equence of our results is that empirical evidence that is interpreted either as a
ejection of or as a support for the shirking model may be simply rejecting or supporting some of the assumptions defining the economic environment where the nodel is cast.

Riassunto JEL J3

ESISTE UN TRADE OFF TRA LIVELLO DI SUPERVISIONE E REMUNERAZIONI?

Lo scopo di questa nota è di mostrare che le implicazioni generate da modelli standard di salari efficienti sulla relazione tra supervisione e remunerazioni sono nolto sensibili alle assunzioni sullo scenario economico che sta dietro a tali modelli. La nota presenta una versione del modello di salari efficienti in cui la relazione tra supervisione e remunerazioni è positiva. La conseguenza principale riguarda l'interpretazione dell'evidenza empirica esistente come rifiuto o meno dei nodelli di salari efficienti.

RICERCHE ECONOMICHE, 46 (1992) pp 335-378

LEARNING AND CONVERGENCE TO EQUILIBRIUM IN REPEATED STRATEGIC INTERACTIONS: AN INTRODUCTORY SURVEY

PIERPAOLO BATTIGALLI
Università Bocconi, Milano
MARIO GILLI
University of Cambridge and *Università Bocconi, Milano*
M. CRISTINA MOLINARI
Università Bocconi, Milano

1. *Equilibrium and Learning in Economic Theory* (*)

According to the broad approach called "methodological individualism", economic phenomena should be explained as (possibly unintended) results of the intentional and goal-directed actions of interacting individuals. Most formal economic models, however, conform to the narrower neoclassical approach whereby each agent is supposed to choose an optimal solution of a well-defined maximum problem, and analysis focuses almost esclusively on equilibria, i.e. on situations where such choices are mutually "consistent". In game theoretic, or strategic models, mutual consistency is embodied in the notion of Nash equilibrium: each agent chooses a strategy which is a best response to his opponents' strategies. This means that each agent has a correct conjecture about his opponents' behavior.

Received May 1992.

(*) We would like to thank Marco Li Calzi for helpful comments and valuable editorial work. The financial support of Università L. Bocconi is gratefully acknowledged.

We take for granted that the notion of equilibrium is a powerful organizing concept and that equilibrium analysis imposes a healthy methodological discipline. But this does not seem to be enough to justify such a concern with equilibrium. We would like equilibrium analysis to be relevant for the prediction and explanation of economic phenomena. Why should it be so? At the risk of oversimplifying, we consider two provisional answers: (i) an equilibrium is a solution of a grand multi-agent decision problem: since the economic agents are rational, they choose their own component in such a solution; (ii) an equilibrium is a steady state of an adjustment process involving tacit and/or explicit learning: no disequilibrium state can persist.

The first answer may be reasonable when applied to situations with few agents who know each other very well. But it strains our credulity when applied to complex environments such as a whole economic system. This leads us to the second answer and to the following general notion of equilibrium proposed by Hahn (1973): at any time, an agent holds a *theory* that includes his conditional predictions concerning the way the economy will develop and the consequences of his own actions. The agent is assumed to abandon this theory if it is sufficiently and systematically falsified. The agent also has a *policy*, i.e. a plan which specifies his actions conditional on whatever information the agent believes he may possibly receive. An economy is in equilibrium when it generates messages which do not cause the agents to change the theories they hold or the policies they pursue (Hahn 1973, p. 38)(1).

Hence the learning process interpretation entails a change in the notion of equilibrium itself. In particular, we must allow for agents with possibly incorrect conjectures, as long as these conjectures are not contradicted by evidence. We must therefore focus our attention on the observational possibilities of the agents. This equilibrium notion is called *conjectural equilibrium*(2): it will play an important role in this survey as a unifying concept.

The first systematic attempt to pursue a learning-process approach to equilibrium is contained in the literature on learning in rational expectations models (see e.g. Blume, Bray and Easley 1982, Bray and Kreps 1987 and Bray 1990). The learning process consists of a sequence of temporary equilibria. The agents try to estimate the true equilibrium relationship between exogenous and endogenous variables, using the observed realizations. Since agents update their estimates, the temporary equilibrium relationship between exogenous and endogenous variables changes over time. Therefore "learning affects what is to be learned". In most learning models the agents use statistical procedures (e.g. ordinary least squares) which would be correct in a stationary environment, but since the environment of each single agent is made nonstationary by the economy-wide learning, the learning procedures are mis-specified. In some other models, however, learning is modeled as a fully rational process, and the resulting sequence of temporary equilibria is just a grand rational expectations intertemporal equilibrium of a dynamic model with incomplete information.

(1) This equilibrium concept has been anticipated by Hayek (1937). See Littlechild (1982) for more on the comparison between Hayek and Hahn.

(2) Hahn used this terminology some years later, applying his equilibrium concept in order to obtain non-Walrasian equilibria (see Hahn 1978 and 1977).

This literature has some limitations imposed by the parametric nature of the models analyzed(3): (i) since the model does not specify what happens when demand is different from supply, the analyst is bound to consider sequences of temporary equilibria, i.e. 'static equilibrium must be assumed within a single period; and (ii), the analysis can only cover learning in "competitive" situations, where no single agent can affect the observed aggregate outcome. On the contrary, strategic models do not suffer from these limitations: they specify a well-defined outcome for every array of choices (thus allowing for disequilibrium situations within a period) and are sufficiently flexible to be applied to virtually all economic situations. Indeed, in the past decade we have witnessed an impressive diffusion of the language, concepts, and techniques of noncooperative game theory in almost all fields of economics. Therefore, the recent developement of literature on learning in strategic models is not surprising.

Our primary concern is to provide a fairly rigorous account of this literature within a common framework which facilitates comparisons between different contributions and stresses their implicit assumptions and limits.

We propose a broad classification of game theoretic learning models and we choose a few contributions which we regard as representative. These contributions are discussed in some detail: we present the most important results (often in a simplified version) and a few simple proofs. Our choice of the material presented reflects the aforementioned methodological approach. We apologize for the many valuable contributions we have been forced to exclude. The reader is urged to consult Eichberger's survey (Eichberger 1990) which provides a different perspective and a different selection of contributions.

For readers who are less conversant with noncooperative game theory, we have included a short and necessarily incomplete introduction to the subject, which is specially tailored to the needs of the survey. But we also advise these readers to see the brilliant and insightful publication by Kreps (1990), which includes a final chapter on learning.

(3) The Walrasian competitive economy with prices determined by equality of demand and supply is a paradigmatic example of a parametric model. See Battigalli (1988) on the notion of equilibrium in parametric and game-theoretic models.

2. *Equilibrium and Solution Concepts in Noncooperative Game Theory: A Short Introduction*

In this section we introduce the game theoretic language and concepts to be used in the survey. Readers already familiar with the theory of noncooperative games may skip subsections 2.1-2.3[4]. Subsection 2.3 states the fundamental problem: how can we justify the basic equilibrium concept of noncooperative game theory? In subsection 2.4 we introduce appropriate game theoretic formalizations of the "conjectural equilibrium" concept.

2.1. *How Strategic Interaction is Modeled: The Notion of a Game*

We consider a set $\{1, 2, ..., n\}$ of interacting agents, or ***players***, and, possibly, a random device called *Nature* denoted by 0. For many purposes the random device 0 can be regarded as a player and we refer to the set $I = \{0, 1, ..., n\}$ as the players' set. Whenever we make no reference to Nature, we mean that either there is no randomness or that randomness is immaterial for the sake of the argument. Without loss of generality we assume that choices are sequential. The key feature of strategic interaction is that any sequence of players' choices (possibly including random choices by Nature) determines an economic or social outcome $z \in Z$ and each player $i \neq 0$ ranks different probability distributions over the set Z of possible outcomes according to the expected value of a utility function $u_i: Z \rightarrow \mathbf{R}$. Therefore the welfare of each agent depends on other agents' actions as well as his own.

The "rules of the game" — which may be physical and/or institutional — specify which sequences of actions can occur and the amount of information a player has when he is given the move. Whenever a choice is randomly selected by Nature, the rules specify the "objective" probability of each possible choice. A mathematically rigorous definition of the rules of the game is beyond the scope of the present paper. Here we only discuss an example in order to give the necessary intuition, introducing notation and graphic conventions to be used later. This example will be used repeatedly in the following subsections.

EXAMPLE 1. Player 2 is a seller who owns an object having no personal value. He meets one out of two possible buyers: player 1, who values the object at one dollar, and player 3, who values the object at three dollars. Nature selects one of the buyers at random. The "objective" probability of player 1 being chosen is π_0. This selection is not observed by the seller. Once a buyer has been selected seller and buyer bargain as follows: the

(4) A few up-to-date textbooks on game theory have been recently published. The most recent ones are Binmore (1992) (see the learning section), Fudenberg and Tirole (1991) and Myerson (1991).

seller proposes either a low price $P_L = 0.9$ or a high price $P_H = 2.9$. Then the buyer either accepts (Y) or (N) rejects this proposal. The utility of an agent is given by his or her final wealth.

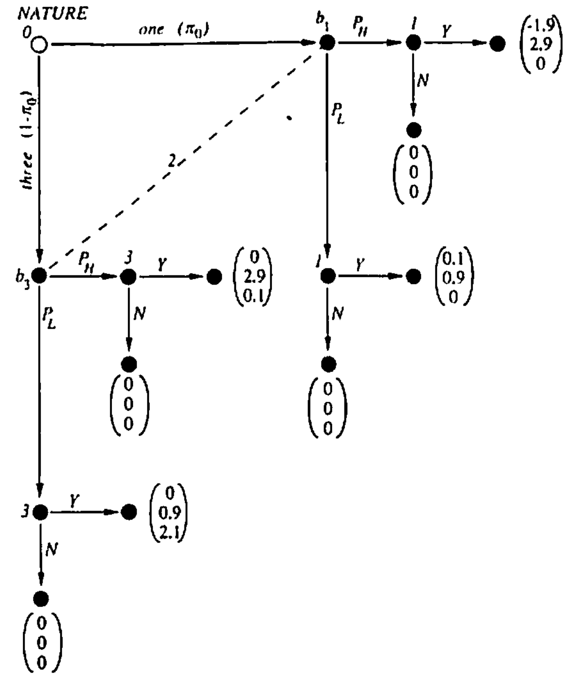

Figure a. - An extensive game

This game is represented in Figure a by means of a tree. Each path of the tree is a possible sequence of random choices and actions. Without loss of generality we can identify outcomes and terminal nodes. To each terminal node z is assigned a vector of utility numbers $(u_1(z), u_2(z), u_3(z))$. Nonterminal nodes, or ***decision nodes***, correspond to actual decision situations. The tree representation implies that each (decision or terminal) node corresponds to a unique sequence of intentional and/or random choices. The state of information h of a player when he has to choose an action (i.e. his subjective decision situation) is given by the set of decision nodes that the player cannot distinguish, and is called *information set*. In the figure such decision nodes are connected by a dotted line. Since the seller does not know which buyer he is facing, his information state corresponds to the set $h = \{b_1, b_3\}$. The set of possible actions when the state of information is h is denoted $A(h)$. In the figure $A(h) = \{P_L, P_H\}$, $A(\{x_0\}) =$

{*one, three*}. Every other decision node x corresponds to an information singleton $\{x\}$ with $A(\{x\}) = \{Y, N\}$.

The detailed mathematical description of the rules of the game is called *extensive form*. If we add to this description the specification of the players' preferences, i.e. their utility functions, we obtain an *extensive game*. A given extensive game is denoted by Γ. We assume hereafter that Γ is a finite game, i.e. its tree contains a finite number of nodes.

We might imagine that a player makes all decisions in advance taking into account all possible contingencies, i.e. that he chooses a plan of action or *strategy*. Let H_i be player i's collection of information sets. Then a strategy for i can be represented as a function

$$s_i : H_i \rightarrow \bigcup_{h \in H_i} A(h), \text{ such that } s_i(h) \in A(h)$$

which assigns a possible action $s_i(h)$ to each information set h. The set of player i's strategies is denoted S_i. S is the Cartesian product of the sets of strategies. We interpret a strategy of the random device as a "state of Nature". If the state of Nature is s_0 and players' actions are given by a strategic profile $(s_1, \ldots, s_n)$, an outcome $z = \zeta(s_0, s_1, \ldots, s_n)$ occurs. We therefore have a well-defined *outcome function* $\zeta : S \rightarrow Z$ mapping strategic profiles into outcomes. If we fix only the strategies of the nonrandom players, we obtain a probability distribution over outcomes denoted $P(\cdot | s_1, \ldots, s_n)$. It is worth noting that the outcome function is not one to one unless each player takes his action without knowing anything about previous choices (including Nature's choices)(5). In this latter case it is as if all choices were simultaneous. We therefore say that an extensive game with a one to one outcome function is a *simultaneous game*.

It is often assumed that players can choose random devices which either implement strategies or select actions with given probabilities. Mathematically, this amounts to enlarging player i's strategy space to the set of distributions $\Delta(S_i)$ or to the set Π_i of arrays of probability distributions $\pi_i = (\pi_i(h) \in \Delta(A(h)), h \in H_i)$ ($\Delta(\cdot)$ denotes the space of distributions on some set). A distribution $q_i \in \Delta(S_i)$ is called *mixed strategy* and an array of distributions $\pi_i \in \Pi_i$ is called *behavior strategy*. Of course, Nature's behavior strategy, π_0 is fixed by the rules of the game. Mixed and behavior strategies of player i also have alternative and related interpretations. They may represent probabilistic beliefs about i of some other player j. We may also think that the actual player i is drawn at random from a large population of agents, that each agent deterministically plays a given s_i, and that q_i is the statistical distribution of strategies in the population. In this case a belief about player i would be a probability distribution g_i on the space of distributions $\Delta(S_i)$, as if i actually chose a mixed strategy.

(5) The reader may easily check this point using example 1.

Fix the strategy s_j of each nonrandom player j except player i. Such a strategic profile is denoted s_{-i}. Then a mixed (behavior) strategy q_i (π_i) induces a probability distribution $P(\cdot | q_i, s_{-i})$ ($P(\cdot | \pi_i, s_{-i})$) over the outcome space Z. A fundamental result due to Kuhn (1953) states the following:

PROPOSITION 1: *Assume that the extensive form is such that player i in every decision situation remembers what he knew and what he did in the preceding decision situations (if any). Then there is no relevant difference between choosing mixed strategies and choosing behavior strategies, because for each q_i there is a corresponding π_i such that for all s_{-i}*

$$P(\cdot | q_i, s_{-i}) = P(\cdot | \pi_i, s_{-i})$$

and vice versa.

When the assumption of the theorem holds, we say that player i has *perfect recall*. We tacitly assume that each player has perfect recall, whenever we do not say otherwise.

We now define a mathematically useful "summary" of an extensive game. Note that for each player i a profile of strategies $s = (s_1, \ldots, s_n)$ induces an expected utility $U_i(s)$ by the formula

$$U_i(s) = \sum_{z \in Z} P(z|s)\, u_i(z)$$

Assume that each player i chooses in advance a strategy s_i, which is automatically implemented by a mechanical device or by obedient subordinates. This choice cannot be observed by the other players. Such a situation corresponds to a simultaneous game denoted by $G = (S_1, \ldots, S_n, U_1, \ldots, U_n)$ and is called the *normal form* of the original extensive game. Clearly the normal form G provides less information on the structure of strategic interaction than the original extensive game Γ, unless Γ is a simultaneous game. We identify a simultaneous game Γ with its normal form G.

2.2. *Rationality, Common Knowledge of Rationality, and their Implications*

Up to this point we did not assume anything about the players except that they have perfect recall and well defined preference orderings. If we want to get some observable implications from the theory, however, we must make some assumptions about the players' behavior. In this section we analyze the consequences of the assumptions that (i) the players are rational and well informed, and (ii) that this is common knowledge(6).

(6) The assumption of common knowledge of the game and of players' rationality has been formally defined in a few papers (e.g. Tan and Werlang 1987). The notion of common knowledge of an event has been mathematically defined in several ways since the seminal paper of Aumann (1976) (see Binmore and Brandeburger 1988).

We say that an event E is *common knowledge* if each player knows E, each player knows that each player knows E, each player knows... We say that Γ is a *game with complete information* if it is common knowledge that the true structure of strategic interaction is given by Γ itself.

Let us consider a two-player simultaneous game with complete information $G = (S_1, S_2, U_1, U_2)$. If a Bayesian-rational player i has a subjective belief $g_i^j \in \Delta(S_j)$(7), he plays a strategy which maximizes his subjective expected utility

$$U_i(s_i, g_i^j) = \sum_{s_j \in S_j} U_i(s_i, s_j)\, g_i^j(s_j)$$

We say that such a strategy is *justified* by the belief. A strategy is (strictly) *dominated* or irrational if it cannot be justified by any belief.

Let $J_i(S_j)$ be the set of undominated strategies of player i. More generally, for each subset T_j of strategies of player j, $J_i(T_j)$ is the subset of strategies of i justified by some belief with support included in T_j. Let $J(T_1 \times T_2)$ be the Cartesian product of these sets. If each player i knows that the opponent plays a strategy in T_j, then the original game can be replaced by a restricted game with strategy space $J(T_1 \times T_2)$, because if player j knows that i is rational, he assigns zero probability to the set of dominated strategies of i. The function $T \rightarrow J(T)$ has an obvious, but important, monotonicity property: if T shrinks, its image $J(T)$ also shrinks (or it remains the same), because fewer strategies can be justified by fewer beliefs. Such monotone set-to-set functions are called *justification operators* (Milgrom and Roberts 1991). The definition of the relevant justification operator for a given game depends on the assumed form of the players' beliefs. For example, if players' beliefs were deterministic the relevant justification operator would be different from J. Here we focus on operator J because it is appropriate for Bayesian rationality, but sometimes it is useful to take an unspecified justification operator as primitive.

If each player knows that the opponent is rational, the original strategy space S can be replaced by $J(S)$ and only strategy pairs in $J(J(S))$ can be rationally chosen. Let J^k denote the k^{th} iteration of operator J, i.e. $J^k(T) = J(J^{k-1}(T))$. The present discussion suggests that if rationality is common knowledge the players choose a pair of strategies in $J^K(S)$, where K is the smallest positive integer such that $J^{K+1}(S) = J^K(S)$. (Note that, by monotonicity of J, $J^{k+1}(S) \subseteq J^k(S)$; i.e. these sets (weakly) shrink as k increases). Since S is finite the integer K is well defined. Such strategies are called *rationalizable* or *iteratively undominated*(8). A game is *dominance solvable* if it has a unique rationalizable outcome.

(7) A belief can also be modeled as a distribution on the set of the opponent's mixed strategies, but the difference is immaterial in this context (see Pearce 1984).

Let R^* be the set of rationalizable strategies. Does common knowledge of rationality imply something more than a play in R^*? Without additional assumptions, it does not. One way to consider this point is the following: if each player knows that his opponent only plays rationalizable strategies, the actual play must be in the set $J(R^*)$, but $R^* = J(R^*)$, i.e. R^* is a fixed point of the operator J. It is also easy to show that R^* is the largest fixed point of J: assume that a subset T of S is also a fixed point of J; then $J(T) = T \subseteq S$; recursively applying monotonicity, we get $J^k(T) = T \subseteq J^k(S)$ for all k. Therefore $T \subseteq R^*$.

The rationalizability approach can easily be extended to n-players simultaneous games, but its extension to nonsimultaneous games is problematic. However, there are many extensive games which can be easily solved by exploiting the additional information on the structure of interaction provided by the extensive form. For example, it is trivial to work out the rational strategies of the buyers in the game of Figure a: player 1 only accepts the low price and player 3 accepts both prices. Knowing this, the seller will propose a low price if the probability π_0 is large and a high price if π_0 is small. But it can be easily checked that the following profile is rationalizable in the normal form for every value of π_0: both buyers accept only the low price P_L and the seller proposes P_L. This shows that we can refine some normal-form results using the extensive form.

2.3. *Nash Equilibrium: Justifications and Problems*

Consider a game with normal form G and assume that there is complete information. It may be the case that there is an "obvious way" to play the game G. For example, G could be dominance solvable, or it could be the normal form of the extensive game in Figure a, which has a trivial solution. Situations of a different kind will be discussed later. But how can we characterize such an obvious way to play the game? Since it is obvious, each player can figure out how his opponents play, and since he is rational his strategy must be justified by his belief. Therefore we have the following necessary condition: $(s_1, ..., s_n)$ is an obvious way to play G only if s_i maximizes $U_i(\cdot, s_{-i})$ for each $i = 1, ..., n$ (equivalently $s \in J(\{s\})$). A profile of strategies with this property is called a *Nash equilibrium*. This

(8) When there are three or more players, we can define two plausible justification operators. If we exclude correlated beliefs about the opponents, we get rationalizable strategies; if we allow for correlated beliefs we get iteratively undominated or *correlated rationalizable* strategies (see e.g. Bernheim 1986). Tan and Werlang (1987) formally show that (correlated) rationalizability is equivalent to common knowledge of rationality. In this survey we ignore the issue of correlation. Hence we do not distinguish between iterated dominance and rationalizability.

definition is easily extended to the case where the players can choose mixed strategies or behavior strategies. A Nash equilibrium in mixed or behavior strategies may also be interpreted as an "obvious belief about the play of G". For example, consider a two-player game and let such a belief be given by the product of the distributions q_1 and q_2. Since the belief is obvious it must be shared by the players. But each player i knows that the opponent j plays rationally. Therefore each strategy in the support of q_j, the belief of player i, must be justified by q_i, the belief of player j. Clearly this condition holds if and only if (q_1, q_2) is a Nash equilibrium in mixed strategies. This implies that if q is a Nash equilibrium in mixed strategies, then $Supp(q) \subseteq J(Supp(q))$.

It is easy to prove that (the support of) every Nash equilibrium is contained in the rationalizable set R^*: if $\bar{s}$ is a Nash equilibrium, then $\{\bar{s}\} \subseteq J(\{\bar{s}\})$, by the monotonicity of J, $J(\{\bar{s}\}) \subseteq J(S)$. Applying monotonicity repeatedly we get

$$\{\bar{s}\} \subseteq J(\{\bar{s}\}) \subseteq J^2(\{\bar{s}\}) \subseteq \ldots \subseteq J^k(\{\bar{s}\}) \subseteq J^k(S)$$

for all k. Hence $\{\bar{s}\} \subseteq R^*$. (A similar proof holds for the support of a Nash equilibrium in mixed strategies.) It is a quite obvious consequence of this fact that if a game is dominance solvable, the rationalizable profile $\bar{s}$ is also the unique Nash equilibrium.

In most applications of noncooperative game theory it is assumed that a Nash equilibrium is played. How can we justify such an assumption? "Classical" justifications typically rely on complete information and common knowledge of rationality. It is assumed that the game is played exactly once (situations of repeated interaction should be appropriately modeled as a grand extensive game) and the players have enough knowledge and ability to analyze the game in a rational manner, simulating their opponents' reasoning. The outcome of such an analysis should be a commonly held "obvious" belief about the play, i.e. a Nash equilibrium. This approach is sometimes called *eductive* (see Binmore 1987).

Note however that complete information and common knowledge of rationality may be insufficent to justify the playing of a Nash equilibrium. What about simultaneous games which are not dominance solvable? Some authors argue that eductive analysis leads to playing a Nash equilibrium if the equilibrium is unique (more generally if the equilibria are exchangeable)(9). We might accept this position for the sake of the discussion. But then what about the case of multiple (nonexchangeable) equilibria? In some cases the multiplicity problem can be solved by eliminating "implausible" equilibria. Take the seller/buyers example. We know that it has

(9) Kaneko and Tagashi (1991) provide an interesting and sophisticated formalization of this argument.

a unique (and trivial) solution, which is of course a Nash equilibrium. But there are other equilibria. We have already mentioned one of them: both buyers accept only the low price and the seller proposes the low price. This equilibrium relies on an implicit "noncredible threat" of the buyer with a high reservation price. But a rational and completely informed seller would not believe such a threat. Examples of this sort have prompted an effort to provide *refinements of the Nash equilibrium concept* (see van Damme 1991). However the multiplicity problem has not been solved for general games by the literature on refinements(10).

In some contexts it is reasonable to assume that there is a "preplay communication" stage, in which completely informed players may reach an agreement about their behavior in the game, but the agreement cannot be enforced (otherwise we would be in the domain of cooperative game theory). Therefore each player trusts the other players and plays as prescribed only if the agreed-upon mode of behavior is a "plausible" Nash equilibrium.

Eductive analysis and/or preplay communication may lead to play a Nash equilibrium when there is complete information. But in many applications of game theory the complete information assumption (which implies the knowledge of the opponents' preferences) is clearly unrealistic. Is there any role for Nash equilibrium analysis in such situations? Harsanyi (1967) proposed a theory which answers this question in the affirmative. Let B_i be the space of possible vectors of player i's characteristics. Such vectors are called *types*. For example, B_i could be the space of possible utility functions of player i. The Cartesian product of these spaces is B and B_{-i} is the space of types of players different from i. Each player knows his own type, but possibly ignores those of the opponents. The belief of player i about his opponents' types may depend on his own type t_i and is denoted $\beta_i(\cdot, t_i) \in \Delta(B_{-i})$. Assume now that the beliefs of the different players are consistent, i.e. there is a *common prior* $\beta \in \Delta(B)$ such that each $\beta(\cdot, t_i)$ is the marginal distribution on B_{-i} conditional on t_i derived from β. According to Harsanyi's theory such a situation is strategically equivalent to a game of complete information where the types are selected at random with an "objective" distribution $\pi_0 = \beta$, and each player learns his own type and implements a strategy. The Nash equilibria of this enlarged game are called *Bayesian*. For example, the game of Figure a can be interpreted as a game of incomplete information where the buyer has two possible reservation values and the seller does not know the true one.

(10) Some authors have proposed eductive theories, selecting a (mixed) Nash equilibrium for each finite game (the most noteworthy attempt can be found in Harsanyi and Selten 1988). Such theories are not widely accepted in the profession. However, note that Harsanyi and Selten's concept of "risk dominance" can be justified by some learning processes (see, for example, Hendon et al. 1991, Myerson 1991 (section 3.7), Kandori et al. 1991 and Young 1992).

Harsanyi's theory is brilliant (11) and it has prompted a variety of interesting economic models. But it shifts the problem one step back when it assumes the existence of a common prior and the equivalence with a situation of complete information for the corresponding enlarged game. Consider a modification of our previous example: if the buyer has to propose a price to the seller, it is important to know the seller's belief. Here the common prior assumption is crucial, but also questionable.

To sum up, the classical arguments for the relevance of the Nash equilibrium concept do not apply in many interesting situations. It is therefore important to explore alternative explanations, relying on the repetition of the game and on the adaptation of players' strategies. An equilibrium is intepreted as a stationary point of the resulting process. Such an interpretation is sometimes called *evolutive* (see Binmore 1987) and takes two different though nonexclusive routes. The approach considered in this survey explicitly analyzes processes where the players learn from their past experience and adapt their behavior accordingly. The other follows evolutionary biology, assuming that the relative frequency of a strategy in a population of individuals with the same characteristics is correlated with its success/fitness. Such correlation may be caused by learning, imitation, and higher rates of reproduction for fitter strategies, and is not explicitely explained. Good references for the evolutionary approach are provided by Friedman (1991), Nachbar (1990) and van Damme (1991, ch. 9).

2.4. *Conjectural Equilibria of Noncooperative Games*

Before moving on to the evolutive (and hence dynamic) analysis of games, we propose a static equilibrium concept, which is meant to represent the stationary states of plausible learning processes. Our definition is a game theoretic formalization of Hahn's notion of conjectural equilibrium (Hahn 1973)(12).

Assume that an extensive game Γ is played repeatedly. At the end of each period each player i gets a message $m_i = \mu_i(z)$ in i's message space M_i, which partially or perfectly reveals the play (terminal node) of that period. Such messages are used to update players' beliefs about their opponents' behavior. We say that there is *perfect monitoring* if each player i exactly observes all the actually chosen actions, i.e. if each μ_i is the identity function on Z. Assume also that the players behave myopically, i.e. they try to maximize their current or short-run expected utility given their updated beliefs (in Section 3 we discuss a variety of assumptions implying myopic behavior). We may assume without loss of generality that the players know the extensive form of Γ (except possibly π_0) and their own utility functions.

We consider two kinds of environments:

(i) the actual set of agents coincides with the player set I and they meet in each period. This is the *non-anonymous* case.

(ii) for each player index i there is a population of agents with the same characteristics, but possibly different beliefs (possibly due to different experiences); each agent meets other $(n-1)$ agents, each one drawn at random from a different population; the identity of the opponents is either unknown or ignored. No player acquires any information about matches which he did not participate in. This is the *anonymous* case.

Let $P(m_i|s_i, g_{-i})$ be the subjective probability of receiving message m_i as assessed by player i when he has a belief g_{-i} and plays s_i. Analogously $P(m_i|s_1, \ldots, s_n)$ is the objective probability of m_i when the profile $(s_1, \ldots, s_n)$ is played. Such probabilities can be derived using the outcome function ζ and the message function μ_i. In the non-anonymous case a belief can be appropriately modeled as a probability distribution on the opponents' strategies, where the "opponents" include Nature, if π_0 is unknown. We say that a profile of strategies $(s_1, \ldots, s_n)$ is a *non-anonymous conjectural equilibrium* of Γ if for each player i we can find a belief $g_{-i} \in \Delta(S_{-i})$ such that

(a) g_{-i} justifies s_i, and

(b) $P(m_i|s_i, g_{-i}) = P(m_i|s_1, \ldots, s_n)$ for all $m_i \in M_i$(13).

Condition (a) says that player i maximizes the expected payoff of the current period, but we will see that this can be consistent with long-run maximization. Condition (b) says that the observed long-run frequencies of messages should not induce i to change his belief and choice (cfr. the informal definition of a conjectural equilibrium in Section 1). Notice that when chance moves are absent the outcome of a strategic profile $(s_1, \ldots, s_n)$ is deterministic. Hence condition (b) can be restated as follows: (b') $P(\mu_i[\zeta(s_1, \ldots, s_n)] \mid s_i, g_{-i}) = 1$.

Clearly every Nash profile s is a conjectural equilibrium justified by correct beliefs, but if the players just learn from the observed messages we should not *a priori* expect them to hold correct beliefs. It is interesting to note, however, that there is a class of games where Nash and conjectural equilibria yield the same outcomes.

(11) It is comparable to the introduction of uncertainty in the general equilibrium model.

(12) To the best of our knowledge the concepts and results of this subsection were originally proposed for non-anonymous games in Battigalli (1987) (see also Battigalli and Guaitoli 1988). Similar concepts and results have been independently proposed by Rubinstein and Wolinsky (1990), Kalai and Lehrer (1991b), Fudenberg and Levine (1990) and Fudenberg and Kreps (1988). A different but related approach is contained in Kaneko (1987). Another interesting non-Nash equilibrium concept relying on learning is provided by Gilboa and Matsui (1991).

(13) Rubinstein and Wolinsky (1990) define a sophisticated version of this concept, the rationalizable conjectural equilibrium, which is appropriate under complete information and common knowledge of rationality.

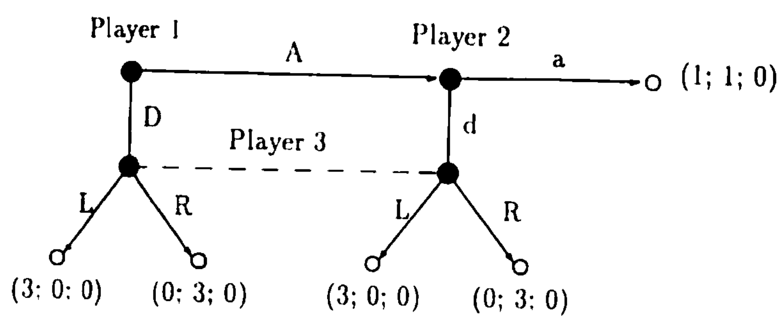

Figure b. - A game in extensive form

PROPOSITION 2: *Assume that Γ has two players and either there is no random move or π_0 is known. Assume also that there is perfect recall and perfect monitoring. Then each conjectural equilibrium of Γ is realization equivalent to a Nash equilibrium in mixed or behavior strategies.*

The proof of this result is quite simple. Let (s_1, s_2) be a conjectural equilibrium supported by beliefs (g_2^1, g_1^2). By perfect recall, player i's belief g_j^i is realization equivalent to a behavior strategy π_j of player j. Since each player perfectly observes the outcome, he observes the actions actually chosen by player j, i.e. the actions selected by s_j at the information sets of j reached by the play. Therefore condition (b) implies that π_j assigns probability one to such actions. The value of π_j at "unreached" information sets does not affect the expected utility of j. Hence $U_i(s_i, g_j^i) = U_i(\pi_i, g_j^i) = U_i(\pi_i, \pi_j)$, where the last equality holds because g_j^i is realization equivalent to π_j. Then condition (a) implies that π_i is justified by π_j. It follows that (π_1, π_2) is a Nash equilibrium in behavior strategies inducing the same distribution on Z (the same outcome if there is no random choice) as (s_1, s_2).

The proof shows that the result can be extended to n-player games with π_0 unknown: for example when the beliefs of any two players about a third opponent (possibly Nature) coincide[14]. If this is not the case, a conjectural equilibrium may induce a non-Nash outcome. This is illustrated in the following example (cfr. Fudenberg and Kreps 1988).

EXAMPLE 2: In the game of Figure b, (A, a) is part of a conjectural equilibrium where player 1 has a correct belief about player 2 and assigns a high probability to R, player 2 has a correct belief about player 1 and assigns a high probability to L (player 3's utility and beliefs are irrelevant in this example, he might be Nature). No common belief about player 3 justifies (A, a) and the corresponding outcome is not a Nash equilibrium outcome (we will elaborate on this example later).

[14] We will see in Section 4.1 that the equivalence result can also be extended to repeated, simultaneous games.

Even in games where conjectural equilibria are realization equivalent to Nash equilibria the conjectural equilibrium concept may be relevant. For example, an "implausible" Nash equilibrium outcome supported by an "noncredible threat" may be interpreted as a conjectural equilibrium in which the threat is erroneously believed (presumably because of incomplete information), but it would not be carried out.

In the anonymous case a belief can be appropriately modeled as a probability distribution on the space of statistical distributions of strategies in the opponents' populations. An *anonymous conjectural equilibrium* is a profile of distributions $(q_1, \ldots, q_n)$ such that for each player/population i and each strategy s_i in the support of q_i there is a belief $g_{-i} \in \Delta[\Delta(S_{-i})]$ such that properties (a) and (b) (with $P(\cdot|s_i, s_{-i})$ replaced by $P(\cdot|s_i, q_{-i})$) hold.

Note that different strategies of the same player/population in the support of the equilibrium distribution can be justified by different beliefs, because they are chosen by different agents. Fudenberg and Levine (1990) consider the special case of perfect monitoring of the outcome and show that, even in two-player games, there are anonymous conjectural equilibria which are not equivalent to Nash equilibria.

We will see in the following sections that we can reasonably expect that a learning process converges to a conjectural equilibrium (if it converges at all). But we have shown that even in the favorable case of perfect monitoring of the outcome a conjectural equilibrium need not be equivalent to a Nash equilibrium. If the observation of the outcome is imperfect a conjectural equilibrium can even be outside the rationalizable set as is shown in the game of Figure c. The strategy b is dominated by a. Hence the game is dominance solvable and the only rationalizable strategies are B and a. However the profile of strategies (A, a) is a conjectural equilibrium when the players, at the end of the game, observe only their own payoff.

	a	b
A	(2; 4)	(2; 3)
B	(3; 4)	(1; 3)

Figure c. - A simultaneous game

	N	C
N	(5; 5)	(0; 6)
C	(6; 0)	(1; 1)

Figure d. - A prisoners' dilemma game

3. *Repeated Games and Learning*

3.1. *How to Model Repeated Strategic Interaction*

Since learning arises in the context of repeated interaction we should consider how traditional game theory analyzes such situations. Imagine that a game with normal form $G = (S_1, ..., S_n, U_1, ..., U_n)$ and no chance moves, called *component game*, is played repeatedly over time. The resulting sequence of component games is itself a game which is called *repeated game*.

To characterize repeated games we need two elements. First of all, it is necessary to specify the information that the players receive at the end of the component game. In fact, when the players can observe the outcomes of past component games, repeated games are a more complex strategic situation than the simple sequence of component games.

To grasp this aspect consider, for example, the prisoner's dilemma game in Figure d, where the players can either confess (C) or not confess (N). In the infinitely repeated prisoners' dilemma, if the players do not observe anything at the end of each component game, their repeated game strategies, called superstrategies, are simply infinite sequences of Cs and Ns (*open loop superstrategies*). If, on the contrary, they can observe, at least partially, the past history of the game, they can play superstrategies which can be much more complicated (*closed loop superstrategies*). An example is the trigger strategy "do not confess on the first iteration and continue to refuse to confess if both have also refused to confess in the past, but revert to confession after the first defection and keep confessing afterwards".

Recall that the signal function for player i in extensive games is $\mu_i: Z \rightarrow M_i$, where Z is the set of terminal nodes of the component game. Using the outcome function ζ we can obtain the signal function in strategic form $\psi_i(s) = \mu_i(\zeta(s))$. This function associates each profile of strategies with the message received by player i at the end of the component game. Let $m_i^t = (\psi_i(s_{(0)}), ..., \psi_i(s_{(t-1)}))$ be player i's observation up to time t and M_i^t the space of all possible messages up to t. A (*pure*) *superstrategy* for player i is given by $r_i = (r_{i,(0)}, r_{i,(1)}, ...)$ where $r_{i,(0)} \in S_i$ and $r_{i,(t)}: M_i^t \rightarrow S_i$. Let R_i be the set of players i's superstrategies. Similarly, a behavior superstrategy for player i is given by an array $\pi_i = (\pi_{i,(0)}, \pi_{i,(1)}, ...)$ such that $\pi_{i,(0)} \in \Delta(S_i)$ and $\pi_{i,(t)}: M_i^t \rightarrow \Delta(S_i)$.

A second element we need, in order to define a repeated game, is the way players evaluate their payoffs arising in different time periods. Let $\delta_i \in (0, 1]$ be player i's discount factor. Assume that the profile of superstrategies $r = (r_1, ..., r_n)$ is played and let m_j^t be player j's resulting observations up to time t (i.e. $m_j^1 = \psi_j(r_{1,(0)}, ..., r_{n,(0)})$, $m_j^2 = (m_j^1, \psi_j(r_{1,(1)}(m_1^1), ..., r_{n,(1)}(m_n^1)))$. Then player i's repeated game utility is:

$$L_i(r) = U_i(r_{(0)}) + \sum_{t=1}^{\infty} \delta_i^t U_i(r_{(t)}(m^t)).$$

For simplicity's sake we assume that the discount factor is the same for each player, that is $\delta_i = \delta$. A repeated game with imperfect monitoring is thus given by $RG = (S_1, ..., S_n, U_1, ..., U_n, \psi_1, ..., \psi_n, \delta)$(15).

Both the definition of Nash equilibrium and that of conjectural equilibrium given in section 3 can be naturally extended to the repeated game RG. A profile of superstrategies $(r_1, ..., r_n)$ (an analogue definition can be given in terms of behavior superstrategies) is a Nash equilibrium if, for each player, r_i is a long-run best reply to r_{-i}, that is, for each player i's superstrategy r_i'

$$L_i(r_i, r_{-i}) \geq L_i(r_i', r_{-i}).$$

The profile of superstrategies $(r_1, ..., r_n)$ is a (non-anonymous)(16) conjectural equilibrium if, for each player i, there is a belief $g_{-i} \in \Delta(R_{-i})$ over his opponents' superstrategies such that:

a) g_{-i} justifies r_i, with respect to the repeated game utility;

b) for each player i the messages observed are consistent with his conditional subjective probability, given his strategy and his belief. Since there are no chance moves this means that, for each time t and sequence of messages m_i^t induced by $r_1, ..., r_n$, $P(m_i^t | r_i, g_{-i}) = 1$.

Repeated games are characterized by a multiplicity of Nash equilibria even when the underlying component game has a unique equilibrium. For example, for sufficiently high values of δ, a pair of trigger strategies in the infinitely repeated prisoners'dilemma is an equilibrium that supports the "cooperative" outcome (N, N). This outcome is doomed to failure in the one-shot version of the game, because C is a strictly dominant action.

Furthermore, many more outcomes can be sustained as equilibria. The folk theorem shows that any outcome that yields each player an individual component game payoff greater than the minimax of his payoffs, can be sustained in a noncooperative equilibrium of the infinitely repeated game with no discounting.

(15) Note that RG can be interpreted either as a standard non-anonymous repeated game or as an anonymous repeated game where at the beginning of each period each agent is selected at random from a large population of individuals characterized by $(S_i, U_i, \psi_i, \delta)$.

(16) We define only a non-anonymous repeated game conjectural equilibrium because, as we will see, in the case of anonymous repeated games learning converges to the conjectural equilibria of the component game.

3.2. *A Classification of Game Theoretic Learning Models*

It is clear from the preceding discussion that the reasoning of the players in a repeated game could be extremely complex. In order to limit this complexity somehow, learning is often associated with some kind of bounded rationality or incomplete information that results in myopic behavior; that is, maximization of current period utility. In order to explain how myopic behavior emerges, we single out four features characterizing learning models, and use a simple model of recursive dynamic programming to show how their combination can result in myopic behavior. Two of these features are "psychological" and concern players' preferences and beliefs. The remaining two, which are "structural", regard the structure of the repeated strategic interaction.

The first psychological feature is highlighted in the literature on learning in rational expectations equilibrium models discussed in Section 1. As we said, one of its powerful insights can be summarized by the statement that "learning affects what is to be learned". This insight can be transferred from the competitive environment of rational expectations models to the strategic environment of game theory. Here, the role that it plays is even greater because, as we know from the previous section, in a repeated game nonstationarity arises not only because players observe past play before they decide their behavior, but also because they can influence their opponents' behavior through reputational or retaliation effects.

Thus, we can assume that either players are sophisticated(17) (that is, they recognize the effects of learning) or they are naive. A *sophisticated player* thinks that his opponents play closed loop superstrategies, and thus his belief about his opponents' play is given by a probability distribution on the set of his opponents' closed loop superstrategies. A *naive player* thinks, on the contrary, that the other players' behavior is not influenced by past history of the play, including his own past strategies. Hence he neither tries to influence the strategies of the other players in the future nor does he realize that his opponents' behavior is nonstationary. A naive player simply thinks that the others play open loop strategies and his belief is represented by a probability distribution over the set of his opponents' pure or mixed strategies of the component game.

A general example of naive learning is *stationary Bayesian learning* according to which the players analyze past observations as if their opponents' play were governed by a fixed, albeit unknown, probability distribution. Other naive learning algorithms are the Cournot dynamics, whereby each player chooses a best reply given the assumption that the others will choose their last period strategy, and the fictitious play dynamics, proposed by Brown (1951), under which agents play a best reply to the empirical frequencies of their opponents' past strategies(18).

(17) The use of the term "sophisticated" is slightly different from that found in the literature (Milgrom and Roberts 1991).

A second "psychological" feature that characterizes learning models is the discount factor that the players use to evaluate future payoffs(19). We can either have players with *myopic preferences* or players with *foresighted preferences*. Players have myopic preferences when they are so impatient that they discount the future by a factor δ so small that long-run maximization is equivalent to a sequence of short-run maximizations. The extreme case of myopic preferences is represented by a discount factor δ equal to zero.

The "structural" features of learning models regard the observation structure available to the players at the end of the component game and the definition of the player set. As regards the observation structure the important distinction to be made is based on its *manipulability*. Player i has a manipulable observation structure when what he observes depends on his past actions. Formally let $S_{-i}(m_i, s_i) = \{s_{-i} \in S_{-i} \mid \psi_i(s_i, s_{-i}) = m_i\}$ be the set of possible strategies of player i's opponents when player i has chosen strategy s_i and observes message m_i. For each player i's strategy s_i, the collection of such sets

$$\Psi_{-i}(s_i) = \{S_{-i}(m_i, s_i);\ m_i \in M_i\}$$

is a partition of S_{-i} representing the possible observations of player i about his opponents' strategies; that is, player i's observation structure. This structure is manipulable if it depends on player i's strategy, that is if for some s_i' and $\hat{s}_i \in S_i$, $\Psi_{-i}(s_i') \neq \Psi_{-i}(\hat{s}_i)$.

For example, the game of Figure b has a manipulable observation structure. Assume that players have perfect monitoring of the outcome. If player 1 plays A and player 2 plays a, no information about player 3's strategy accrues to 1 and 2 because player 3's information set is not reached. To obtain evidence about player 3's strategy either D or d should be chosen.

Note that when the game is nonsimultaneous, as in Figure b, perfect monitoring of outcomes does not involve perfect observation of the opponents' component game strategies and there is at least one player with a manipulable observation structure.

Active experimentation is a possible consequence of manipulability.

(18) For a summary of convergence results see Moulin (1986, ch. 6) as regards Cournot dynamics, and Fudenberg and Kreps (1991) for fictitious play. As for the latter, an example of cycling for two-player games with three strategies is Shapley (1964). Note that originally fictitious play was not interpreted as a learning process but as a way to compute Nash equilibria.

(19) To a certain extent the discount factor can be regarded as structural since it depends not only on the players' characteristics but also on the length of the periods.

In fact, the players could have an incentive to experiment in order to get more precise information about their opponents' behavior; namely, they can rationally sacrifice some expected payoffs in the current period to acquire more information and so, hopefully, higher payoffs in the future.

In the literature on learning in games it is usually assumed that players have perfect monitoring of the component game outcome. This assumption rules out the manipulability of the observation structure in simultaneous games. Note, however, that this is a strong requirement in many strategic situations of economic interest. The game of Figure c, for example, shows that if the players observe only their own payoff and they cannot observe their opponents' payoffs realization (an assumption which is not implausible) they need not have perfect monitoring of the actions of the game, even if there is common knowledge of the utility functions. If player 1 plays A, his payoff is 2, regardless of player 2's action. So player 1 does not know whether a or b was played. As we will see, many common convergence results would not hold with imperfect monitoring of past outcomes.

The last structural feature concerns the player set. We know that it can either be interpreted as a fixed set of individuals or as a set of roles. In the latter case players are drawn at random from different populations and do not know the identity of their opponents. For this reason this case is called *anonymous*. The main effect of anonymity in a repeated game is that it rules out the incentive to affect the opponents' future behavior, because the probability of meeting again players already met in the past is negligible if the populations are large[20].

The role played by these four elements in the determination of players' behavior can be made clear by representing the decision problem of each player with a simple recursive dynamic programming model. Suppose that each player updates his belief about his opponents' current behavior according to an updating rule $\gamma_i : S_i \times M_i \times \Delta_{-i} \rightarrow \Delta_{-i}$, where Δ_{-i} is an appropriately chosen subset of $\Delta[\Delta(S_{-i})]$. This means that the current belief is a sufficient statistic of past information. Such an updating rule can be derived via Bayes' rule by using appropriate statistical models. The current belief is denoted by g_{-i} and the updated belief is $g_{-i}(\cdot|s_i, S_{-i}(m_i, s_i)) = \gamma_i(s_i, m_i, g_{-i})$. With this notation we stress the dependence of the updated belief on the indirect observation about opponents' behavior induced by the signal m_i when s_i is played. At the beginning of a period, the future updated belief of player i when he is going to play s_i is a random distribution, because the signal has not yet been observed. This random distribution is denoted $g_{-i}(\cdot|s_i, \Psi_{-i}(s_i))$.

[20] We are assuming that no agent has any information about the games in which he did not play. Therefore, if the probability of meeting the same opponents in the future is negligible, the probability of affecting their behavior is also negligible.

Note that if either player i is naive or the game is anonymous (with large populations), the updating rule γ_i does not depend on the first argument; that is, it is independent of the strategy s_i. Since a nonmanipulable observation structure implies that $\Psi_{-i}(s_i)$ is independent of s_i, we can conclude that, overall, when the information is nonmanipulable and learning is either naive or the game is anonymous, player i's random updated belief does not depend on his own strategy s_i.

Using the techniques of recursive programming, we can define a value function $V_i: \Delta_{-i} \rightarrow \mathbf{R}$ as follows:

$$V_i(g_{-i}) = max_{s_i \in S_i} \{U_i(s_i, g_{-i}) + \delta E_{g_{-i}}[V_i(g_{-i}(\cdot|s_i, \Psi_{-i}(s_i)))]\},$$

where $E_{g_{-i}}$ denotes the expected value with respect to the current belief g_{-i}.

TABLE 1. - *A classification table*

PSYCOLOGICAL \ STRUCTURAL		NON-ANONYMOUS REPEATED GAMES: MANIPULABLE OBSERVATION STRUCTURE	NON-ANONYMOUS REPEATED GAMES: NON-MANIPULABLE OBSERVATION STRUCTURE	ANONYMOUS REPEATED GAMES: MANIPULABLE OBSERVATION STRUCTURE	ANONYMOUS REPEATED GAMES: NON-MANIPULABLE OBSERVATION STRUCTURE
SOPHISTICATED PLAYERS	FORESIGHTED PREFERENCES	REPUTATION STICK / CARROT	REPUTATION STICK / CARROT	ACTIVE LEARNING	MYOPIC BEHAVIOR
SOPHISTICATED PLAYERS	MYOPIC PREFERENCES	MYOPIC BEHAVIOR	MYOPIC BEHAVIOR	MYOPIC BEHAVIOR	MYOPIC BEHAVIOR
NAIVE PLAYERS	FORESIGHTED PREFERENCES	ACTIVE LEARNING	MYOPIC BEHAVIOR	ACTIVE LEARNING	MYOPIC BEHAVIOR
NAIVE PLAYERS	MYOPIC PREFERENCES	MYOPIC BEHAVIOR	MYOPIC BEHAVIOR	MYOPIC BEHAVIOR	MYOPIC BEHAVIOR

Then a strategy s_i' is dynamically optimal given i's current belief g_{-i} if

$$U_i(s_i', g_{-i}) + \delta\, E_{g_{-i}}[V_i(g_{-i}(\cdot|s_i', \Psi_{-i}(s_i')))] = V_i(g_{-i}).$$

A look at the value function clarifies how the combination of the four features introduced above can result in myopic behavior. It is clear that player i's behavior is myopic when $\delta = 0$[21]. But it is also myopic when player i's random belief $g_{-i}(\cdot|s_i, \Psi_{-i}(s_i))$ does not depend on s_i, either directly or indirectly; that is, when the observation structure is non-manipulable and learning is either naive or the game is anonymous. In this case the expression that has to be maximized with respect to player i's strategy is given by the sum of two terms: the current utility and a constant term.

The consequences, in terms of behavior, of the four aspects listed so far are summarized in Table 1.

4. *Sophisticated Learning in Non-Anonymous Repeated Games*

In this section we consider models of sophisticated learning. The key assumption here is that players recognize that their opponents' choices are influenced by past history. Despite this common feature, sophisticated learning models encompass fairly different forms of learning ranging from equilibrium to disequilibrium learning.

4.1. *Foresighted Preferences: Kalai and Lehrer's Model*

Kalai and Lehrer (1991a) consider an infinitely repeated n-person simultaneous game with a finite set of strategies where each player has foresighted preferences.

Monitoring of the outcome is perfect. Therefore, since the game is simultaneous, the message at time t is given by the profile of strategies of the component game played at t, denoted by $s_{(t)} = (s_{1,(t)}, \ldots, s_{n,(t)})$. A history of the play up to t is $s^t = (s_{(1)}, \ldots, s_{(t)}) \in S^t$. Let $H = \cup_{t=0}^{\infty} S^t$ be the set of all histories. A pure (closed-loop) superstrategy for player i is given by $r_i: H \rightarrow S_i$.

Players are sophisticated, i.e. each player thinks that the other players play closed loop superstrategies. Then, at the beginning of the play, each player's belief is given by a probability distribution over the set of his opponents' closed loop pure superstrategies. We assume that beliefs satisfy independence across players, therefore they can be represented as a profile of mixed superstrategies (one for each opponent).

(21) Actually it is sufficient that δ be smaller that the ratio of the smallest expected short-run loss deriving from the choice of a strategy that does not maximize the current utility and the largest expected long-run gain derived from such choice.

For simplicity's sake we present Kalai and Lehrer's model assuming that players have beliefs with finite support and that they choose pure superstrategies. This last assumption implies that a profile of superstrategies $r = (r_1, \ldots, r_n)$ corresponds to a single infinite history of play s^{∞}.

Suppose that players' beliefs are *compatible* with the superstrategy profile r. By this we mean that they assign positive probability to s^{∞}, the real history of play generated by r[22]. This assumption implies that players never observe something they thought impossible, and thus permits Bayesian updating of beliefs, but it also places restrictions on unobservable events. Consider, for example, the repeated prisoners' dilemma of Figure d. Suppose that player 1 always plays N and player 2 thinks that, in each component game, player 1 randomizes with probabilities 1/2 on C and 1/2 on N. Then there is no finite history of the play that contradicts player 2's belief but the latter is not compatible with player 1's strategy, as it assigns probability zero to the infinite sequence $N, N, N, \ldots$

Compatibility, however, is weaker than the requirement that players' beliefs contain a "grain of truth"; that is, assign positive probability to the superstrategies actually played by the opponents. Consider once again the prisoners' dilemma and suppose that each player plays his trigger strategy. Then the only infinite history with positive probability is the sequence $(N, N, \ldots)$. Hence player 1's belief that assigns probability one to the event that player 2 plays "tit for tat"; that is, he plays cooperative in the first iteration and afterwards mimics his opponent's past action, is compatible with the actual play but does not contain a 'grain of truth'. The following result holds:

PROPOSITION 3: *If the profile of superstrategies $\hat{r}$ actually played is compatible with the players' beliefs, under Bayesian updating there is a finite time T after which players can anticipate the future play correctly.*

Let G_{-i} be the support of player i's (initial) belief. If two profiles of superstrategies r and $\hat{r}$ determine the same play we write $P(r) = P(\hat{r})$. The set G_{-i} can be partitioned in two subsets $\hat{G}_{-i} = \{r_{-i} \in G_{-i} : P(\hat{r}_i, r_{-i}) = P(\hat{r}_i, \hat{r}_{-i})\}$ and $\overline{G}_{-i} = \{r_{-i} \in G_{-i} : P(\hat{r}_i, r_{-i}) \neq P(\hat{r}_i, \hat{r}_{-i})\}$. A superscript t indicates that the same sets are defined for the updated belief at date t. By the compatibility assumption, $\hat{G}_{-i}$ is not empty. Consider now $\overline{G}_{-i}$. Each profile of superstrategies of player i's opponents $r_{-i} \in \overline{G}_{-i}$, combined with player i's superstrategy $\hat{r}_i$, determines a history different from what player i actually observes. Then, after the first period t in which the play path

(22) If, as in Kalai and Lehrer's model, players are allowed to choose behavior superstrategies, the play results in a probability distribution over the set H. In this case private beliefs are compatible with the strategies if the "objective" probability distribution induced by the strategies is absolutely continuous with respect to the subjective distributions induced by the beliefs.

generated by $\hat{r}$ is different from that of $(\hat{r}_i, r_{-i})$, r_{-i} would be assigned probability zero and thus removed from the updated support G^t_{-i}. Since G_{-i} is finite we can conclude that there is a finite time T_i after which $\overline{G^t_{-i}}$ is empty. Therefore, after T_i, $G^t_{-i} = \hat{G}_{-i}$ and even if player i does not know the superstrategies actually played by his opponents, he exactly forecasts the future play of the game on the play path. Choose T to be the largest T_i. As stated by the theorem above, after T, even if the players' beliefs are not fully correct, they cannot be contradicted by the actual play.

The proof makes clear that the result depends, neither on the assumption of perfect monitoring nor on optimizing behavior hypotheses. These assumptions, however, play an important role in the next proposition.

Indeed, optimizing behavior implies that, after T, the play converges to a conjectural equilibrium (23). Consider player i's belief; since it is given by a profile of mixed superstrategies, in games of perfect recall, by Kuhn's theorem (see Section 2.1), we can consider its equivalent profile of behavior superstrategies, denoted g_{-i}. Let $(r_i|s^t)$ denote player i's superstrategy played, according to r_i, in the subgame starting after history s^t. Then, after T, each player i plays a best reply $r_i|s^t$ to his belief $g_{-i}|s^t$ and $(r_i|s^t, g_{-i}|s^t)$ results in the same play of the profile of superstrategies actually played.

In the spirit of the equivalence result between conjectural equilibria and Nash equilibria of Section 2.4 Kalai and Lehrer prove that:

PROPOSITION 4: *For each conjectural equilibrium there is a Nash equilibrium which induces the same history. Hence, after time T, the play converges to a Nash equilibrium history.*

In order to prove proposition 4 let $r = (r_1, \ldots, r_n)$ be a conjectural equilibrium supported by beliefs $(g_{-1}, \ldots, g_{-n})$ and consider the following behavior superstrategies. On the play path, that is, along the history generated by profile r, play according to r. Out of the play path distinguish two cases. If only one player, say j, has deviated from r, each player other than j plays a local mixed strategy according to j's belief g_{-j}. If more than one player have deviated play arbitrarily (since Nash equilibrium is not concerned with multiple defections). This profile of behavior superstrategies is a Nash equilibrium that results in the same outcome as conjectural equilibrium r.

The propositions above regard a very specific case (finite support beliefs and deterministic play) and are thus only intended to illustrate Kalai and Lehrer's approach. Kalai and Lehrer extend these propositions to the case

(23) Or a private beliefs equilibrium, as Kalai and Lehrer called it.

of general beliefs and behavior superstrategies and show that after a finite time beliefs are accurate only up to a small error term ε and play only approximates that induced by a Nash ε-equilibrium.

Note that Kalai and Lehrer's model does not necessarily encompass equilibrium dynamics. Indeed the sequence of component game outcomes before T is not necessarily part of any equilibrium path of the repeated game. This is perhaps the most important difference with respect to the model described in the next section which provides an example of learning in equilibrium (24).

4.2. *Myopic Preferences: Jordan's Model*

Jordan (1991) studies a repeated simultaneous game $G = (S_1, \ldots, S_n, U_1, \ldots, U_n)$ with incomplete information in which the players know their own payoff function but not the payoff functions of their opponents. The issue of incomplete information, in contrast with many alternative learning models including that of Kalai and Lehrer, is tackled assuming that players have a common prior on the space of possible payoff functions, as in Harsanyi's standard theory of incomplete information games.

Let us suppose that payoff functions are normalized to the unit ball in $\mathbf{R}^{|S|}$ and let $B = \{U_i \in \mathbf{R}^{|S|} : [\Sigma_{s \in S} U_i^2(s)]^{1/2} \leq 1\}$ be the space of payoff functions (25). A common prior satisfying independence across players is a product measure $\beta = \beta_1 \times \ldots \times \beta_n$ on the space of players' types B^n such that, for each player i, β_i is a Borel probability distribution on B.

It is assumed that players have perfect monitoring of the outcome of the stage game and that they have myopic preferences.

To describe the evolution of the system under learning Jordan defines a *Bayesian Strategy Process* (*BSP*). If X is a discrete set and φ a Borel probability distribution on $X \times Y$, let $\varphi(x)$ for $x \in X$ denote the marginal probability $\varphi(\{x\} \times Y)$. As before $s_{(t)}$ denotes the profile of component game strategies played at time t and s^t the history of the play up to t.

A *BSP* for the prior β is a joint probability measure φ on the space $B^n \times S^\infty$ of payoff functions and histories of the play such that:

a) the marginal distribution on B^n agrees with β;

b) if φ assigns positive probability to s^t, then for each profile of strategies $s = (s_1, \ldots, s_n) \in S$, $\varphi(s|s^t) = \varphi_1(s_1|s^t) \cdot \ldots \cdot \varphi_n(s_n|s^t)$;

c) for each player i, each t and each triple $(U_i; s^t, s_i)$ in the support of φ, s_i maximizes $\Sigma_{s_{-i}} U_i(\cdot, s_{-i})\, \varphi_{-i}(s_{-i}|s^t)$ on S_i.

(24) Another process of learning in equilibrium is provided by Crawford and Haller's (1990) analysis of repeated coordination games.

(25) This assumption is necessary to prove convergence. To see how the normalization can be done without affecting the strategic situation see Jordan's example, p. 64.

In order to understand the definition of *BSP* it is useful to see how such a process can be built iteratively. Suppose that, at the initial stage, the players play a Bayesian Nash Equilibrium determined by the prior β (see Section 3.3). Let $s_{(0)}$ be the strategies observed by the players at the end of the first iteration. Suppose also that, in case of multiple Bayesian Nash Equilibria, there is a selection mechanism which couples each n-tuple of payoff functions with an equilibrium strategic profile. Then in the following period player i knows that, for each of his opponents $j \neq i$, the payoff function U_j belongs to the set of payoff functions of player j associated, through the selection mechanism, with strategy $s_{(0),j}$ and accordingly updates his prior using Bayes' rule. Given the updated priors, a new Bayesian Nash Equilibrium is determined, and the process is repeated. Hence a *BSP* is an example of learning in equilibrium as in each iteration players play a Bayesian Nash equilibrium[26].

In each round the players can derive the updated probability $\varphi(\cdot|s^t)$ they assign to the next period strategy profiles $s_{(t+1)}$ conditional on history s^t, from the updated beliefs and the selection mechanism. Let $(U_i)_{i \in I}$ be the "true" vector of payoff functions. A strong condition on the learning process requires that, in the long run, players get their beliefs on their opponents' strategies right. Under such a condition the sequence of probability distributions $\{\varphi(\cdot|s^t)\}_{t>0}$ converges to the set $N\ E((U_i)_{i \in N}) \subseteq \Delta(S)$ of Nash equilibria of the one-shot game. Formally, define a distance on the set $\Delta(S)$ of mixed strategies such that, given $q,\ q' \in \Delta(S)$, $||q - q'|| = \max \{|q(s) - q'(s)| : s \in S\}$. Given a set $A \subset \Delta(S)$ let $||q - A|| = \inf \{||q - q'|| : q' \in A\}$. Jordan proves that:

PROPOSITION 5: *For any BSP nonconvergence occurs only on a set of payoff functions and histories of play with φ-probability zero, that is:*

$$\varphi\{((U_i)_{i \in N},\ s^\infty) : \lim_{t \to \infty} ||\varphi(\cdot|s^t) - N\ E((U_i)_{i \in N})|| = 0\} = 1.$$

If, moreover, the priors are sufficiently uniform[27] *convergence occurs for every vector of payoff functions.*

Notice that this form of learning is different from the one described in Kalai and Lehrer's model. Indeed here learning involves fully correct beliefs (with respect to the "objective" common prior β) about the opponents' strategies: when learning is over, if the preferences are myopic, we obtain a sequence of one-shot equilibria and, since monitoring of the strategies is perfect, the players have correct beliefs. In Kalai and Lehrer's model, on the contrary, after learning beliefs are correct only on the path of play because closed loop superstrategies cannot be perfectly monitored.

(26) Compare *BSP* with Kalai and Lehrer's model which deals with learning in disequilibrium. A classical example of learning in equilibrium for rational expectations equilibrium models is Bray and Kreps (1987).

(27) To avoid technicalities we refer to Jordan's definition 3.4.

5. *Models of Naive Learning*

In this section we discuss some learning models in which players do not fully take into account the effects of learning on opponents' behavior, i.e. players are characterized by naive learning processes. An extreme form of "naiveté" is to believe that the opponents' behavior is governed by a fixed, albeit unknown, objective probability distribution, but we allow for more general processes. The common feature of all these processes is that each player believes that the future behavior of his opponents is not affected by his own current actions[28]. This is justified by some form of bounded rationality or by anonymity.

5.1. *Non-Anonymous Repeated Games with Myopic Behavior*

In non-anonymous repeated games the current action of a player affects his opponents' information, and hence his opponents' future actions. However, limited information and/or limited cognitive and computational abilities can prevent the players from taking this effect into account. In this case they use naive learning procedures. If their observation structures are not manipulable, the induced behavior is myopic. Otherwise there is room for experimentation (provided that the players are sufficiently patient). But even in this case we can invoke some sort of bounded rationality in order to justify heuristic procedures, in which the players are largely guided by short run considerations and occasionally make experiments.

5.1.1. *Milgrom and Roberts' General Learning Processes*

Learning is a highly situation-dependent creative process. Thus only a general formulation of learning behavior can encompass the variety of learning strategies that "intelligent" people can possibly use. For this reason Milgrom and Roberts (1991) choose not to specify a single, detailed description of a learning rule, and instead focus on some general properties that plausible learning processes should have.

(28) Milgrom and Roberts (1991) is a noteworthy exception, because they analyze both naive and sophisticated processes.

The analysis is set in the framework of a class of non-anonymous and simultaneous repeated games with perfect monitoring. Two general properties which the sequence of plays over time $\{s_{(t)}\}$ must satisfy are identified. They lead respectively to *adaptive* and *sophisticated learning*. Players use adaptive learning when they mostly learn from past observations while sophisticated learning allows them to make full use of both past information and knowledge of the game and of players' rationality through eductive reasoning(29). Here we will focus on adaptive learning which (implicitly) assumes myopic behavior.

The basic analytical tool is the *justification operator*. Recall that a justification operator is a monotone correspondence from the power set of the strategy space into itself. Since behavior is myopic, the appropriate operator is the best response correspondence J defined in Section 2.2.

A history of plays $\{s_{(t)}\}$ is *consistent with adaptive learning* if and only if the players eventually choose only strategies that are justified by beliefs that assigns near-zero probability to opponents' strategies that have not been played for a sufficiently long time. That is, any "initial" history of plays is eventually inconsequential in players' models of how others act. Formally, $\{s_{(t)}\}$ must be such that for all dates $\hat{t}$, there exists a date $\bar{t}$ such that for all $t \geq \bar{t}$, $s_{(t)} \in J(\{s_{(k)} : \hat{t} \leq k < t\})$.

Note that typical naive learning processes, such as stationary Bayesian learning, fictitious play or Cournot dynamics, induce histories which are consistent with adaptive learning. But the same is true of some extremely sophisticated learning processes such as the hyperrational Bayesian stochastic processes analyzed by Jordan.

The relevance of Milgrom and Roberts' analytical categories is made apparent by three important results. The first limits the scope of positive equilibrium convergence.

We say that the play path $\{s_{(t)}\}$ *converges* to a distribution $q \in \Delta(S)$ if the empirical frequencies generated by $\{s_{(t)}\}$ approach q and there is a date T after which no profile outside the support of q is played(30).

PROPOSITION 6: *If the play path $\{s_{(t)}\}$ converges to a Nash equilibrium in mixed strategies, then the sequence is consistent with adaptive learning. If a sequence $\{s_{(t)}\}$ converges to a point $\bar{s}$ and it is consistent with adaptive learning, then $\bar{s}$ is a Nash equilibrium.*

In order to prove proposition 6 assume that $\{s_{(t)}\}$ converges to a Nash equilibrium in mixed strategies q. Then there is a date T such that for all

(29) See Gul (1991) for a similar approach.

(30) As we are assuming that the component game is finite this is the appropriate definition. The analysis and the results can be extended to compact-continuous games.

$\hat{t} \geq T$, $\{s_{(k)} : k \geq \hat{t}\} = Supp(q)$. Since $Supp(q)$ is finite, for all such $\hat{t}$ there is a later date $\bar{t}$ such that for all $t \geq \bar{t}$, $\{s_{(k)} : \hat{t} \leq k < t\} = Supp(q)$. By the choice of $\bar{t}$ and the best reply property of Nash equilibria it follows that for all $t \geq \bar{t}$ $s_{(t)} \in Supp(q) \subseteq J(Supp(q)) = J(\{s_{(k)} : \hat{t} \leq k < t\})$, i.e. $\{s_{(t)}\}$ is consistent with adaptive learning. Now assume that $\{s_{(t)}\}$ converges to a profile of strategies $\bar{s}$. Then for all t large enough $s_{(t)} = \bar{s}$ is also consistent with adaptive learning, it follows that $\bar{s} \in J(\{\bar{s}\})$, i.e. $\bar{s}$ is a Nash equilibrium.

The previous proposition states that equilibrium convergence results cannot be obtained with a broader class of learning models than those encompassed by consistency with adaptive learning. Since positive equilibrium convergence results can be proved for very sophisticated learning processes, as in Jordan's model, we see that such processes generate histories of play which are consistent with adaptive learning.

The second result provides an evolutive foundation of rationalizability (iterated dominance), a concept which is essentially eductive.

PROPOSITION 7: *If $\{s_{(t)}\}$ is consistent with adaptive learning, then the players eventually play only rationalizable strategies, i.e. for every positive integer m there exists a date t_m after which (i.e. for all $t \geq t_m$) $s_{(t)} \in J^m(S)$.*

The proof is quite simple. We can state by definition that $J^0(S) = S$. Hence the inductive proposition is trivially true for $m = 0$. Now assume that for all $t \geq t_m$, $s_{(t)} \in J^m(S)$. By compatibility with adaptive learning, we know that there must be a date t_{m+1} such that for all $t \geq t_{m+1}$, $s_{(t)} \in J(\{s_{(k)} : t_m \leq k < t\})$. By the induction hypothesis, $\{s_{(k)} : t_m \leq k < t\} \subseteq J^m(S)$; by the monotonicity of J, $J(\{s_k : t_m \leq k < t\} \subseteq J(J^m(S)) = J^{m+1}(S)$. Hence for all $t \geq t_{m+1}$, $s_{(t)} \in J^{m+1}(S)$ and the theorem follows by induction.

A corollary of these results is that if a game is dominance-solvable, a history of plays $\{s_{(t)}\}$ is consistent with adaptive learning if and only if it converges to the unique Nash equilibrium.

The scope of proposition 7 is somehow limited. Indeed, if payoffs were drawn at random, the expected number of dominated strategies would approach zero as the number of players and the number of strategies get larger. Therefore for generic large games, adaptive learning implies no restrictions. However, Milgrom and Roberts (1990) show that rationalizability implies surprisingly strong restrictions on adaptive processes in the class of games with strategic complementarities(31).

Assume for simplicity that the strategy sets are compact intervals in R and that each player has a well-defined best response function(32). We

(31) See also Krishna (1991).

(32) This means that for each strategic profile s_{-i} there is one and only one strategy s_i which maximizes $U_i(\cdot, s_{-i})$.

say that a game displays *strategic complementarities* if the best response functions are upward-sloping, i.e. the marginal return of increasing one's strategy rises when the opponents' strategies increase. This class of games is important in many economic applications. Examples are the stability problem in the Arrow-Debreu model when the "gross substitutes" condition is satisfied, and Cournot and Bertrand competition, under some regularity conditions on the demand and cost functions. Milgrom and Roberts (1990) prove that:

PROPOSITION 8: *In games with strategic complementarities the (inf and sup) bounds on players' joint behavior implied by iterated dominance and pure Nash equilibria coincide. In particular, when there is a unique pure Nash equilibrium the sequence of strategic profiles converges to this equilibrium if and only if it is consistent with adaptive learning.*

The result is indeed a generalization of the standard result on stability for the Cournot dynamics illustrated in Figure e, where R_1 and R_2 are the reaction functions for firm 1 and 2, respectively (the picture may represent a Bertrand duopoly with differentiated products). Figure e shows that with the same reaction functions the iteration of the dominance operator also converges to the unique Nash equilibrium. If $[p_1, P_1]$ is the set of possible strategies (prices) for firm 1, then the set of possible best reply

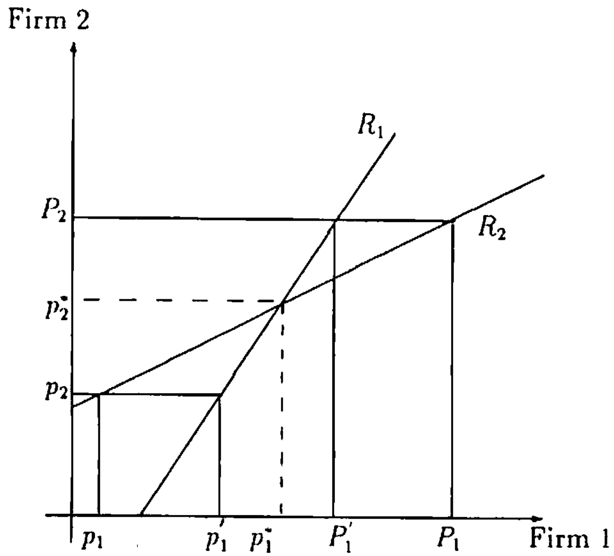

Figure e. - A Bertrand duopoly with differentiated products

for firm 2 is $[p_2, P_2]$. Thus firm 1 can consider only its best reply set $[p_1', P_1']$ and so on, until convergence to the unique Nash equilibrium (p_1^*, p_2^*) is obtained.

The definition of consistency with adaptive learning can be easily extended to general games with imperfect monitoring of the strategies. Proposition 6 can be adapted to this more general context, by replacing Nash equilibrium with conjectural equilibrium. Proposition 7, however, crucially depends on the assumption of perfect monitoring of the strategies. This is made apparent by the example of Figure c where player 1 only observes his own realized payoff. In this case a history of plays, though consistent with adaptive learning (in a properly defined sense), may converge to the non-rationalizable conjectural equilibrium (A, a). Furthermore, in this more general framework, the assumption of myopic behavior must rely on myopic preferences, since otherwise there are incentives for active learning. The latter is the main theme of the next two sections.

5.1.2. *Fudenberg and Kreps' Model: Boundedly Rational Players and Experimentation*

In a forthcoming monograph Fudenberg and Kreps (1988)(33) present a theory of learning in extensive non-anonymous games(34) which has a distinguishing feature. In almost any model of learning the first step is the description of a rational or reasonable learning behavior. The outcomes yielded by this behavior are then analyzed. Fudenberg and Kreps' model follows the opposite route since the specification of learning behavior is driven by the desire to sustain the Nash equilibrium concept and its refinements. As the assumptions about this behavior have unappealing aspects, doubts are cast on the relevance of the Nash equilibrium itself.

A formal presentation of Fudenberg and Kreps' model is beyond the scope of this survey, in light of the preliminary nature of their results. We therefore only describe, mainly by means of examples, a few insights which are highlighted in their paper about the problems implied by convergence to Nash equilibrium.

Nash equilibrium requires that players hold correct beliefs about the strategies played by their opponents. Thus, if we want to justify it, our

(33) Different versions of the monograph are available. We refer to *A Theory of Learning, Experimentation and Equilibrium in Games*, Draft 0.11, July 1988. Note that in the most recent version of the monograph, *Learning and Equilibrium in Games*, Partial Draft 0.62, August 1991, which is substantially different from Draft 0.11, attention is focused on simultaneous games and fictitious play while the theory of learning in extensive games we present here is not touched upon.

(34) They consider a subclass of extensive games, i.e. games where, in any possible play, no player can ever be called upon to move twice.

main concern is to provide a theory of how and when correct beliefs might emerge. We know that the repetition of the game by itself might not be enough for players to understand how the opponents play at unreached information sets, even if the outcome is observed. (Recall the game of Figure b and what we said about manipulability.) Thus, if we want to avoid getting stuck in a non-Nash outcome, we need a theory explaining how players can receive evidence of what happens off the path of play.

The theory provided by Fudenberg and Kreps is that players are uncertain about their beliefs and that they experiment to see whether they are correct. Moreover players are boundedly rational. Hence they are not able to work out a grand dynamic strategy to experiment optimally. They simply choose, most of the time, strategies which are optimal in the short run and they occasionally try other strategies. Occasional experiments, however, must be frequent enough to enable players to accumulate sufficient evidence about opponents' behavior.

The first assumption of Fudenberg and Kreps' model responds to this criterion. Roughly, it says that experiments are such that for each information set that occurs infinitely often, each action is played infinitely often.

Note that this assumption is stronger than it seems. Indeed, for it to be satisfied it is not enough that a player plans to experiment infinitely often; he must actually do so. To grasp this point consider the game of Figure f. Suppose that at time t player 1 assigns probability less than 1/2 to the event that player 2 plays A. Then his optimal strategy is D. Suppose also that he experiments action A with probability $1/t$. This experimentation rate tends to zero but it does not vanish too quickly ($\Sigma_{t=0}^{\infty} 1/t$ is divergent). Therefore player 1's action A is chosen an infinite number of times and player 2's information set occurs infinitely often. However, if player 2 experiments likewise, although he plans to experiment infinitely often, her strategy A is chosen only a finite number of times ($\Sigma_{t=0}^{\infty} 1/t^2$ is finite)[35].

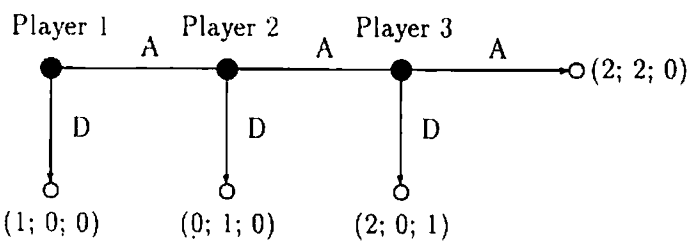

Figure f. · A game in extensive form

[35] In game of perfect information, as in the game of Figure f, the first assumption implies that each information set is reached infinitely often. This is not true in general; see for example Figure 1.3 in Fudenberg and Kreps (1988).

Besides being boundedly rational, players are naive: it is assumed that at information sets that occur infinitely often they asymptotically form their conjectures on their opponents' behavior on the basis of a simple count of how often the various actions have been played in the past, assuming that the opponents' strategies are independently implemented. Such a behavior is a sort of asymptotical fictitious play. Note that this statistical procedure would be reasonable if the sequence of observations at an information set were exchangeable but, in this model, players' behavior may be highly nonstationary[36].

The last assumption about players is that they are asymptotically almost certain to play actions that almost maximize their expected short run payoffs, given their conjectures. Optimization is *almost* certain because there is always a small probability of non-optimal actions as experimentation persists in the long run.

Note that in order to conform to this last assumption the experimentation rate must vanish with time, because the probability of short-run non-optimal behavior must be asymptotically very low. We can then conclude that the rate of experimentation involved by the behavior assumptions given is of a very special kind. The number of experiments must decrease over time because it is assumed that asymptotically players almost always choose optimal strategies. However it must not decrease too quickly, or else enough evidence would not be collected about opponents' behavior.

Given the persistence of experimentation in the long run, we cannot demand that play approaches some definite stationary mixed strategy in this model. Indeed, given that the players are experimenting, there is always a small chance that the behavior will take away from any point. For this reason the definition of stable and unstable strategies is given in terms of high probability.

A profile of strategies $q \in \Delta(S)$ is *locally stable* with respect to players' behavior if there exists a partial history of outcomes such that, with conditional probability close to one, the outcome distribution induced by such behavior converges to the outcome distribution associated with q.

A profile of strategies $q \in \Delta(S)$ is *locally unstable* with respect to players' behavior if, with probability one, the outcome distribution induced by such behavior eventually stays outside a neighborhood of the outcome distribution associated with q.

Fudenberg and Kreps(1988) prove that a non-Nash outcome is locally unstable with respect to all behaviors satisfying the previous assumptions

[36] It is difficult to say how players behave when evidence clearly rejects the hypothesis of stationary and independently chosen strategies even if it seems natural to assume that they abandon it. For this reason Fudenberg and Kreps consider only histories of play that pass a consistency test.

and a Nash equilibrium outcome is locally stable for some behavior that satisfies the previous assumptions.

The second part of the above result does not give any idea of the range of behaviors that make a particular Nash equilibrium stable. Even if no general results are given by Fudenberg and Kreps, the sketch of the proof suggests that behavior must be carefully tailored for each particular Nash equilibrium.

An interesting line of research departing from Fudenberg and Kreps' model is what they call a complete theory; that is, a theory which, by imposing conditions on the likelihood of various experiments, leads to different sorts of equilibrium refinements. In a complete theory, then, experimentation plays a fundamental role. However, until it is not endogenously derived, as in this model, it is open to the criticism of being *ad hoc*. Endogenous experimentation is the focus of the model of the next section.

5.2. *Anonymous Repeated Games*

In anonymous repeated games with large populations of agents for each player/role, the current choice of a single agent almost surely does not affect his future opponents' behavior. Therefore the agent correctly neglects this effect and behaves myopically, unless there is room for active learning. It is still generally true that "learning affects what is to be learned", making the statistical distribution of strategies in the population nonstationary. But we focus on learning procedures that would be correct only in a stationary environment.

5.2.1. *The Model of Fudenberg and Levine: Steady State Learning and Optimal Experimentation*

Fudenberg and Levine (1990) present a model of learning in finite extensive games in which each player knows the extensive form and his own payoff function, but possibly ignores his opponents' payoff functions.

The game is anonymous: thus, for each player/role we can consider the corresponding population. It is assumed that each population consists of a continuum of agents with total mass one composed as follows: each agent lives T periods, in each period $1/T$ new agents enter into each population and $1/T$ agents of age T exit. Thus each population is stationary and there are $1/T$ agents in each generation. There is a double infinite sequence of periods ..., -1, 0, 1, ... In every period each agent is randomly and independently matched with one agent from each of the opponents' population. From the assumptions above the probability of meeting a player of age t is $1/T$.

In the first part of their paper Fudenberg and Levine define an equilibrium concept relevant in this context of anonymous learning, called self-confirming equilibrium. This notion corresponds to that of anonymous conjectural equilibrium defined in Section 2.4 when the players observe the outcome of the component games in which they have played. The definition is repeated here for convenience in a somewhat different form.

A profile of distributions $(q_1, ..., q_n)$ is a *self-confirming equilibrium* if for each player i and each strategy s_i in the support of q_i there exists a belief $g_{-i} \in \Delta(S_{-i})$ (possibly depending on s_i) such that

(a) g_{-i} justifies s_i,

(b) for each opponent j and each information set $h \in H_j$, if h is reached with positive probability according to the profile (s_i, q_{-i}), then the subjective belief (s_i, g_{-i}) assigns to each action $a \in A(h)$ the 'correct' conditional probability, i.e. the same conditional probability that can be derived by (s_i, q_{-i}).

A self-confirming equilibrium does not necessarily correspond to a Nash equilibrium (recall the game of Figure b). However, because of the manipulability of the observation structure, there are incentives for active learning. Are these incentives strong enough to rule out non-Nash outcomes?

The answer provided by Fudenberg and Levine in the second part of their paper refers to a particular setting, since they focus on the case of a steady state population. In this case stationary Bayesian learning is a correct procedure of statistical inference and the optimal rate of experimentation can be endogenously determined by dynamically optimal superstrategies. No convergence explanation is offered for the steady state assumption but we argue that this approach is both analytically convenient and methodologically defensible. Indeed, from an analytical point of view the correctness of the agents' procedures makes it possible to apply the powerful limit theorems of mathematical statistics. On the other hand, from a methodological standpoint, the encompassing of the steady states of some appropriate dynamic system can be regarded as a minimum requirement for the relevance of a static equilibrium concept.

Suppose that each agent believes that he faces a time invariant distribution of opponents' strategies (this turns out to be true in a steady state), but he is not sure about the true distribution. At the beginning of each period his (updated) belief is represented by a strictly positive density function on the space $\times_{j\neq i}\Delta(S_j)$. Assume also that each agent belonging to population i is born with the same prior g^0_{-i}.

Let y_i denote player i's personal history including his past strategies and the terminal nodes reached in the games he has played. Let Y_i be the set of all such histories. The state of the system at the beginning of each period is given by a vector $\theta = (\theta_1, \ldots, \theta_n) \in \Delta(Y_1) \times \ldots \times \Delta(Y_n)$, which specifies the fractions of each population i in each experience category y_i.

Assume that the behavior of all i-agents is described by the same superstrategy $r_i : Y_i \to S_i$. Then one can compute an associated function $\theta \to f[\theta]$ describing the dynamic behavior of the system: $f[\theta]_i(y_i)$ is the fraction of population i in the experience category y_i if the previous state was θ given that the agents' behavior is governed by the rules $r_1, \ldots, r_n$. A *steady state* of the system is a state θ such that $\theta = f[\theta]$, i.e. it is a fixed point of f. Since f is a polynomial function from a compact and convex set into itself, a steady state exists for every n-tuple of rules. Clearly in a steady state the fractions of each population playing each strategy (as determined by θ and the rules r_i) are also stationary.

Assume that each agent maximizes the expected present value of his payoffs given his initial belief g^0_{-i}, denoted $E \sum_{t=1}^{T} \delta^t u_i(z_{ti})$, with $\delta \in [0, 1)$. This is a well-defined dynamic programming problem, with at least one deterministic solution r_i. Fudenberg and Levine study the limit properties of steady states associated with dynamically optimal rules as the parameters T (life span) and/or δ (discount rate) increase to infinity and one respectively. They prove that:

PROPOSITION 9: *For fixed full-support priors g^0, and $\delta < 1$ each limit point θ of the associated sequence of steady states as $T \to \infty$ corresponds to a self-confirming equilibrium q^0.*

Note that here we do not have "convergence in real time" as in the previous models. In the steady state "convergence in real time" has already done its job and the theorem only characterizes the limit properties of a sequence of dynamic models indexed by T, the length of agents' life span.

The proof of proposition 9 is based on the idea that long-lived agents will eventually stop experimenting and they will play to maximize their current expected payoff, because the option value, or the expected future gain associated with experimentation, becomes too small. Therefore condition (a) in the definition of a self-confirming equilibrium holds in the limit. Furthermore the agents' beliefs about opponents' actions actually chosen (hence observed) almost certainly become approximately correct. Since all the actions on the limit path are chosen many times, condition (b) also holds in the limit.

Note, however, that it may well be the case that players stop experimenting too soon and do not accumulate enough evidence about opponents' behavior off the limit path. Therefore the limit point need not correspond to a Nash equilibrium. This is the multi-agent analogue of the observation that in a two-armed bandit problem an impatient agent can repeatedly choose the wrong arm. (See, for example, Rothschild 1974).

Will the experimentation rates of long-lived and extremely patient players be high enough so as to approximate a Nash equilibrium? The following proposition provides a partially affirmative answer:

PROPOSITION 10: *For fixed priors g^0_{-i}, there is a function $T(\cdot)$ such that if $\delta_m \to 1$ and $T_m \geq T(\delta_m)$ (roughly speaking, this means that the lifetimes go to infinity more quickly than the discount factor tends to 1), every limit point θ of the associated sequence of steady states corresponds to a Nash equilibrium q^0.*

The answer provided here is only partial for two reasons: (i) it considers a quite special double limit, and (ii) it does not say anything about the plausibility of the limit equilibrium. It is worth noting that, at least in simple games such as that of Figure a, the limit equilibrium is the plausible one [37]. But general results of this sort have not yet been proved.

5.2.2. *Canning's Model: An Anonymous Repeated Game with Noise*

Canning (1989) [38] has developed a theory of learning and equilibrium for simultaneous games with anonymous interaction.

The basic model is quite simple. There are two N-agent populations. In every period each agent is randomly paired with an agent from the other population to play a finite two-person [39] simultaneous game $G = (S_1, S_2, U_1, U_2)$. It is assumed that each agent knows both strategy sets and his payoff function but not the payoff function of his opponent. Each agent has perfect monitoring of the strategy played by his current opponent; however, he does not observe the strategies played by the other agents in his opponents' populations. If we assume that the agents have bounded memory of length T, each agent's experience can be described by a history of length T given by the T-uple of strategies played by the opponents in the previous T stages. A state of the system, denoted by $x \in X$, is described by an array, formed by $T \times N \times 2$ elements, that contains a history of T elements for each agent in each population. Note that, because of the assumptions of finite populations and bounded memory, there is a finite set of states.

At the end of each period, each agent can die with probability α. Whenever an agent dies, he is replaced by a new agent with a null history.

---

(37) The assumption of a strictly positive prior rules out weakly-dominated strategies and it is easy to check that the implausible equilibria of the aforementioned games display weakly-dominated strategies.

(38) The paper has been subsequently divided in two parts, Canning (1990a) and (1990b).

(39) The generalization to n-person games is straightforward.

At every stage each agent plays a pure strategy, arbitrarily selected from the best reply correspondence, given his beliefs. Canning assumes that the agents' beliefs are given by the empirical distribution of past observations. Then he considers a sort of fictitious play where, however, the computation of the frequencies is limited to the last T periods. Whenever an agent has observed nothing yet, his belief is given by a uniform probability distribution over his opponent's strategies.

Note that since the players are naive empiricists with non-manipulable observation structures their myopic behavior is justified independently of their intertemporal preferences.

Let $\omega_t \in \Delta(X)$ be the probability distribution over states at time t. The elements listed so far, given by the matching rule for the players, the best reply mappings, the rule for belief updating and the probability of death, completely determine a Markov operator $P : \omega_t \rightarrow \omega_{t+1}$ through which a probability measure over states at time t induces another probability measure over states at time $t+1$.

Note that, whatever the state of the system, there is a positive probability α^{2N} that all agents die and hence the state is reached where all (newborn) agents have null history. By standard theorems of probability theory, this is sufficient to obtain the following:

PROPOSITION 11: *There exists a unique probability distribution* ω^* *such that* $P\omega^* = \omega^*$. *Furthermore, the proportion of time that the system is in state* x *converges with probability one to* $\omega^*(x)$ *for any initial state.*

This proposition contrasts sharply with the nonconvergence results of the literature on fictitious play (see e.g. Shapley 1964). The intuitive reason of this dramatic change is that the introduction in a deterministic dynamic system of a history-independent random term (death and replacement in our case) yields a stochastic dynamic system with better convergence properties[40]. To grasp this intuition consider the following example.

An agent characterized by infinite horizon and no discounting has to cut a tree and sell it at some date $t = 0,1,2, \ldots$. In each period the tree grows by one unit. At each date t the agent has to decide whether to cut or not, given the state of the system, i.e. the size of the tree s_t. Clearly the agent has no incentive to cut the tree and we obtain a divergent determistic dynamic system $s_{t+1} = s_t + 1$. Now assume that at each date the agent cuts the tree (or, alternatively, the tree falls) with a very small probability $p > 0$. Take an arbitrary probability distribution ω_t over the possible sizes $s = 0,1,2, \ldots$ at date t and compute ω_{t+1}. Clearly $\omega_{t+1}(0) = p$ and $\omega_{t+1}(k) = \omega_t(k-1)(1-p)$ (the probability of being in state $k-1$ at t times the probability of not being cut at period t). By induction we get the unique stationary distribution $\omega^*(k) = p(1-p)^k$, which is also the long-run distribution of sizes.

[40] See Canning (1990b). Kandori et al. (1991) and Young (1992) also rely on this.

The second part of Canning's paper focuses on the definition of a social equilibrium and the relation between this notion and Nash equilibrium.

Let $q_i(x) \in \Delta(S_i)$ be the empirical distribution of strategies in population i given state x: $q_i(x)$ is simply a mixed strategy that represents the average behavior of population i, given that each agent chooses a (pure) best reply to the empirical frequencies determined by x (or a best reply to the uniform distribution, if he is newly born). Define its expected value with respect to $\omega \in \Delta(X)$ as follows:

$$q_i(\omega) := \sum_{x \in X} q_i(x)\, \omega(x).$$

Then a *social equilibrium* of the system is defined as the mixed strategy pair

$$q(\omega^*) = (q_1(\omega^*), q_2(\omega^*)).$$

By the previous results it follows that:

PROPOSITION 12: *A social equilibrium exists and is unique for every* $\alpha > 0$. *Furthermore the time average of the pair of mixed strategies* $(q_1(x_t), q_2(x_t))$ *actually played at time t converges to the social equilibrium.*

As the time average of the strategies played converges it seems reasonable that the players learn the social equilibrium if given enough time; that is, if they have a low probability of death and a long memory. However as T gets large and α small, the convergence speed slows down because the longer the life and the memory, the greater the probability of meeting the same opponents and hence the correlation between players' belief. Thus to obtain convergence it is necessary to limit somehow this correlation by increasing the number of agents in each population.

On the basis of a sligthly modified version of his model[41] Canning shows that:

PROPOSITION 13: *In the case of an infinite number of agents, for* α *sufficiently small and T sufficiently large, an equilibrium distribution of histories yields a mixed strategy pair which is close to a Nash equilibrium of the game G. Moreover, the result also holds for large (but finite)* N.

[41] He supposes that in each round only two players, instead of the entire population, are called to play and every matching is assigned equal probability by the matching rule. Moreover time is assumed to be continuous.

Proposition 13 resembles Fudenberg and Levine's proposition 10. In both cases convergence to equilibrium is obtained when the life span goes to infinity and the presence of newborn agents — that is, agents that take history-independent actions — helps stabilize the system and avoid the kind of cycling characteristic of Shapley's example. Here, however, there is no room for experimentation, because G is a simultaneous game and there is perfect monitoring of the strategies played. In this case, self-confirming equilibria coincide with Nash equilibria and there is no need to consider extremely "patient" players, in order to achieve convergence to Nash equilibria.

REFERENCES

AUMANN, R. (1976): "Agreeing to Disagree", *Annals of Statistics*, 4, pp. 1236-1239.

BATTIGALLI, P. (1987): "Comportamento razionale ed equilibrio nei giochi e nelle situazioni sociali", unpublished dissertation, Milano: Bocconi University.

BATTIGALLI, P. (1988): "Il concetto d'equilibrio nei modelli strategici e parametrici", *Giornale degli Economisti e Annali di Economia*, 47, pp. 99-118.

BATTIGALLI, P. and D. GUAITOLI (1988): "Conjectural Equilibria and Rationalizability in a Macroeconomic Game with Incomplete Information", Working Paper IEP N. 1988-6, Bocconi University.

BERNHEIM, D. (1986): "Axiomatic Characterizations of Rational Choice in Strategic Environments", *Scandinavian Journal of Economics*, 88, pp. 473-488.

BINMORE, K. (1987): "Modeling Rational Players: Part I", *Economics and Philosophy*, 3, pp. 179-214.

BINMORE, K. (1992): *Fun and Games. A Text on Game Theory*. Lexington, Massachussets: D. C. Heath and Company.

BINMORE, K. and A. BRANDENBURGER (1988): "Game Theory and Common Knowledge", Discussion Paper, London School of Economics.

BLUME, L., M. BRAY and D. EASLEY (1982): "Introduction to the Stability of Rational Expectation Equilibrium", *Journal of Economic Theory*, 26, pp. 313-317.

BRAY, M. (1990): "Rational Expectations, Information and Asset Markets", in F. HAHN, ed., *The Economics of Missing Markets, Information and Games*. Oxford: Oxford University Press.

BRAY, M. and D. KREPS (1987) "Rational Learning and Rational Expectations", in G. FEIWEL, ed., *Arrow and the Ascent of Modern Economic Theory*. New York: Macmillan.

BROSSEAU, V. and A. KIRMAN (1991): "The Dynamics of Learning in Mis-Specified Models", mimeo, E.U.I., Florence.

BROWN, G.W. (1951): "Iterative Solution of Games by Fictitious Play", in T.C. KOOPMANS, ed., *Activity Analysis of Production and Allocation*. New York: Wiley.

CANNING, D. (1989): "Social Equilibrium", mimeo, Pembroke College, Cambridge University.

CANNING, D. (1990a): "Social Equilibrium", Economic Theory Discussion Paper N. 152, University of Cambridge.

CANNING, D. (1990b): "Average Behavior in Learning Models", Economic Theory Discussion Paper, N. 156, University of Cambridge.

CRAWFORD, V. P. and H. HALLER (1990): "Learning How to Cooperate: Optimal Play in Repeated Coordination Games", *Econometrica*, 58, pp. 571-550.

EICHBERGER, J. (1992): "Nash Equilibrium and Learning", in J. BORLAND, J. CREEDY and J. EICHBERGER, eds., *Recent Developments in Game Theory*. Brokfield, Vermont: Gower.

EICHBERGER, J., H. HALLER and F. MILNE (1990): "Naive Bayesian Learning in 2 x Matrix Games", Discussion Paper, N. 9071, Center for Economic Research, Tilbu University.

FRIEDMAN, D. (1991): "Evolutionary Games in Economics", *Econometrica*, 59, pp. 637-66

FUDENBERG, D. and D. KREPS (1988): "A Theory of Learning, Experimentation ar Equilibrium in Games", mimeographed monograph (partial draft 0.11), MIT and Stanfor

FUDENBERG, D. and D. KREPS (1991): "A Theory of Learning, Experimentation ar Equilibrium in Games", mimeographed monograph (partial draft 0.56), MIT and Stanfor

FUDENBERG, D. and D. LEVINE (1990): "Steady State Learning and Self Confirmi Equilibrium", mimeo, MIT and UCLA.

FUDENBERG, D. and J. TIROLE (1991): *Game Theory*. Cambridge: The MIT Press.

GILBOA, I. and A. MATSUI (1991): "Social Stability and Equilibrium", *Econometrica*, 5 pp. 859-867.

GUL, F. (1991): "Rationality and Coherent Theories of Strategic Behavior", mimeo, Stanfo University.

HAHN, F. (1973): *On the Notion of Equilibrium in Economics. An Inaugural Lecture*. Cambridg Cambridge University Press.

HAHN, F. (1977): "Exercises in Conjectural Equilibria", *Scandinavian Journal of Economics*, 7 pp. 210-226.

HAHN, F. (1978): "On Non-Walrasian Equilibria", *Review of Economic Studies*, 45, pp. 1-1

HARSANYI, J. (1967): "Games with Incomplete Information Played by 'Bayesian' Player Part I", *Management Science*, 14, pp.159-182.

HARSANYI, J. and R. SELTEN (1988): *A General Theory of Equilibrium Selection in Game* Cambridge: MIT Press.

HAYEK, F.A. von (1937): "Economics and Knowledge", *Economica*, 4, pp. 33-54.

HENDON, E., H. J. JACOBSEN, M. T. NIELSEN and B. SLOTH (1990): "A Learning Proce for Games", Discussion paper, 90-20, Institute of Economics, University of Copenhage

HENDON, E., H. J. JACOBSEN, M. T. NIELSEN and B. SLOTH (1991): "Learning, Tracing an Risk Dominance", mimeo, University of Copenhagen.

JORDAN, J. (1991): "Bayesian Learning in Normal Form Games", *Games and Economic Behavio* 3, pp. 60-81.

KALAI, E. and E. LEHRER (1991a): "Rational Learning Leads to Nash Equilibrium. (A Ne Extended Version)", Working Paper N. 925, Northwestern University.

KALAI, E. and E. LEHRER (1991b): "Private Beliefs Equilibrium", Working Paper N. 92 Northwestern University.

KANDORI, M., G. MAILATH and R. ROB (1991): "Learning, Mutation and Long Run Equilibri in Games", Discussion Paper 71, University of Pennsylvania.

KANEKO, M. (1987): "The Conventionally Stable Sets in Non-Cooperative Games with Limite Observations I: Definitions and Introductory Arguments", *Mathematical Social Science* 19, pp. 93-128.

KANEKO, M. and N. TAKASHI (1991): "Final Decision, the Nash Equilibrium and the Solv bility in Games with Common Knowledge of Logical Abilities. Logical Investigation int Game Theory III", mimeo, Virginia Polytechnic Institute and State University.

KREPS, D. (1990): *Game Theory and Economic Modelling*. Oxford: Clarendon Press.

KRISHNA, V. (1991): "Learning in Games with Strategic Complementarities", mimeo, Harvar Business School.

KUHN, K.W. (1953): "Extensive Games and the Problem of Information", in K.W. KUHN an A. TUCKER, eds., *Contributions to Game Theory*. Vol. 2. Princeton: Princeton Universit Press.

LITTLECHILD, S.C. (1982): "Equilibrium and the Market Process", in I. KIRZNER, ed *Method, Process and Austrian Economics. Essays in Honor of L. Von Mises*. Lexingto D.C. Heath and Company.

MARIMON, R., E. MCGRATTAN and T.J. SARGENT (1990): "Money as a Medium of Exchang in an Economy with Artificially Intelligent Agents", *Journal of Economic Dynamics an Control*, 14, pp. 329-373.

MILGROM, P. and J. ROBERTS (1990): "Rationalizability, Learning and Equilibrium in Games with Strategic Complementarities", *Econometrica*, 58, pp. 1255-1277.

MILGROM, P. and J. ROBERTS (1991): "Adaptive and Sophisticated Learning in Normal Form Games", *Games and Economic Behavior*, 3, pp. 82-100.

MIYASAWA, K. (1961): "On the Convergence of the Learning Process in a 2 x 2 Non-Zero-Sum Two-Person Game", Research Memorandum, 33, Princeton University.

MOULIN, H. (1986): *Game Theory for Social Sciences*. New York: New York University Press.

MYERSON, R.B. (1991): *Game Theory. Analysis of Conflict*. Cambridge: Harvard University Press.

NACHBAR, J. H. (1990): "'Evolutionary' Selection Dynamics in Games: Convergence and Limit Properties". *International Journal of Game Theory*, 19, pp. 59-89.

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

RAMPA, G. (1991): "Giochi, concorsi di bellezza e slot machines. Riflessioni sui processi di apprendimento", *Economia Politica*, 7, pp. 429-455.

ROTHSCHILD, M. (1974): "A Two-Armed Bandit Theory of Market Pricing", *Journal of Economic Theory*, 9, pp. 185-202.

RUBINSTEIN, A. and A. WOLINSKY (1990): "Rationalizable Conjectural Equilibrium: Between Nash and Rationalizability", Working Paper N. 19-90, The Foerder Institute for Economic Research.

SELTEN, R. (1991): "Evolution, Learning and Economic Behavior", *Games and Economic Behavior*, 3, pp.3-24.

SHAPLEY, L. (1964): "Some Topics in Two-Person Games", in M. DRESHER, L. SHAPLEY and A. TUCKER, eds., *Advances in Game Theory*. Princeton: Princeton University Press.

SMITH, L. (1991): "Error Persistence, and Experiential Versus Observational Learning", Working Paper N. 19-91, The Foerder Institute for Economic Research.

TAN, T. and S. WERLANG (1987): "The Bayesian Foundations of Solution Concepts of Games", *Journal of Economic Theory*, 45, pp. 370-391.

VAN DAMME, E. (1991): *Stability and Perfection of Nash Equilibria* (2nd edition). Berlin: Springer.

YOUNG, H. P. (1992): "The Evolution of Conventions", mimeo, University of Maryland.

Summary JEL C72, C73, D83

LEARNING AND CONVERGENCE TO EQUILIBRIUM IN REPEATED STRATEGIC INTERACTIONS: AN INTRODUCTORY SURVEY

In recent years, there has been increasing interest in the hypothesis that an appropriate concept of equilibrium might be obtained as the result of a learning process in game theory. We propose a classification of this literature based on four features of the players' psychology and the structure of the game together with an annotated review of some contributions. We provide an introduction to the relevant concepts of game theory, and learning results are often given in simplified versions to facilitate the comparison between different approaches. The *conjectural equilibrium* notion, first proposed by Hahn, emerges as a unifying concept.

Riassunto JEL C72, C73, D83

APPRENDIMENTO E CONVERGENZA ALL'EQUILIBRIO IN SITUAZIONI DI INTERAZIONE STRATEGICA RIPETUTA: UNA RASSEGNA INTRODUTTIVA

L'ipotesi che individui razionali possano convergere ad un equilibrio come risultato di un processo di apprendimento è stata un'idea centrale in teoria dei giochi negli ultimi anni. Qui proponiamo una classificazione di questa letteratura basata su quattro caratteristiche della psicologia dei giocatori e della struttura dei giochi. Dopo aver introdotto una facile "guida" dei concetti di teoria dei giochi rilevanti, presentiamo una selezione di alcuni fra i contributi più significativi, spesso presentando i risultati in maniera semplificata per facilitare i confronti fra i diversi approcci. L'*equilibrio congetturale*, inizialmente proposto da Hahn, risulta essere un concetto organizzatore di tutta la letteratura.