Discussion Paper No. 895

RATIONAL LEARNING LEADS TO NASH EQUILIBRIUM

by

Ehud Kalai*

and

Ehud Lehrer*

March 1990

---

*Department of Managerial Economics and Decision Sciences, J. L. Kellogg Graduate School of Management, Northwestern University, 2001 Sheridan Road, Evanston, Illinois 60208.

The authors wish to thank Robert Aumann, Larry Blum, David Easley, Itzhak Gilboa, Sergiu Hart, James Jordan, Dov Monderer, Dov Samet, and Vernon Smith for helpful discussions and suggestions. Kalai's research was partly supported by Grant No. SES-9011790 from the National Science Foundation, Economics.

Abstract

Rational Learning Leads to Nash Equilibrium

by

Ehud Kalai and Ehud Lehrer

Two players are about to play a discounted infinitely repeated bimatrix game. Each player knows his own payoff matrix and chooses a strategy which is a best response to some private beliefs over strategies chosen by his opponent. If both players' beliefs contain a grain of truth (each assigns some positive probability to the strategy chosen by the opponent), then they will eventually (a) accurately predict the future play of the game and (b) play a Nash equilibrium of the repeated game. An immediate corollary is that in playing a Harsanyi-Nash equilibrium of a discounted repeated game of incomplete information about opponents' payoffs, the players will eventually play an equilibrium of the real game as if they had complete information.

1. Introduction

The concept of Nash (1950) equilibrium has become central in game theory, economics, and other social sciences. Yet the process by which the players learn to play it, if they do, is unknown. This is not surprising for a game which is played only once since the players do not have much opportunity to learn. However, in infinitely repeated games, where the players do have enough time to learn the behavior of their opponents, one would expect them to learn to play a Nash equilibrium.

If a repeated game involves incomplete information, a second issue of learning arises. The justification of a Nash equilibrium now requires the existence of a commonly known prior distribution of the uncertain parameters in the game. Other than general verbal discussions about the players having a common prior when they were very young, or before they were born (behind the veil of ignorance), we have no satisfactory way of dealing with this problem.

In this paper we study an infinitely repeated two person game with discounting. We assume that each player knows his own payoff matrix and chooses a strategy which is a best response to his private beliefs about his opponent's strategy. We show that if both players' beliefs contain a grain of truth (each assigns some positive probability to his opponent's strategy) then eventually:

(i) they will accurately predict the future play of the game, and

(ii) they will play according to a Nash equilibrium of the repeated game.

Moreover, this learning takes place without reliance upon a common

prior distribution or common knowledge of the strategies played or other parameters of the game. Every Nash equilibrium can result from such learning.

While the experimental literature of Smith and others (see, for example, McCabe-Rassenti-Smith, 1989) have shown that agents in repeated interactive situations do learn to play Nash equilibrium, no theoretical explanation for this phenomenon has been provided. This is in spite of the existence of a continuously growing game theoretic literature on repeated games with or without complete information (see Aumann (1981) and Mertens (1986) for surveys that are already outdated; see the forthcoming book by Mertens-Sorin-Zamir (1990) for state-of-the-art knowledge on repeated games with and without complete information), and the interest in the topic of learning in economics (e.g., Blum-Bray-Easley (1982), Easley-Kiefer (1986), Grandmont-Laroque (1990), Jordan (1985), McLennan (1987), Woodford (1990), and references therein).

Motivated by the important applications of Nash equilibrium in economics, recent researchers have started studying processes of learning to play it. For example, papers by Brock-Marimon-Rust-Sargent (1989), Canning (1989), Crawford (1988), Fudenberg-Kreps (1988), Fudenberg-Levine (1990), Jordan (1989, 1990), Linhart-Radner-Schotter (1989), Milgrom-Roberts (1989), Selten (1988), and Stanford (1990) have studied such processes. In several of these papers the authors construct specific learning rules and specific dynamic environments and show that together they lead to a Nash equilibrium after a long enough period of real or fictitious play. Thus, these papers show that positive theories of learning to play Nash equilibrium can be constructed.

In this paper we show that for two rational players, participating in a real (not fictitious) repeated game, learning to play Nash equilibrium is unavoidable. Our assumptions regarding the rationality of the players involve two items.

1. The players seek to maximize their overall expected present and future payoffs evaluated under discounting. Learning is not a goal in itself here but is, rather, a consequence of an overall payoff maximization plan. It is obtained as the real game progresses and in this sense it may be thought of as learning by playing, paralleling the economic literature on "learning by doing" (see Arrow (1962)).

2. Learning is modeled by Bayesian updating of prior beliefs. This follows the traditional approach of games of incomplete or imperfect information, e.g., Kuhn (1953), Harsanyi (1967), and Aumann and Maschler (1967), which was recently used by Jordan (1989) to construct a process, converging to Nash equilibrium, in an incomplete information game played by myopic players.

We depart from the standard assumptions of game theory by not requiring that the players have common knowledge or even common beliefs about each others' strategies and the uncertain parameters of the game (we do not prohibit such assumptions but they are not necessary in our model). This assumption is replaced by a weaker one requiring that the initial private beliefs of each player assign some positive probability to the strategy actually used by his opponent. Even though stronger versions of this assumption are used in almost all papers using the incomplete information approach of Harsanyi, we think that this assumption can be weakened and we discuss in the body of the paper some examples that are important for future

research.

An interesting corollary to the main result of this paper relates to Harsanyi-Nash equilibria of a two person infinitely repeated game under discounting with two sided incomplete information about opponents' payoff matrices. The corollary states that at any such equilibrium the players will eventually play according to a Nash equilibrium of the infinitely repeated underlying realized game (the one with complete information) as if the uncertainty was not present. (In other words, players learn to play the "right" game!) This corollary is closely related to Jordan's (1989) results. If we consider the case where the discount parameter is close to zero, i.e., the players are myopic, our corollary reconfirms the result of Jordan regarding convergence to a Nash equilibrium of the one shot game.

Before moving to the formal model and general results, we discuss here the special case of finitely many pure strategies. This discussion should serve to explain the approach and intuition before getting involved with the general notations and the probability computations.

We start with two players facing a finite bimatrix game and assume that each knows his own payoff matrix. The players will play this stage game infinitely many times with perfect monitoring and evaluate their payoff streams according to some fixed discount parameter. A player's strategy for such a repeated game is <u>rational</u> if (a) he has a finite-support probability distribution describing his beliefs about the pure strategy to be used by his opponent, and (b) his own pure strategy is an optimal response to these beliefs. Two rational strategies are <u>compatible</u> if each player's beliefs assign a positive probability to his opponent's strategy.

The main result of this paper is that, given any two such compatible

rational strategies of the repeated game, after a finite time T the players must play according to a Nash equilibrium of the repeated game. We do not claim that the players will necessarily learn the identity of the true game (i.e., the opponent's payoff matrix), but their actions along the play path will be the same as the actions of players playing a Nash equilibrium of the real fully known repeated game (the one with complete information). Moreover, each player will be predicting correctly his opponent's actions along the equilibrium play path.

To give a sketch of the proof for this special case, assume that player two's beliefs about player one's strategy are described by a private prior probability distribution over finitely many pure strategies $f_1, f_2, \ldots, f_N$ of player one. Assume without loss of generality that player one is actually using the strategy f_1. Player two is using a strategy g. For each of the strategies f_i the play path generated by (f_i, g) either coincide forever with the play path of (f_1, g) or there is some time when they differ. This means that there is a finite time T by which player two can rule out some of the strategies not used by player one and after T no further ruling out is possible. In other words, the remaining strategies f_j not ruled out by player two by time T all generate the same play path with g as f_1 does. In effect we just argued that after the time T, player two will accurately predict the future play of the game. Applying the same argument to player two we conclude that there is a time T after which both players accurately predict the future play of the game.

To construct an equilibrium of the repeated game that coincides with the play path of the original game after time T we do the following. After histories that are continuations of the play path after time T we let the

players take the same actions as in the original play. After other histories, we let each player play the mixed strategy coinciding with the beliefs of his opponent about him.

It is clear that the strategies just constructed yield the same equilibrium path as in the original game after time T. It is also clear that they are an equilibrium of the repeated game, since now each player's beliefs (to which he is best responding) coincide with his opponent's actual strategy.

Notice that in the above model it was never necessary for a player to know his opponent's payoffs. The only interaction between the two players was through the flow of information regarding each other's actions and in the assumption of compatibility. This assumption essentially replaces the common knowledge of solution assumption. While this assumption still has a flavor of commonality, it is much weaker than full common knowledge. All that it requires is that a player's private prior distribution regarding the strategy of an opponent be dispersed enough to allow positive probability to what the opponent actually does.

With the above result the move to the case of incomplete information games is easy. Assuming that the $m \times n$ bimatrix game was drawn from a finite set of such games using some commonly known prior distribution and that each player is informed about his own payoff matrix before playing the repeated game. In a pure strategy Nash equilibrium of this incomplete information repeated game, each player chooses one repeated game pure strategy for each one of his types (matrix realizations). Moreover, each player's beliefs about his opponent's type are given by the distributions derived from the original prior and conditioned on his own realization.

With the pair of realized types prescribing the strategy to be played by the players we are back in the situation described above. Thus, after a finite time T each player will predict correctly his opponent's future actions on the play path and they will be following a play path of a Nash equilibrium of the repeated game really drawn. This is the case since the strategies of the true types actually drawn, are rational with respect to their beliefs regarding the behavior of the opponent's other types. And they are compatible because of the common prior used by both players.

It is important to note that in the above discussions we started with pure strategies and constructed quasi-pure strategies, i.e., strategies that are pure on the play path but randomize off it. This is not surprising. Players learn to predict with high precision future behavior on the equilibrium path after observing enough past behavior. On the other hand, in the above model they do not learn to predict their opponent's behavior off the equilibrium path. The randomization off the equilibrium path reflects this uncertainty.

In the remaining sections we generalize the above results to mixed strategies and a broader class of beliefs.

2. The Model and Assumptions

2.1 The Repeated Game

Two players are about to play an infinitely repeated game. The stage game is described by the following components.

1. Two finite sets Σ_1, Σ_2 of actions with $\Sigma = \Sigma_1 \times \Sigma_2$ denoting the set of action combinations.
2. Two payoff functions $u_i : \Sigma \to \mathbb{R}$.

We let H_t denote the set of histories of length t, i.e., Σ^t, and $\bar{H} = \cup_t H_t$ be the set of all histories. A (behavior) strategy of player i is a function $f: \bar{H} \to \Delta(\Sigma_i)$ with $\Delta(\Sigma_i)$ denoting the set of probability distributions on Σ_i. Thus, a strategy specifies how a player randomizes over his choices of actions after every history of past actions (see Appendix 1, "Behavior and Mixed Strategies," for important elaborations).

We assume that each player knows his own payoff function and that the game is played with perfect monitoring, i.e., the players are fully informed about all realized past action combinations at each stage.

2.2 The Payoffs

Let λ_i, $0 < \lambda_i < 1$ be the discount factor of player i and let x_i^t denote player i's payoff in stage t. If player one plays f and player two plays g then the payoff of player i in the repeated game is defined as

$$U_i(f,g) = (1 - \lambda_i) \sum_{t=0}^{\infty} E_{f,g}(x_i^{t+1})\lambda_i^t,$$

where $E_{f,g}$ denotes the expected value calculated with respect to the probability measure induced by (f,g).

2.3 Behavior Assumptions

Let $\tilde{g}$ be a strategy denoting player one's belief over the strategy that player two will play in the repeated game. If player two plays the strategy g we say that $\tilde{g}$ contains a grain of g (or that $\tilde{g}$ contains a grain of truth) if $\tilde{g}$ can be obtained from a mixed strategy choosing g with probability α and some other strategy $\hat{g}$ with probability $1 - \alpha$ for some positive number α.

$0 < \alpha \leq 1$. (Appendix I contains useful elaborations on the definitions of mixed and behavior strategies and their relation to each other.)

As usual, we say that a strategy of player one, f, is a best response to a strategy of player two (a belief of how player two plays) $\tilde{g}$, if $U_1(\bar{f},\tilde{g}) - U_1(f,\tilde{g}) \leq 0$, for all strategies $\bar{f}$ of player one. We say that f is an ε-best response ($\varepsilon \geq 0$) if the same inequalities hold but with ε replacing 0 in the right side. The corresponding definitions apply to player two.

For our main result we will be assuming that each player plays a best response strategy to some beliefs over the strategy of his opponent and that each player's belief contains a grain of truth regarding the strategy actually chosen by the opponent. Formally, we assume that the players play the pair of strategies (f,g) with f (resp. g) being best responses to some strategy $\tilde{g}$ (resp. $\tilde{f}$) and that $\tilde{g}$ (resp. $\tilde{f}$) has a grain of truth.

Before proceeding to the statements of the results it is useful to elaborate and show some examples where the behavior assumptions described above are relevant.

2.4 Remarks and Examples

Assuming that a player's beliefs are described by a behavior strategy of the opponent, allows for a larger set of beliefs than may seem at first. This is due to Kuhn's theorem (see Appendix I) stating that every mixed strategy can be represented by a single behavior strategy. Thus, if a player's belief consists of a probability distribution over a whole family of possible opponent's strategies, then it can be reduced to an aggregate belief represented by one behavior strategy.

The assumption that the players' beliefs contain a grain of truth is important. Later we show an example illustrating that without this assumption, or other alternative assumptions of this type, there are beliefs that do not allow for any learning over time.

The following examples illustrate situations where two players best respond to beliefs containing a grain of truth.

Example 1: A Nash equilibrium in a two person repeated game.

If (f,g) is such an equilibrium we have $\tilde{g} = g$ and $f = \tilde{f}$ and each player best responding to his beliefs which are the full truth.

Example 2: A variation of the prisoner's dilemma game.

Each player has two possible actions: c (cooperate) and d (double-cross). The stage game payoffs of player one are given by the traditional prisoner's dilemma payoffs: $u_1(c,c) = 3$, $u_1(d,c) = 4$, $u_1(c,d) = 0$, and $u_1(d,d) = 1$. Player two, on the other hand, strictly prefers cooperation to noncooperation, no matter what player one does and his payoffs are given by $u_2(c,c) = 4$, $u_2(c,d) = 3$, $u_2(d,c) = 1$ and $u_2(d,d) = 0$.

We let C denote the constant cooperating strategy, D denote the constant double-crossing strategy, and c-tft denote the strategy in which a player starts by cooperating and continues by mimicking his opponent's previous action.

Suppose player one believes that player two's strategy can be represented by the following mixed strategy: D with probability .90, C with probability .05, and c-tft with probability .05. The following learning strategy, L, is a best response to these beliefs. Start with the action d.

If player two played d (now player one's posterior belief is that player two is playing D with probability one) continue with D forever. If player two started with c, however, player one does not know if player two plays C or c-tft and each has a posterior probability of .5. Waiting one period and observing player two's response to player one's initial d will indicate which of the two strategies player two really uses. If player one's discount parameter is sufficiently low, then playing d while waiting is optimal. Now, if player two cooperated, he must be a constant cooperator (i.e., he is playing C), and player one's optimal response from here on is to play D. If player two played d, however, player one's posterior shows him to be a tft-er and player one's best response is to play c to bring back cooperation and then to play c-tft.

It is easy to check that, for a proper discount factor, the above learning strategy is a best response of player one to his initial beliefs. This shows, in particular, that best responding to beliefs involving several possibilities is not totally passive and can call for some <u>active experimentation</u>.

To complete this example, assume that player two believes that player one plays c-tft with probability .99 and the learning strategy L with probability .01. He plays c-tft, which is an optimal response to these beliefs.

Note that the resulting play in this example is $(\begin{smallmatrix}d\\c\end{smallmatrix})(\begin{smallmatrix}d\\d\end{smallmatrix})(\begin{smallmatrix}c\\d\end{smallmatrix})(\begin{smallmatrix}c\\c\end{smallmatrix})(\begin{smallmatrix}c\\c\end{smallmatrix})\ldots$. The play $(\begin{smallmatrix}c\\c\end{smallmatrix})(\begin{smallmatrix}c\\c\end{smallmatrix})\ldots$ is an equilibrium path of this game and learning to play it took three steps. The fast convergence is obtained here because the players' beliefs were restricted to a small number of low complexity strategies.

Example 3: A repeated game with incomplete information.

We consider a repeated game, described as before by a set of action combinations $\Sigma = \Sigma_1 \times \Sigma_2$. However, we assume that the pair of payoff functions (u_1,u_2) is drawn from a finite set $(u_1^k)_{k\in K} \times (u_2^\ell)_{\ell\in L}$ according to a commonly known prior distribution $(p_{k,\ell})_{(k,\ell)\in K\times L}$. We assume that each player is told his own realized payoff function and they play a Harsanyi-Nash equilibrium (q,r) of the incomplete information infinitely repeated game.

We recall that such an equilibrium consists of two vectors, $q = (q^1,q^2,\ldots,q^{|K|})$ and $r = (r^1,r^2,\ldots,r^{|L|})$ of strategies for the repeated game. The interpretation is that player one (resp. player two) will play the strategy q^k (resp. r^ℓ) if his k^{th} (resp. ℓ^{th}) "type" is realized. At such equilibrium each strategy, q^k, of player one (and, similarly, player two) is best response to the mixed strategy obtained by mixing the opponent's type strategies $r^1,r^2,\ldots,r^{|L|}$ according to the conditional probability distribution $p_{\bullet|k}$. Thus, once a realization $(\bar{k},\bar{\ell})$ is obtained we have a situation described by our assumptions with $f = q^{\bar{k}}$, $g = r^{\bar{\ell}}$ and $\tilde{g}$ being induced by the vector of strategies r with the probabilities $p_{\bullet|\bar{k}}$ and $\tilde{f}$ being induced by the vector q with the probabilities $p_{\bullet|\bar{\ell}}$. The assumption of common prior distributions guarantee that the players' beliefs have a grain of truth.

The examples above bring to focus one difference between our behavioral assumptions and the ones of traditional game theory. We have players' uncertainty expressed over strategies of the opponent rather than on what is considered in traditional game theory to be the fundamentals, i.e., the

unknown parameters of the game. Traditional game theory would require that players have some distribution over the possible games (as in Example 3) and, if the games have a multiplicity of equilibrium, some selection criterion would determine the chosen one. As Example 3 illustrates, our model is more general. Since a distribution over games and equilibria selected will yield, after a (possibly large) number of additional computations, a distribution over opponents' strategies.

3. Statement of the Main Results

By a path we mean an infinite sequence of action combinations, i.e., an element of $H = \Sigma^{\mathbb{N}}$. For any path p and time $t \in \mathbb{N}$ we denote by p_t the t-prefix of p (the element in H_t consisting of the first t action combinations of p).

Let (f,g) and (f',g') be two pairs of strategies. We denote by μ, μ' the measures on H (endowed with the σ-algebra generated by all finite histories) induced by (f,g) and (f',g'), respectively. We also use μ and μ' to denote probabilities over finite histories. For example, $\mu(h)$ will denote the probability that the history h will be played when the strategies (f,g) are used.

Definition: Let $\varepsilon > 0$. We say that (f,g) plays ε-like (f',g') if for every history h $|\mu(h) - \mu'(h)| \leq \varepsilon$.

In a subsequent section we will introduce a stronger notion of similar play which we call playing the same up to ε. It will require that the two pairs of strategy combinations induce the same probability measure on a

subset of infinite paths of measure at least $1 - \varepsilon$ according to both of them. Theorem 2 will actually be proven with this stronger notion of similarity.

Definition: Let f be a strategy, $t \in \mathbb{N}$ and $h \in H_t$. The induced strategy f_h is defined as follows:

$$f_h(h') = f(hh') \text{ for any } h' \in H_r,$$

where hh' is the concatenation of h with h', i.e., the history of length $t + r$ whose first t elements coincide with h followed by the r elements of h'.

The following theorem states that players with beliefs containing a grain of truth eventually learn to predict accurately the future play of the game. We state it only for player one and omit the symmetric statement for player two.

Theorem 1: Given $\varepsilon > 0$, let f and g be strategies of player one and player two, respectively, and let $\tilde{g}$ be a strategy representing player one's beliefs. Suppose $\tilde{g}$ contains a grain of g. For almost all paths p (according to the measure induced by (f,g)) there is a time T such that for all $t \geq T$, $(f_{p_t}, \tilde{g}_{p_t})$ plays ε-like (f_{p_t}, g_{p_t}).

In other words, the probability of any history being played according to what one believes, i.e., according to a measure induced by himself and his beliefs about his opponent, is essentially the same as the probability

of the history actually being played.

Notice that, in Theorem 1, we did not make any assumptions on the strategy f of player one. It essentially states that Bayesian updating by itself will lead to a correct prediction of the important parts of player two's strategy, namely, player two's actual future play in response to f. It does not state that player one would learn to predict player two's future randomizations in response to actions not taken by himself. (It is true, though, that this learning will take place for any strategy of player one. However, the length of time required may be different for every strategy.) For this reason, the next theorem is only obtained for Nash equilibrium and not for subgame perfect equilibrium.

Theorem 2: Suppose f and g are best responses to beliefs $\tilde{g}$ and $\tilde{f}$, respectively, and that $\tilde{g}$ and $\tilde{f}$ contain a grain of g and f, respectively. Then for every $\varepsilon > 0$ and for almost all (with respect to the probability measure induced by (f,g)) paths p there is a time $T = T(p,\varepsilon)$ such that for every $t \geq T$ there exists an ε-equilibrium (f',g') of the repeated game satisfying (f_{p_t}, g_{p_t}) plays ε-like (f',g').

In other words, given any $\varepsilon > 0$, with probability one there will be some time T after which the players will play ε-like an ε-Nash equilibrium. This means that if players start with beliefs containing only a grain of truth about their opponent's strategies then, in the long run, their individually rational behavior must be essentially the same as behavior described by an ε-Nash equilibrium.

We turn now to implications of the main theorem in the theory of repeated games with incomplete information. Returning to the set up

described in Example 3, we start with a stage game whose action combinations are given by a set $\Sigma = \Sigma_1 \times \Sigma_2$. We assume that a pair of utility functions (u_1,u_2) is drawn from a finite set $(u_1^k)_{k\in K} \times (u_2^\ell)_{\ell\in L}$ according to a commonly known prior distribution $(p_{k,\ell})_{(k,\ell)\in K\times L}$. Each player is informed of his own realized index, say, $\bar{k}$ and $\bar{\ell}$, and proceeds to play the repeated game. As usual, it is implicitly assumed that the model and the prior distribution are common knowledge.

A strategy q for player one in such a game consists of a choice of a repeated game behavior strategy as discussed earlier, for each possible realized "type" of himself, $q = (q^k)_{k\in K}$. Similarly, a strategy of player two is a vector of repeated game strategies, $r = (r^\ell)_{\ell\in L}$. The pair of strategies (q,r) is a Harsanyi-Nash equilibrium if each type k of player one is best responding according to his utility u_1^k by playing q^k against the mixed strategy of player two, $\tilde{r}^k$ obtained by mixing $(r^1,r^2,\ldots,r^{|L|})$ with the probabilities $p_{\bullet|k}$ describing the conditional distribution on |L| given the realized k value. The analogous requirement holds for player two types. Each strategy r^ℓ must be a best response to the strategy $\tilde{q}^\ell$ describing his conditional beliefs of the strategy of player one.

For every realized choice of k and ℓ, $\bar{k}$, $\bar{\ell}$, we now have a situation described by the main theorem with $f = q^{\bar{k}}$, $g = r^{\bar{\ell}}$, $\tilde{g} = r^{\bar{k}}$ and $\tilde{g} = q^{\bar{\ell}}$. Thus, the following result follows immediately from the previous theorem.

Corollary 1: Let (q,r) be a Harsanyi-Nash equilibrium of the game described above. For any realized pair of types $(\bar{k},\bar{\ell})$ let the real game be described by $(u_1^{\bar{k}},u_2^{\bar{\ell}})$. For any $\varepsilon > 0$ and almost all paths p drawn by $(q^{\bar{k}},r^{\bar{\ell}})$ there is a time $T = T(p,\varepsilon)$ such that for every $t \geq T$ there exists an ε-Nash

equilibrium (q',r') of the real game $(u_1^{\bar{k}},u_2^{\bar{\ell}})$ satisfying $(q_{p_t}^{\bar{k}},r_{p_t}^{\bar{\ell}})$ plays ε-like (q',r').

In other words, equilibrium strategies of the incomplete information game will eventually converge to equilibrium strategies of the complete information realized game. The players are led by optimal strategies to "learn" complete information equilibrium behavior, even if they do not learn the identify of the true game.

4. Proof of Theorem 1

The following notations and observations will be used for proving both theorems.

Denote by $\tau,\tilde{\sigma},\tilde{\tau}$ the probability measures on H, induced by (f,g), $(\tilde{f},g)$, and $(f,\tilde{g})$, respectively. $(f,\tilde{g})$ and $(\tilde{f},g)$ define two sequences of random variables $\{X_t\}_t$ and $\{Y_t\}_t$ attaining values in[1] Δ^2. $X_t(p)$ and $Y_t(p)$ are the posteriors of PI and PII over $\{g,\hat{g}\}$ and $\{f,\hat{f}\}$, respectively, after observing the history p_t. As the posteriors are derived from a Bayesian updating, $\{X_t\}$ and $\{Y_t\}$ are martingales on $(H,\tilde{\sigma})$ and $(H,\tilde{\tau})$, respectively (see, for example, Hart (1985)). Furthermore, since all the coordinates of X_t and Y_t are bounded between 0 and 1, by the martingale convergence theorem (Shiryayev, 1984) they converge almost surely to X_∞ and Y_∞. Before proceeding, we need the following lemma.

Lemma 1:

(a) With probability 1 the first coordinate of X_∞ is positive, and

(b) with probability 1 the first coordinate of Y_∞ is positive.

---

[1] Δ^2 is the unit simplex in $\mathbb{R}^2$.

In other words, X_∞ assigns, almost surely, a positive probability to the strategy player two really plays--that is, g. And Y_∞ assigns, almost surely, a positive probability to the strategy player one really plays--that is, f.

Proof: We prove (a). By a similar method one can get a proof of (b). In order to simplify the notations we denote $g_1 = g$, $g_2 = \hat{g}$, $\alpha_1 = \alpha$, and $\alpha_2 = 1 - \alpha_1$. For $h \in H_t$ denote by $\text{prob}_{g_j}(h)$ the probability that the pair of strategies (f,g_j) will result in the history h. Understanding h also as a set of paths we may define a measure τ_2 as follows: $\tau_2(h) = \text{prob}_{g_2}(h)$.

With a slight abuse of notation,[2] but without any confusion, the posterior of g_j after h is $X_t(h)(j)$. Thus,

$$X_t(h)(1) = \text{prob}_{g_1}(h)\alpha / \Sigma_{j=1}^{2} \text{prob}_{g_j}(h)\alpha_j \text{ for all } h \in H_t.$$

Thus,

$$(1) \qquad X_t(h)(i)\ \text{prob}_{g_2}(h)\alpha_2 = \text{prob}_{g_1}(h)\alpha(1 - X_t(h)(1))$$

Since $\text{prob}_{g_2}(h)\alpha_2 = \tilde{\tau}(h) = \alpha\tau(h) = \alpha_2\tau_2$ and $\text{prob}_{g_1}(h) = \tau(h)$ we can rewrite (1) as follows:

$$(2) \qquad \int_A X_t(1)d\tau_2 = \alpha \int_A (1 - X_t(1))d\tau, \text{ for all } A \in \mathcal{F}_r,$$

where $\mathcal{F}_r$ is the field generated by histories of length r, $r \leq t$.

---

[2] Recall that X_t is defined on infinite paths and not on finite histories.

Notice that τ and τ_2 are absolutely continuous with respect to $\tilde{\tau}$. Thus, the limit X_∞ satisfies (by the bounded convergence theorem):

(3a) $\int_A X_t(1)d\tau_2 \to_{t\to\infty} \int_A X_\infty(1)d\tau_2$ for all $A \in \mathcal{F}_r$

and

(3b) $\alpha \int_A (1 - X_t(1))d\tau \to_{t\to\infty} \alpha \int_A (1 - X_\infty(1))d\tau$ for all $A \in \mathcal{F}_r$.

From (2), (3a) and (3b) we derive

(4) $\int_A X_\infty(1)d\tau_2 = \alpha \int_A (1 - X_\infty(1))d\tau$ for all r and $A \in \mathcal{F}_r$.

As (4) hold for r and every $A \in \mathcal{F}$, we deduce

(5) $\int_A X_\infty(1)d\tau_2 = \alpha \int_A(1 - X_\infty(1))d\tau$ for all measurable sets A in H.

For $A = \{X_\infty(1) = 0\}$ the left side of (5) is 0 and the right side is $\alpha \int_A d\tau = \alpha\tau(A)$. Since $\alpha > 0$, $\tau(A) = 0$, which concludes the proof of Lemma 1. //

To complete the proof of Theorem 1, let $\varepsilon > 0$ and let $\sigma = \tau$. We have to show that for almost all (w.r.t σ) paths p there is a time $T = T(p,\varepsilon)$ s.t. for any $t \geq T$ $(f_{p_t}, \tilde{g}_{p_t})$ plays ε-like (f_{p_t}, g_{p_t}). Define $A = \{p | X_\infty(p)(1) = 1\}$. The conclusion of Theorem 1 clearly holds for almost every $p \in A$. As for $p \notin A$, by Lemma 1, $\sigma\{p | X_\infty(p)(1) > 0\} = \sigma\{X_\infty(p)(1) > 0\} = 1$. Thus, for almost all $p \notin A$ there is a time $T = T(\varepsilon,p)$ s.t. if $t \geq T$

$|X_t(p)(1) - X_\infty(p)(1)| < \varepsilon \min[X_\infty(p)(1),(1 - X_\infty(p)(1))]/16$. Therefore, for any $s,t \geq T$

$$(6) \qquad |X_t(p)(1) - X_s(p)(1)| < \varepsilon \min[X_\infty(p)(1),(1 - X_\infty(p)(1))]/8.$$

Moreover,

$$(7a) \qquad X_t(p)(1) > X_\infty(p)(1)/2$$

and

$$(7b) \qquad 1 - X_t(p)(1) > (1 - X_\infty(p)(1))/2.$$

Let h_t be the history that satisfies $p_t = p_s h_t$ where $s \geq T$. Notice that

$$(8) \qquad X_t(p)(1) = X_s(p)(1) \, \mathrm{prob}_{f_{p_s},g_{p_s}}(h_t) \,/$$

$$/[X_s(p)(1) \, \mathrm{prob}_{f_{p_s},g_{p_s}}(h_t) + (1 - X_s(p)(1)) \, \mathrm{prob}_{f_{p_s},g_{p_s}}(h_t)].$$

One derives (the tedious calculation is omitted) from (6), (7a), (7b) and (8) that

$$|\mathrm{prob}_{f_{p_s},g_{p_s}}(h_t) - \mathrm{prob}_{f_{p_s},g_{p_s}}(h_t)| < \varepsilon, \text{ for all } s > T,$$

which concludes the proof. //

5. Proof of Theorem 2

We use the notation used in the previous section and we add the notations $f_1 = f$ and $f_2 = \tilde{f}$. Denote for any T and $\eta > 0$

$$B_\eta^T = \{p \in H | \text{there are } s,t \geq T \text{ s.t. } \|X_s(p) - X_t(p)\|_\infty > \eta\}$$

The set B_η^i is the set of the "bad" paths. Notice that the martingale convergence theorem implies $\tilde{\tau}(B_\eta^t) \to_{t\to\infty} 0$ for all $\eta > 0$. Thus,

(9) $$\tau(B_\eta^t) \to_{t\to\infty} 0 \text{ for all } \eta > 0.$$

Lemma 2: Fix $\eta > 0$. For almost all (with respect to τ) $p \in H$ there exists a time $t = t_X(p,\eta)$ s.t. if $T > t$ then

(10) $$\tau(B_\eta^T | p_T) = \tau\{p' \in H | p'_T = p_T \text{ and } p' \in B_\eta^T\}/\tau\{p' | p'_T = p_T\} < \eta.$$

Proof: Suppose to the contrary that there is a positive measure set (w.r.t. τ) $D \subseteq H$ satisfying: for all $p \in D$ there is an infinite sequence $\{T_n\}$ of times s.t. $\tau(B_\eta^{T_n} | p_{T_n}) \geq \eta$.

Define for any t

$$D^t = \{p \in D | \tau(B_\eta^t | p_t) \geq \eta\}.$$

Thus, any $p \in D$ is included in infinitely many D^t. I.e. (recall that $D^t \subseteq D$),

$$D = \cap_{n=1}^{\infty} \cup_{t \geq n} D^t.$$

In other words,

$$D = \cup_{t \geq n} D^t \text{ for all } n.$$

By the definition of D^t,

$$\tau(B_\eta^n) \geq \tau(\cup_{t \geq n} D^t)\eta = \tau(D)\eta \text{ for all } n.$$

This contradicts (9). //

Similarly to B_η^T define for $\{Y_t\}$:

$$C_\eta^T = \{p \in H | \text{there are } s,t \geq T \text{ s.t. } \|Y_s(p) - Y_t(p)\|_\infty > \eta\}.$$

Similarly to Lemma 2, there is $t = t_Y(p,\eta)$ s.t. if $T > t$, then

(10') $$\tau(C_\eta^T | p_t) < \eta \text{ for almost all } p \in H.$$

Define,

$$Z_\eta^T = B_\eta^T \cup C_\eta^T.$$

(10) and (10') imply

(10") $\tau(Z^T_\eta | p_t) < 2\eta$ if $T > t(p,\eta) = Max\{t_X(p,\eta), t_Y(p,\eta)\}$.

Denote $W^T_\eta = H \backslash Z^T_\eta$, the set of η-"good" paths.

For Lemma 3 recall that $X_t(p)$ is the posterior of player one over $\{g_1, g_2\}$ after observing the history p_t. $X_t(p)(j)$ denotes the probability assigned to $g_j, j \in \{1,2\}$.

Lemma 3: There is a function $\delta: (0,1) \to (0,1)$ s.t. $\delta(\eta) \to_{\eta \to 0} 0$ and for every $p \in W^T_\eta$ the following hold:

(a) $X_t(p)(1), X_t(p)(2) > \eta$, then

$$\| g_1(p_t) - g_2(p_t) \|_\infty < \delta(\eta) \text{ for all } t \geq T.$$

(b) If $Y_t(p)(1), Y_t(p)(2) > \eta$, then

$$\| f_1(p_t) - f_2(p_t) \|_\infty < \delta(\eta) \text{ for all } t \geq T.$$

In words, for every η-"good" path p, if player 1 believes after time t (after observing p_t) that player two plays with a probability greater than η the strategies g_1 and g_2, then g_1 and g_2 behave about the same (up to $\delta(\eta)$).

Proof: Let $p \in W^T_\eta$.

For a history $h \in H_t$ and $b = (a_1, a_2) \in \Sigma$, the concatenated history starting with h and proceeding with b is denoted by hb. Let b be the one satisfying $p_{t+1} = p_t b$. The Bayesian updating implies:

(11) $$X_{t+1}(p_t b)(i) = \text{prob}(i|p_t b) - \text{prob}(i|p_t,a_2) =$$
$$g_i(p_t)(a_2)\ \text{prob}(i|p_t)/\Sigma_{j=1}^{2}\ g_j(p_t)(a_2)\ \text{prob}(j|p_t)$$
$$= g_i(p_t)(a_2)X_t(p_t)(i)/\Sigma_{j=1}^{2}\ g_j(p_t)(a_2)X_t(p_t)(j),$$
$$\text{for all } i \in \{1,2\}.$$

Since p_t and $p_t b$ are both prefixes of p, $X_{t+1}(p_t b)(i)$ and $X_t(p_t)(i)$ are close to each other up to η. As (11) holds for every $i \in \{1,2\}$, we conclude that if $x_t(p_t)(1)$ and $X_t(p_t)(2)$ are greater than η, then $g_1(p_t b)$ is close to $g_2(p_t b)$ up to $\delta(\eta)$, which tends to zero as $\eta \to 0$. //

Now we are in a position to complete the proof of Theorem 2. Actually, we are going to prove more than what is required.

Definition: Let μ and μ' be the measures induced by (f,g) and (f',g') on H, respectively. We say that (f,g) and (f',g') play the same up to ε if μ and μ' coincide on a set of paths, say, A, of a measure greater or equal to $1 - \varepsilon$, i.e., $\mu(B) = \mu'(B)$ for all measurable sets $B \subseteq A$ and $\mu(A) = \mu'(A) \geq 1 - \varepsilon$.

Sometimes we say that (f',g') plays the same as (f,g) up to ε.

Remark: It is clear that if (f,g) and $(f'g')$ play the same up to ε then (f,g) plays ε-like (f',g').

Fix $\varepsilon > 0$. For almost every $p \in H$ we will find a time T and we will construct for every $t > T$ an ε-Nash equilibrium (f',g') which plays the same

as (f_{p_t}, g_{p_t}) up to ε.

By Lemma 1, for almost all $p \in H$, $X_\infty(p)(i)$ and $Y_\infty(p)(i)$ are positive.

Let η be a small positive number, to be specified later. By Lemma 2, for almost all $p \in H$ there is $t(p,\eta)$ satisfying (10").

For a fixed $t > t(p,\eta)$ define (f',g') as follows. For any history $h \in H_r$ if $p_t h$ is a prefix of a point in W^t_η let $f'(h) = f(p_t h)$ and $g'(h) = g(p_t h)$. Otherwise, define $f'(h) = \sum_{j=1}^{2} Y_{t+r}(p_t h)(j) f_j(p_t h)$ and $g'(h) = \sum_{j=1}^{2} X_{t+r}(p_t h)(j) g_j(p_t h)$. In words, f' is defined on "good" histories as the induced strategy of f, and on "bad" ones f' is defined to be exactly the expected strategy (from player two's point of view) player one is about to play.

Since (f',g') plays along $p \in W^t_\eta$ exactly as (f,g) does (which is the same as (f', g_{p_t}) and (f_{p_t}, g') play), (f',g') and (f_{p_t}, g_{p_t}) induce the same probability distribution on $V^t_\eta = \{p' \mid p_t p' \in W^t_\eta\}$. In other words, (f',g') and (f,g) induce measures that differ only on $H \backslash V^t_\eta$. Moreover, (f',g') and (f,g) assign to V^t_η the same probability. Since t satisfies (10"), (f',g') and (f_{p_t}, g_{p_t}) play the same up to 2η. It remains to show that (f',g') is an ε-equilibrium. We will show that f' is an ε-best response against g'. A similar argument would work for g'.

For p_t define an auxiliary strategy $\bar{g}$ of player two as follows:

$$\bar{g}(h) = \sum_{j=1}^{2} X_{t+r}(p_t h)(j) g_j(p_t h) \text{ for all } h \in H_r.$$

I.e., $\bar{g}$ is the strategy player 1 expects player two to play after the history h.

Since f,g are best responses to the respective beliefs that contain a

grain of truth, the induced strategy f_{p_t} is a best response against $\bar{g}$ in the repeated game.

That is, for any strategy k of player 1:

$$U_1(k,\bar{g}) \leq U_1(f_{p_t},\bar{g}). \tag{12}$$

By the definition of g', $g'(h) = g(h)$ whenever $p_t h$ is <u>not</u> a prefix of a path in W^i_η. Suppose $p_t h$ is a prefix of a path in W^t_η. If $X_\infty(p)(2) > 0$ and if η is smaller than $\mathrm{Min}_{1\leq i\leq 2}\{X_\infty(p)(i)\}/2$, then Lemma 3 implies

$$\|\bar{g}(h) - g'(h)\|_\infty < \delta(\eta) + \eta. \tag{13}$$

In words, the mixed action played according to $\bar{g}$ differs from that played by g' by at most $\delta(\eta) + \eta$. On the other hand, if $X_\infty(p)(2) = 0$ then $X_t(p)(2) < \eta$, which implies (13) in this case. Since U_1 is continuous one can find $\delta(\eta) + \eta$ small enough that

$$|U_1(k,\bar{g}) - U_1(k,g')| < \varepsilon/3 \text{ for all } k. \tag{14}$$

(13) and (14) imply

$$U_1(k,g') - \varepsilon/3 \leq U_1(k,\bar{g}) \leq$$
$$\leq U_1(f_{p_t},\bar{g}) \leq U_1(f_{p_t},g') + \varepsilon/3.$$

Thus,

(15) $$U_1(k,g') \leq U_1(f_{p_t},g') + 2\varepsilon/3.$$

(f_{p_t},g') induce the same probability distribution on V_η^t as (f',g') do. Moreover, they assign the same probability to V_η^t, which is, by (10"), at least $1 - 2\eta$.

Hence,

(16) $$U_1(f_{p_t},g') \leq U_1(f',g') + 2\eta.$$

Combine (15) and (16) to obtain

(17) $$U_1(k,g') \leq U_1(f',g') + 2\eta + 2\varepsilon/3 \text{ for all } k.$$

If η is chosen so that $\eta < \varepsilon/6$, (15) and the fact that (f',g') and (f_{p_t},g_{p_t}) play the same up to 2η conclude the proof of Theorem 2. //

6. Remarks

In this section we include some remarks about the assumptions of the model, the scope of the results, and possible extensions.

6.1 Is Having a Grain of Truth Necessary for Learning?

Without the assumption that the players' beliefs contain a grain of truth regarding the opponent's strategy, the learning described by Theorems 1 and 2 will not hold. To illustrate this point we present an example which addresses the learning aspect alone. Consider a repeated game with player one having to choose between ℓ and r in every stage. Suppose player one

chose the constant pure strategy L of always playing ℓ. Now, let us assume that player two believes that player one chose his pure strategy according to the following procedure: after every history player one randomized and chose ℓ with probability λ and r with probability $1 - \lambda$. This procedure induces for player two a belief probability distribution over the set of pure strategies of player one which assigns zero probability to the strategy L actually chosen by player one. Thus, player two's beliefs do not contain even a grain of truth.

It is easy to see that player two cannot learn anything about the future actions of player one. This is due to the fact that player one's future choices are assumed to be independent of his past choices--a situation that prohibits learning. Thus, the only hope of player two in predicting player one's future actions correctly would have been if player one's future actions were the same under every possible realization, which is not the case here.

The discussion above shows that, without the grain of truth assumption, our results may not hold. On the other hand, we know that weaker assumptions than the full grain of truth could suffice for approximate learning. Assume, for example, that player one plays a constant <u>behavior</u> strategy by which he randomizes with probability λ on ℓ and $(1 - \lambda)$ on r after every history of the game. Player two knows that this is the type of player one's strategy but does not know which λ player one is using. He assumes that player one chose λ according to a uniform distribution on the interval $[0,1]$. Again, player two's beliefs do not contain a grain of truth about player one's strategy but it seems that after long enough play player two could approximate the true λ and have a fairly accurate prediction of

player one's future play.

6.2 The Necessity of Knowing Your Own Preferences

The assumption that each player knows his own preferences is crucial. For example, Blum and Easley (1990) show a repeated game with incomplete information where the players never converge to play an equilibrium of the complete information game. Thus, their results contradict our corollary. The difference lies in the fact that, in their example, players do not know their own payoff matrix.

6.3 Do the Players Know When They Have Learned?

Our main results says that the players learn to predict future play and to play Nash equilibrium. Let T be the time by which this learning took place. Does a player know at time T that he has learned? Does he know that his opponent has learned? If the answer to these questions are positive, then does a player know that his opponent knows that he himself has learned? In other words, the entire hierarchy of knowledge and common knowledge questions can be asked.

In general, there is no common knowledge about the time of the learning in our model unless additional assumptions are imposed--for example, if the beliefs of both players were common knowledge. With the current assumptions, it is true that the players know the time of their own learning and in general know no more than that.

6.4 Extensions

Stochastic Games

The results of this paper can be proven for the more general model of stochastic games (see Shapley (1953)). The informational assumptions required are that each player knows his own payoff matrices as well as the transition probabilities and the state realizations of the stochastic game. Both theorems and the coroliary hold with essentially the same proofs. The notations, however, become somewhat more complex with no significant gain to our understanding.

General Continuous Payoff Functions

The proof of Theorem 2 applies to discounted repeated games as well as for repeated games with payoff functions that are continuous with respect to the product topology.

References

Arrow, K. J. (1962), "The Economic Implications of Learning by Doing," Review of Economic Studies, 29, 155-73.

Aumann, R. J. and M. Maschler (1967), "Repeated Games with Incomplete Information: A Survey of Recent Results," in Mathematica, ST-116, Ch. III, pp. 287-403.

Aumann, R. J. (1981), "Survey of Repeated Games," in Essays in Game Theory and Mathematical Economics in Honor of Oskar Morgenstern, Bibliographisches Institut, Mannheim/Wien/Zurich, 11-42.

Aumann, R. J. (1964), "Mixed and Behaviour Strategies in Infinite Extensive Games," in M. Dresher, L. S. Shapley and A. W. Tucker (eds.), Advances in Game Theory, pp. 627-650, Annals of Mathematics Studies, 52, Princeton University Press.

Blum, L., M. Bray and D. Easley (1982), "Introduction to the Stability of Rational Expectations Equilibrium," Journal of Economic Theory, 26, 313-317.

Blum, L. and D. Easley (1990), "Bayesian Learning in Repeated Games: An Example of Non-Convergence to Nash Equilibrium," mimeo.

Brock, W., R. Marimon, J. Rust and T. Sargent (1988), "Informationally Decentralized Learning Algorithms for Finite-Player, Finite-Action Games of Incomplete Information," mimeo.

Canning, D. (1989), "Convergence to Equilibrium in a Sequence of Games with Learning," mimeo.

Crawford, V. (1988), "Learning and Mixed Strategy Equilibria in Evolutionary Games," University of California, San Diego.

Easley, D. and N. Kiefer (1986), "Controlling a Stochastic Process with Unknown Parameters," Econometrica, forthcoming.

Fudenberg, D. and D. Kreps (1988), "A Theory of Learning, Experimentation and Equilibrium in Games," mimeo.

Fudenberg, D. and D. Levine (1990), "Maintaining a Reputation When Strategies are Mixed," mimeo.

Grandmont, J. M. and G. Laroque (1990), "Economic Dynamics With Learning: Some Instability Examples," CEPREMAP.

Harsanyi, J. C. (1967), "Games of Incomplete Information Played by Bayesian Players, Part I," Management Science, 14, 159-182.

Hart, S. (1985), "Nonzero-sum Two-person Repeated Games with Incomplete Information," Mathematics of Operations Research, 10, 117-153.

Jordan, J. S. (1989), "Bayesian Learning in Normal Form Games," University of Minnesota, forthcoming in Games and Economic Behavior.

Jordan, J. S. (1990), "The Exponential Rate of Bayesian Learning in Repeated Games," University of Minnesota.

Kuhn, H. W. (1953), "Extensive Games and the Problem of Information," in H. W. Kuhn and A. W. Tucker (eds.), Contributions to the Theory of Games, Vol. II, pp. 193-216, Annals of Mathematics Studies, 28, Princeton University Press.

Linnart, P., R. Radner and A. Schotter (1989), "Behavior and Efficiency in the Sealed-bid Mechanism," New York University.

McCabe, K. A., S. J. Rasseti and V. C. Smith (1989), "Lakates and Experimental Economics," mimeo.

McLennan, A. (1987), "Incomplete Learning in Repeated Statistical Decision Problems," University of Minnesota.

Mertens, J. F., S. Sorin and S. Zamir (1990), Repeated Games, to be published.

Mertens, J. F. (1987), "Repeated Games," Proceedings of the International Congress of Mathematicians (Berkeley, 1986), 1528-1577, American Mathematical Society, 1987.

Milgrom, P. and J. Roberts, "Adaptive and Sophisticated Learning in Repeated Normal Form Games," mimeo.

Nash, J. F. (1950), "Equilibrium Points in n-person Games," Proceedings of the National Academy of Sciences USA, 36, pp. 48-49.

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

Selten, R. (1988), "Adaptive Learning in Two Person Games," University of Bonn discussion paper.

Shapley, L. S. (1953), "Stochastic Games," Proceedings of the National Academy of Sciences of the USA, 39, 1095-1100.

Shiryayev, A. N. (1984), Probability, Springer-Verlag.

Stanford, W. G. (1990), "Pre-Stable Strategies in Discounted Duopoly Games," forthcoming in Games and Economic Behavior.

Woodford, M. (1990), "Learning to Believe in Sunspots," Econometrica, 58, 277-307.

Appendix I

Behavior and Mixed Strategies

Two types of randomizations have been considered for strategies of extensive form games. One is described by behavior strategies, where a player randomizes over his choice of actions at each one of his information sets. The other is described by mixed strategies, where a player randomizes initially over his choice of a pure strategy (a strategy for the whole game).

Since a pure strategy can be thought of as a special case of behavior strategy (restricted to probabilities 0 or 1 in choosing actions) extending randomizations over pure strategies to ones over behavior strategies results in a larger class of mixed strategies. This larger class obviously contains all behavior strategies and thus we have two classes of strategies: (a) the behavior strategies and (b) the larger class of mixed strategies which allow for randomizations over the choice of a behavior strategy.

The distinction between the two classes of randomizing strategies could have created difficulties in modeling extensive form games. Fortunately, Kuhn's theorem (see Kuhn (1953), Aumann (1964), and Selten (1975)) shows that no such difficulties should arise for games with "perfect recall" (all repeated games have perfect recall). It states that for such games the differences are only in presentation and that the class of mixed strategies is really not larger. More specifically, to every mixed strategy, randomizing over the choice of a behavior strategy, there are "equivalent" behavior strategies. This equivalence is strong in the sense that for any strategy of the opponent, the probabilistic structure over histories is

identical when the player uses a mixed strategy or an equivalent behavior strategy.

For the purposes of this paper, both representations of a strategy are important. We therefore elaborate here on the description of the behavior strategies equivalent to a given mixed strategy.

Consider, for example, the repeated prisoner's dilemma game where each player can choose the actions c or d at every stage of the game. We consider two simple behavior strategies, MC and MD. Under MC (mostly c), after every history of play, player one randomizes and chooses the action c with probability .90 and d with probability .10. Under MD he does the opposite. Suppose player one chooses a mixed strategy μ by randomizing 2/3 to 1/3 on the choice of C and D. For this μ there will be a unique equivalent behavior strategy f described as follows.

At the first stage f randomizes (2/3 × .90) + (1/3 × .10) on the action c and the compiement (2/3 × .10) + (1/3 × .90) on d. On subsequent stages f also randomizes between the actions c and d as prescribed by the strategies MC and MD. However, rather than using the prior distribution of (2/3, 1/3) over the application of MC or MD, it must use a Bayes' updated posterior about MC or MD given the history of the game. For example, suppose the action combination $\binom{c}{x}$ was observed in stage one (x could be either c or d here), then the Bayes' updated posterior distribution over the initial choices of MC or MD is given by (.90 × 2/3)/[(.90 × 2/3) + (.10 × 1/3)] ≈ .95 and 1 − .95 = .05. Thus, after a history $\binom{c}{x}$ f would randomize .95 × .90 + .05 × .10 = .86 on the action c and .95 × .10 + .05 × .90 = .14 on the action d. It is clear that this Bayesian updating can be done here after every history of play and the behavior strategy f, equivalent to the

mixed strategy μ, is uniquely determined.

Non-uniqueness may arise when the behavior strategies used in the mixture μ do not take all possible actions. For example, if a third action b was present in the repeated prisoner's dilemma game but MC and MD still put positive probabilities on c and d only. Now, after a zero probability history, for example $\binom{b}{x}$, the Bayesian updating described earlier is meaningless. A strategy f is equivalent to the mixed strategy μ if and only if it satisfies the randomizations according to Bayes' rule after every positive probability history. Thus, all strategies equivalent to a given mixed strategy coincide on the histories that are possible according to some of the strategies in the mixture. They differ only on their behavior after histories which are impossible. From the point of view of this paper, which deals only with Nash equilibrium, this non-uniqueness is inessential.

It is important to stress, for some future computations in the paper, that the following structure exists. Suppose f is a behavior strategy obtained by a mixed strategy μ, mixing a finite number of behavior strategies. Suppose also that h is a positive probability history according to f (and some strategy of the opponent). Then the strategy induced by f after h is obtained as the mixture of the strategies induced after h in the component of μ each weighted by its conditional probability given that the history h was played.