each other, with the property that (s_1, s_2) believes (μ_1, μ_2) in the sense of definition 4.3.1.

Proof: Consider the pair $\mu_1 = marg_{A_1}[\Phi_2(s_2)]$ and $\mu_2 = marg_{A_2}[\Phi_1(s_1)]$. We know that $\forall a_1 \in supp\ \mu_1$, a_1 is a best response to μ_2, since player 2 thinks player 1 is rational, and $marg_{S_1}[\Phi_2(s_2)] = \{s_1\}$. Similarly, $\forall a_2 \in supp\ \mu_2$, a_2 is a best response to μ_1. Thus (μ_1, μ_2) is a mixed-strategy* equilibrium of the game u. Conversely, suppose (μ_1, μ_2) is a mixed-strategy Nash equilibrium of the game u. One can construct the infinite hierarchies of beliefs (s_1, s_2) which will believe (μ_1, μ_2) by rationalising in each round every point in the support of one of the mixed strategies by the mixed strategies of the opponent. These infinite hierarchies of beliefs will obviously satisfy the requirements of the theorem.

Q E D.

Unfortunately the result above is not true for games with more than two players. Consider a situation with three players. Each player has beliefs about the actions of the other players. Suppose these beliefs satisfy the following condition: for each player i, the support of the beliefs about the actions of player $(k \neq i)$ is contained in the set of best responses of player k against player k's beliefs about action of players who are not k. If there were only two players, the condition above would imply that the two players believed a mixed-strategy Nash equilibrium, according to definition 4.3.1. With three players, the situation changes. It is not necessarily true that these players have a common prior. Thus, even when all three players know each other, it is possible that they do not believe a Nash equilibrium: this is so because they may hold priors about the actions of others which are not consistent with a common prior. The next example will illustrate this point in formal terms.

4.3.4 Example (Common Knowledge of Rationality and Knowledge of Each Other Does Not Imply Nash Beliefs in Three-person Games). There are three players. The pure strategy sets are: $A_1 = \{u, d\}$, $A_2 = \{a, b\}$ and $A_3 = \{L, R\}$. The payoffs are given by the two matrices below. The matrix on the left corresponds to player three playing L; the matrix on the right, R.

III-L

		II: a	II: b
I	u	(3,2,0)	(2,4,2)
I	d	(1,3,2)	(3,2,-4)

III-R

		II: a	II: b
I	u	(4,-3,1)	(0,-1,0)
I	d	(0,1,-3)	(5,0,6)

Define $\mu_{ij} \in \Delta(A_j)$, for $i \neq j$, and $i,j = 1,2,3$, by:

$\mu_{12} = (1/2a,\ 1/2b)$, $\mu_{13} = (1/2L,\ 1/2R)$;

$\mu_{21} = (1/3u,\ 2/3d)$, $\mu_{23} = (1/3L,\ 2/3R)$;

$\mu_{31} = (2/3u,\ 1/3d)$, $\mu_{32} = (2/3a,\ 1/3b)$.

Then, we have:

A_1 = set of best responses to $\mu_{12} \otimes \mu_{13} = v_1$;

A_2 = set of best responses to $\mu_{21} \otimes \mu_{23} = v_2$;

A_3 = set of best responses to $\mu_{32} \otimes \mu_{31} = v_3$.

We now construct three infinite herarchies of beliefs (s_1, s_2, s_3) such that for every i: $marg_{A_j x A_k}[\Phi_i(s_i)] = \mu_{ij} \otimes \mu_{ik}$ for j, $k \neq i$, with $j \neq k$. These hierarchies of beliefs will be such that rationality is common knowledge and for all i:

supp $marg_{S_i x S_k}[\Phi_i(s_i)] = \{(s_j, s_k)\}$ for $j \neq k$, and j, $k \neq i$ (this means that each player knows the other two players). The construction is simultaneous. The first order beliefs, s_{11}, s_{12}, s_{13} are given by v_1, v_2, v_3, respectively. The higher order beliefs will be all constructed in the same fashion as the second order beliefs. For example,

$s_{21} \in \Delta(A_2 x A_3 x S_{12} x S_{13})$ is given by: $s_{21} = s_{11} \otimes \delta\{(s_{12}, s_{13})\}$, where $\delta\{.\}$ is the probability measure which puts mass 1 on the set $\{.\}$. The hierarchies of beliefs thus built are clearly consistent and satisfy the properties required above. However,

$\mu_{21} \neq \mu_{31}$, $\mu_{12} \neq \mu_{32}$, $\mu_{13} \neq \mu_{23}$. Therefore the triple (s_1, s_2, s_3) does not believe a mixed-strategy Nash equilibrium.

This subsection presented a very intuitive set of behavioural assumptions under which Nash equilibrium is played in a two-person game. This same set of assumptions is not sufficient to generate Nash belief in a three-person game (and, therefore, n-player, n>2). The next subsection will provide sufficient conditions for Nash equilibrium which are a generalisation of the conditions of Theorem 4.3.3 and of Theorem 4.1.3 for two-person games. Also, a set of sufficient conditions for Nash behaviour is provided for n-person strictly concave games which generalise Theorem 4.1.3 when applied strictly concave games.

4.4 The Exchangeability Hypothesis and Nash Equilibrium

In this subsection we generalise Theorem 4.3.3 and Theorem 4.1.3 (in the case of two-person games or strictly concave games).

There is an assumption about psychologies which is crucial for these generalisations. This is the exchangeability hypothesis. Formally, we have:

<u>4.4.1 Definition</u> The Exchangeability Hypothesis. We say that the exchangeability condition holds for player i if $s_i \in E_{1i} =$
$= \{s_i \in S_i \mid \forall k \neq i: \text{supp marg}_{A_k x S_k}[\Phi_i(s_i)] = C_k x D_k$ for some C_k contained in A_k, and D_k in $S_k\}$. We say it is common knowledge in the eyes of agent i when $s_i \in \bigcap_{m \geq 1} E_{mi}$, where:
$\forall m \geq 2$: $E_{mi} = \{s_i \in E_{m-1,i} \mid \forall k \neq i: s_k \in \text{supp marg}_{S_k}[\Phi_i(s_i)] \Rightarrow$
$\Rightarrow s_k \in E_{m-1,k}\}$.

In words: the exchangeability hypothesis means that if an action by player k, a_k, is considered possible by player i, then he also considers it possible when player k is of any of the types s_k he believes player k can be. This is certainly a very strong hypothesis, but it is weaker than requiring that the beliefs of player i about actions of other players and types (or psychologies) of other players be independently distributed.

The first result that we provide generalises Theorem 4.3.3 and Theorem 4.1.3 for the case of only two players. An additional assumption about beliefs is needed. This assumption says that each player considers it possible that the infinite hierarchy of beliefs of the other player is what it really is. That is to say, the players are not totally wrong about each other:

<u>4.4.2 Definition</u> An n-tuple of players' psychologies $(s_1, \ldots, s_n)$ is said to satisfy <u>direct consistency</u> when for all i, it happens that $s_{-i} \in \text{supp marg}_{S_{-i}}[\Phi_i(s_i)]$.

With these two hypotheses we have, then:

4.4.3 Theorem Let u be a two-person game. Suppose (s_1, s_2) are such that: (i) $s_i \varepsilon K_{1i} \cap E_{1i}$, for i = 1,2; and (ii) (s_1, s_2) are directly consistent. Then (s_1, s_2) believes a Nash equilibrium. Conversely, any Nash equilibrium can be believed by psychologies which obey (i) and (ii).

Proof Let $a_2 \varepsilon$ supp $\text{marg}_{A_2}[\Phi_1(s_1)]$. By $s_1 \varepsilon K_{11} \cap E_{11}$, a_2 is a best response against any belief in supp $\text{marg}_{S_2}[\Phi_1(s_1)]$. In particular, by direct consistency, a_2 is a best response to belief s_2. So that a_2 is best response to $\text{marg}_{A_2}[\Phi_1(s_1)]$. In the same way, a_1 ε supp $\text{marg}_{A_1}[\Phi_2(s_2)]$ implies a_1 is a best response to $\text{marg}_{A_1}[\Phi_2(s_2)]$. Thus, (μ_1, μ_2) given by $(\text{marg}_{A_1}[\Phi_2(s_2)], \text{marg}_{A_2}[\Phi_1(s_1)])$ is a Nash equilibrium. Hence, (s_1, s_2) plays the mixed strategy Nash equilibrium (μ_1, μ_2). The converse follows from the converse of Theorem 4.3.3.

Q E D.

We can also generalise Theorem 4.1.3 when applied for strictly concave games. This involves the exchangeability hypothesis, as well as a plausible assumption: the assumption that each player thinks that the other players may think that a Nash equilibrium is being played.

4.4.4 Theorem Let $s_i \varepsilon K_{1i} \cap E_{2i}$. Suppose u is a game where $(\tilde{a}_1, \ldots, \tilde{a}_n)$ is a Nash equilibrium such that every action $\tilde{a}_j$ is the unique best response against $\tilde{a}_{-j}$ (in particular any strictly concave game will do). Assume that s_i is an element of the set $\{s_i \varepsilon S_i \mid \forall k \neq i$: there exists $s_k \varepsilon$ supp $\text{marg}_{S_k}[\Phi_i(s_i)]\}$, such that $\text{marg}_{A_{-k}}[\Phi_k(s_k)] = \delta_{\{\tilde{a}-k\}}$. Then, if player i is rational, he will choose $\tilde{a}_i$, the Nash action. Notice that uniqueness of Nash equili bria is not required. Conversely, any Nash equilibrium with the properties above may be played by hierarchies of beliefs with the properties above.

<u>Proof</u> Since there exists $s_k \in$ supp $\text{marg}_{S_k}[\Phi_i(s_i)]$, such that $\text{marg}_{A_{-k}}[\Phi_k(s_k)] = \delta_{\{\tilde{a}-k\}}$, and since $\tilde{a}_k$ is unique best response to $\tilde{a}_{-k}$, then supp $\text{marg}_{A_k}[\Phi_i(s_i)] = \{\tilde{a}_k\}$ by K_{1i} and E_{2i}. But player i is rational, and again, $\tilde{a}_i$ is unique best response to $\tilde{a}_{-i}$, so that player i chooses $\tilde{a}_i$. The converse of the theorem is a direct consequence of the converse of Theorem 4.1.3.

Q.E.D.

It is interesting to notice that there are several instances where Nash equilibria are believed (in the sense of definition 4.3.1) in which the exchangeability hypothesis is necessary. To see that, suppose $(s_1, \ldots, s_n)$ are psychologies of an n-person game u. Assume that $(\mu_1, \ldots, \mu_n)$ is a mixed-strategy Nash equilibrium of the game u, and that $(s_1, \ldots, s_n)$ believes $(\mu_1, \ldots, \mu_n)$ in the sense of definition 4.3.1, that is to say: for all i $\text{marg}_{A_{-i}}[\Phi_i(s_i)] = \otimes_{k\neq i}\mu_k$. Two hypothesis will imply the necessity of exchangeability. The first hypothesis assumes that every $t_k \in$ supp $\text{marg}_{S_k}[\Phi_i(s_i)]$ is a such that t_k thinks the Nash equilibrium $(\mu_1, \ldots, \mu_n)$ is believed. This hypothesis requires very little justification: it is very unlike Nash equilibrium to suppose it is being played without supposing other people think so also. To contradict it would be the same as saying that the players got to the Nash point by mere coincidence, which sounds extremely odd. The other assumption is less intuitive. It is a principle of a priori ignorance. Given that a belief t_k is considered possible by player i, any a_k which is a best response to t_k must be considered possible of being played by t_k, in the eyes of player i. Notice that we do not require player i to consider all best responses equally likely. We only need player i to consider that all actions which are best responses for t_k are possible of being played by t_k. We conclude this section by stating the "necessity" of the exchangeability hypothesis:

<u>4.4.5 Theorem</u> Suppose $(s_1, \ldots, s_n)$ is such that they believe the mixed-strategy Nash equilibrium $(\mu_1, \ldots, \mu_n)$. Assume that the players think that other players think this Nash equilibrium is believed. Finally, suppose that the ignorance principle holds, that is to say: $\forall i, \forall k \neq i$: $t_k \in \text{supp marg}_{S_k}[\Phi_i(s_i)]$, and if a_k is a best response to $\text{marg}_{A_{-k}}[\Phi_k(t_k)]$ then $(a_k, t_k) \in \text{supp marg}_{A_k \times S_k}(\Phi_i(s_i))$. Then: $\forall i$: $s_i \in E_{1i}$.

<u>Proof</u> Immediate.

Q E D.

5. CONCLUSION

The main point of this paper was to emphasise the fact that the coordination required to play Nash equilibrium is very strong. Not only Bayesian rationality has to be common knowledge, but also the actions chosen by the players, before they are chosen. Hence, one is led to two alternative ways of justifying Nash equilibria as reasonable outcomes of games under consideration. It is the case that in some games the coordination required to achieve Nash behaviours is very wild. Subsection 4.2 dealt with this case. The second manner of justifying Nash behaviour is to try to find more intuitive coordination mechanisms to achieve Nash behaviour. For a two-person game we have a result by Armbruster and Böge (1979). If rationality is common knowledge and each player knows the other player, we obtain Nash behaviour. Thus, for two players, Nash equilibrium and rationalisability can be seen as epistemological poles: if the players know only that rationality is common knowledge, they play a rationalisable action (see Bernheim (1984), Pearce (1984), and Tan and Werlang (1984)). If, additionally, they know everything else about the other player, they play a Nash equilibrium.

We showed that this does not hold for more than two players. Finally, we provided a generalisation of this result.

BIBLIOGRAFIA

ARMBRUSTER, W. and W. BÖGE. "Bayesian Game Theory". *Game Theory and Related Topics*, ed. O. Moeschlin and Pallaschke, Amsterdan: North-Holland, 17-28, 1979.

AUMANN, R. "Subjectivity and Correlation in Randomized Strategies". *Journal of Mathematical Economics*, vol. 1:67-96, 1974.

__________. "Correlated Equilibrium as an Expression of Bayesian Rationality", unpublished, June, 1985.

BANKS, J.S. and J. SOBEL. "Equilibrium Selection in Signaling Games". Discussion Paper 85-9, Department of Economics, University of California, San Diego, 1985.

BASU, K. "Strategic Irrationality in Extensive Games". School of Social Science, The Institute for Advanced Study, Princeton, mimeo, 1985.

BERNHEIM, D. "Rationalisable Strategic Behavior". *Econometrica*, vol. 52, nº 4: 1007-1028, July, 1984.

BILLINGSLEY, P. *Convergence of Probability Measures*. New York: John Wiley and Sons, 1968.

BINMORE, K.G. "Equilibria in Extensive Games". *The Economic Journal* 95:51-59, 1984.

__________. "Modelling Rational Players". Mathematics Department, London School of Economics, mimeo, July, 1985.

BÖGE, W. "Gedanken Über die Angewandte Mathematik". *Mathematiker Über die Mathematik*, M. Otte (ed.). Heidelberg: Springer, 1974.

BÖGE, W. and T.H. EISELE. "On Solutions of Bayesian Games". *International Journal of Game Theory*, vol.8, Issue 4:193-315, 1979.

BRANDENBURGER, A. and E. DEKEL. "Hierarchies of Beliefs and Common Knowledge". Department of Economics, Stanford University, mimeo, July, 1985.

CHO, I.K. "A Refinement of the Sequential Equilibrium Concept". Princeton University, forthcoming Econometrica, mimeo, May, 1985.

DELLACHERIE, C. and P.A. MEYER. *Probabilities and Potential* A. Mathematical Studies nº 29. New York: North-Holland, 1978.

FARRELL, J. "Credible Neologisms in Games of Communication". Economics Working Paper 85-4, GTE Laboratories, June, 1985.

GROSSMAN, S. and M. PERRY. "Perfect Sequential Equilibrium". Forthcoming, 1985.

HARSANYI, J.C. "The Tracing Procedure: A Bayesian Approach to Defining a Solution for n-Person Non-cooperative Games". *International Journal of Game Theory* 4:61-95, 1975.

HARSANYI, J.C. and R. SELTEN. "A Noncooperative Solution Concept with Cooperative Applications" (draft chapters 1-5), Bielefeld discussion papers, 1980/84.

HILDENBRAND, W. *Core and Equilibria in a Large Economy*. Princeton: Princeton University Press, 1974.

KALAI, E. and D. SAMET. "Persistent Equilibria". *International Journal of Game Theory* 13:129-144, 1984.

KOHLBERG, E. and J.F. MERTENS. "On the Strategic Stability of Equilibria". Harvard Business School, mimeo, 1985.

KREPS, D. "Signalling Games and Stable Equilibria". *Quarterly Journal of Economics*. Forthcoming, 1985.

KREPS, D. and R. WILSON. "Sequential Equilibria". *Econometrica* 50: 863-894, 1982.

LUCE, D. and H. RAIFFA. *Games and Decisions*. New York: Wiley, 1957.

McLENNAN, A. "Justifiable Beliefs in Sequential Equilibrium". *Econometrica* 47:1337-1351, 1985.

MERTENS, J.F. and S. ZAMIR. "Formulation on Bayesian Analysis for Games with Incomplete Information". *International Journal of Game Theory* 14:1-29, 1985.

MILGROM, P. and R. WEBER. "Topologies on Information and Related Strategies in Games with Incomplete Information". *Game Theory and Topics*, ed. Moeschlin and D. Pallaschke, North-Holland, Amsterdam, 1979.

MOULIN, H. "Dominance Solvability and Cournot Stability". *Mathematical Social Science* 7:83-102, 1984.

MYERSON, R. "Refinements of the Nash Equilibrium Concept". *International Journal of Game Theory* 7:73-80, 1978.

__________. "Bayesian Equilibrium and Incentive Compatibility: An Introduction". J.L. Kellog Graduate School of Management, Discussion Paper nº 584, Nortwestern University, 1983.

__________. Private Communication, 1984.

__________. "Acceptable and Predominant Correlated Equilibria". Center for Mathematical Studies. Discussion Paper nº 591, Northwestern University, 1985.

PEARCE, D. "Rationalizable Strategic Behaviour and the Problem of Perfection". *Econometrica* 52, nº 4:1029-1050, July, 1984.

RENY, P. "Rationality, Common Knowledge and the Theory of Games", Ph.D. Thesis, Princeton University, 1985.

SAVAGE, L.J. *The Foundations of Statistics*. New York: Wiley, 1954.

SELTEN, R. "Spieltheoretische Behandlung eines Oligopolmodells mit Nachfrägetragheit". *Zeitschrift für Gesamte Staatswissenschaft*, 121, 1965.

__________. "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games". *International Journal of Game Theory*, Vol. 4:25-55, 1975.

TAN T.C.C. and S.R.C. WERLANG. "The Bayesian Foundations of Rationalisable Strategic Behaviour and Nash Equilibrium Behaviour". Chicago University, mimeo, October, 1984.

__________. "On Aumann's Notion of Common Knowledge - An Alternative Approach". Working Paper 85-26. Graduate School of Bussiness, University of Chicago, 1985.

VAN DANME, E. *Refinements of the Nash Equilibrium Concept*. Lecture Notes in Mathematical Economics and Social System nº 219, Berlin: Springer-Verlag, 1983.

WERLANG, S.R.C. "Common Knowledge and Game Theory", Ph.D. Dissertation. Princeton University, March, 1986.