Rational Learning Leads to Nash Equilibrium

Ehud Kalai; Ehud Lehrer

Econometrica, Vol. 61, No. 5 (Sep., 1993), 1019-1045.

Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28199309%2961%3A5%3C1019%3ARLLTNE%3E2.0.CO%3B2-Z

Econometrica is currently published by The Econometric Society.

Econometrica, Vol. 61, No. 5 (September, 1993), 1019–1045

RATIONAL LEARNING LEADS TO NASH EQUILIBRIUM

BY EHUD KALAI AND EHUD LEHRER[1]

Each of n players, in an infinitely repeated game, starts with subjective beliefs about his opponents' strategies. If the individual beliefs are compatible with the true strategies chosen, then Bayesian updating will lead in the long run to accurate prediction of the future play of the game. It follows that individual players, who know their own payoff matrices and choose strategies to maximize their expected utility, must eventually play according to a Nash equilibrium of the repeated game. An immediate corollary is that, when playing a Harsanyi-Nash equilibrium of a repeated game of incomplete information about opponents' payoff matrices, players will eventually play a Nash equilibrium of the real game, as if they had complete information.

KEYWORDS: Repeated games, Nash equilibrium, rational learning, Bayesian learning, subjective equilibrium.

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 not fully understood. This is not surprising for games played only once where players have no observations to guide them and learning theories are restricted to model thought processes. However, in repeated interaction, where the players do have enough time to observe the behavior of their opponents, one can hope to obtain a statistical learning theory that leads to Nash equilibrium.

While some experimental work (see, for example, Smith (1990), McCabe et al. (1991), Linhart et al. (1989), Roth et al. (1991), and Prasnikar and Roth (1992)) supports the supposition that agents in repeated games do learn to play Nash equilibrium, no satisfactory theoretical explanation for this phenomenon exists. This is in spite of continuously growing game theoretic literature on repeated games with or without complete information (see Aumann (1981), and Mertens (1987), for surveys that are already outdated; see the forthcoming book by Mertens et al. (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., Blume et al. (1982), Jordan (1985), Easley and Kiefer (1988), Bray and Kreps (1987), McLennan (1987), Grandmont and Laroque (1990), Woodford (1990), and references therein).

[1]The authors wish to thank Robert Aumann, Larry Blume, David Easley, Itzhak Gilboa, Sergiu Hart, Martin Hellwig, James Jordan, Dov Monderer, Dov Samet, Sylvain Sorin, Vernon Smith, Robert Wilson, and anonymous referees for helpful discussions and suggestions. This research was partly supported by Grants Nos. SES-9011790 and SES-9022305 from the National Science Foundation, Economics, and by the Department of Economics and the C. V. Starr Center for Applied Economics of New York University.

The construction of processes that converge to Nash equilibrium is not a new topic to game theorists. Robinson (1951), Miyasawa (1961), and Shapley (1964) studied convergence properties of fictitious play. More recently, however, Fudenberg and Kreps (1988) recognized that such mechanisms can be used as a basis to model learning by players with bounded rationality, and a large literature on the subject has developed. A sample of such papers includes: Selten (1988), Crawford (1989), Canning (1992), Jordan (1991, 1992), Brock et al. (1988), Milgrom and Roberts (1991), Stanford (1991), and Fudenberg and Levine (1993a).

Most of this literature, however, builds on assumptions not applicable to our subject of repeated play among a small number of subjectively rational agents. The dynamic models studied in this literature are often ones of fictitious play or of random matching in a large population, and the behavior of the players is often modeled to be "myopic" or "bounded" in other ways. For our players, this implies shortcomings of the following types.

1. In trying to predict future opponents' behavior, a boundedly rational player ignores the fact that his opponents are also engaged in a dynamic learning process.

2. A myopic player would not perform a costly experiment no matter how high the resulting expected future payoffs are.

3. A myopic player ignores strategic considerations regarding the future. For example, even if he believes his opponent in a repeated prisoners' dilemma game is playing a trigger strategy (see Example 2.1), his consideration of immediate payoff may lead him to a choice of a long run inferior action.

In order to overcome these types of flaws, we take a traditional decision theoretic approach to the problem. We assume that the players are engaged in a standard perfect-monitoring infinitely repeated game with discounting. Thus, each one possesses a fixed matrix which specifies his payoff for every action combination taken by the group of players. In every period every player chooses his individual action, and the vector of chosen actions, through the individual payoff matrices, determines the period payoffs for all the players. Perfect monitoring means that before making the choice of a period's action, the player is informed of all the previous actions taken. Each player possesses a discount factor that he uses in evaluating future payoffs. His goal is to maximize the present value of his total expected payoff.

The players are assumed to be *subjectively rational* in the following sense. Each one starts with subjective beliefs about the individual strategies used by each of his opponents. He then uses these beliefs to compute his own optimal strategy. The strategies used for play, and for describing beliefs about an opponent's play, are behavior ones. Thus, they allow randomization in the choices of periods' actions (Section 3.4 elaborates on this topic and Kuhn's theorem, i.e., the fact that a probability distribution over many strategies can be replaced by a single behavior one). It is important to note that, unlike the stronger notion of rationalizable strategies (see Bernheim (1986) and Pearce (1984)), the knowledge assumptions implicit in the definition of a subjectively

rational strategy are weak. In order to choose one, a player must only know his own payoff matrix and discount parameter, but need not have any information about opponents' payoff matrices, nor assume anything about their rationality.

The main message of the paper is the following. If the players start with a vector of subjectively rational strategies, and if their individual subjective beliefs regarding opponents' strategies are "compatible with the truly chosen strategies," then they must converge in finite time to play according to an ε-Nash equilibrium of the repeated game, for arbitrarily small ε. Moreover, their Bayes-updated posterior subjective beliefs regarding future play of the game will become accurate with time. In other words, they will learn to predict the future play of the game and to play ε-Nash equilibrium for arbitrarily small ε. Some features and assumptions of the model should be emphasized.

1. The players' objective is to maximize, relative to their individual subjective beliefs, their long term expected discounted payoff. Learning is not a goal in itself here but is, rather, a consequence of overall individual payoff maximization plans. Learning is acquired as the real game progresses. 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 takes place through Bayesian updating of the individual 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). However, since the use of Bayesian updating is a consequence of expected utility maximization, assumption 2 is already a consequence of assumption 1.

3. We depart from the standard assumptions of game theory by not requiring that the players have full knowledge of each others' strategies, nor do they have commonly known prior distributions on the unknown parameters of the game. (We do not prohibit such assumptions but they are not necessary in our model.) Rather, we replace these assumptions by a weaker one of compatibility of beliefs with the truth. This assumption requires that players' subjective beliefs do not assign zero probability to events that can occur in the play of the game. In mathematical language, this means that on future play paths, the probability distribution induced by the chosen strategies must be absolutely continuous with respect to the probability distributions induced by the private beliefs of the players, i.e., any positive probability set of paths must be assigned some positive probability by each player. As an example, one may think of a situation where the beliefs about an opponent's strategy assign a small positive probability to the strategy actually chosen. In this case, we say that the beliefs contain a *grain of truth*, and compatibility of the beliefs with the truth is assured. Further discussion of these assumptions and their necessity will follow in subsequent sections.

An important corollary to the main result of this paper deals with Harsanyi-Nash equilibria of an n-person infinitely repeated game under discounting with n-sided incomplete information about opponents' payoff matrices. Assuming that the number of possible payoff matrices is finite or countable, the grain of truth condition stated above is satisfied. It follows that at such an equilibrium

the players will eventually play according to an ε-Nash equilibrium of the infinitely repeated realized game (the one with complete information) as if the uncertainties were not present. This corollary and its relation to Jordan's (1991) results will be discussed later in this paper.

As mentioned earlier, myopic theories of simultaneous learning involve fundamental difficulties due to the fact that what is being learned keeps changing. If players assume that their opponents' actions are fixed, yet the opponents, too, learn and change their own actions as a result of what they learn, inconsistencies are likely to occur. Indeed, as is shown, for example, by Kirman (1983) and by Nyarko (1991), learning may never converge or, worse yet, it may converge to false beliefs. Dynamic approaches, like the one taken in this paper, have the potential to overcome this difficulty. They attempt to learn the strategies (or, more precisely, the reaction rules that guide the opponents), of the *infinitely repeated game*. These strategies, which do not change, already contain the fixed learning rules.

Also, existing results of game theory suggest learning to play Nash equilibrium in a repeated game should be easier than in a one shot game. Considering the extreme case with completely patient players, i.e., discount factor equals one, the folk theorem tells us that all feasible individually rational payoffs are Nash payoffs. This suggests that many, yet certainly not all, play paths are Nash paths. Thus, our result regarding convergence to Nash paths seems to be more meaningful for moderate or low discount parameters.

A second difficulty, associated with learning models, concerns experimentation. In order to avoid getting stuck at suboptimal solutions, a well designed process should occasionally try randomly generated experimentation. For example, the randomly generated mutants in the evolutionary models play such a role. However, as can be seen in Fudenberg and Kreps (1988), in a rational choice model the optimal determination of when and how to experiment is difficult. The subjectively rational approach suggested here overcomes this difficulty: every action in the current model, including experimentation, is evaluated according to its long run contribution to expected utility. And maximization, relative to posterior probability distribution regarding opponents' strategies, yields well defined criterion for determining choices. Thus, a player, with a given discount parameter, will experiment when he assesses that the information gained will contribute positively to the present value of his expected payoff. Given the strategic nature of the interaction in our model, for some subjective beliefs regarding opponents' strategies, a player may find it in his interest to choose randomly when and how to experiment. While, in general, optimal experimentation computed against subjective beliefs does not lead to an individually optimal solution (see the multi-arm bandit example in the last section), it does so in the current paper due to the special assumptions of perfect monitoring and knowledge of own payoff matrices.

In addition to perfect monitoring, knowledge of own payoff matrices, and compatibility of beliefs with the truth, our model contains several additional restrictive assumptions. Independence of strategies is imposed in two places.

First, it is implicitly assumed that players' actual strategies are chosen independently. In other words, players' choices cannot depend on random events (unless they are completely private) since such dependencies may lead to correlated strategies, which are assumed away in the model. But this assumption of independence is also imposed on the subjective beliefs of the players. In his beliefs regarding opponents' strategies, a player assumes that the opponents choose their strategies independently of each other.

Also, the assumption that players maximize their expected payoffs is quite strong for infinitely repeated games. While this assumption is common in game theory and economics, the solution of such a maximization problem, in infinitely repeated games, may be very demanding.

Our preliminary studies regarding the relaxation of the various assumptions above, with the exception of truth-compatible beliefs, indicate that bounded learning will lead the players to correlated equilibrium (see Kalai and Lehrer (1992)) rather than Nash equilibrium. The need and possibilities of relaxing the truth compatibility assumption are discussed in several of the next sections.

Our proof of the convergence to playing an ε-Nash equilibrium is divided into three steps. The first establishes a general self-correcting property of Bayesian updating. This is a modified version of the seminal Blackwell and Dubins' (1962) result about merging of opinions. We give an independent easy proof of their result and an alternative characterization of their notion of merging.

When applied to our model, the self-correcting property shows that the probability distributions describing the players' beliefs about the future play of the game must converge to the true distribution. In other words, the beliefs and the real play become realization equivalent. At such time all learning possibilities have been exhausted. Remaining disagreement of the beliefs and the truth may only exist off the play path, and therefore will never be observed. We refer to such a situation of no further learning as subjective equilibrium. The notion of such an equilibrium and the fact that it may yield play different from Nash equilibrium were observed earlier in models of dynamic optimization, e.g., the multi-arm bandit literature (see, for example, Rothchild (1974)), and in a repeated game set-up by Fudenberg and Kreps (1988). Also, in a different learning model developed independently of ours, Fudenberg and Levine (1993b) developed and studied a closely related notion called self-confirming equilibrium. We refer the reader to Battigalli et al. (1992) for a survey of the history of this concept.

The last step of the proof shows that in our model, the behavior induced by a subjective equilibrium, and even its perturbed versions, approximates the behavior of an ε-Nash equilibrium. Since this last step has independent interest of its own, and since proving it involves long computations not related to learning, we leave it to a companion paper (see Kalai and Lehrer (1993a)).

Section 2 of this paper contains examples and additional elaborations on the approach taken and the assumptions made. The reader can skip it and move directly to Sections 3 and 4 which contain the formal presentation of the model and of the main results. Section 5 is devoted to the self-correcting property of

false priors by means of Bayesian updating. Section 6 contains applications to repeated games with incomplete information and the relation to Jordan's (1991) results. Finally, in Section 7, we give further elaborations on some of the assumptions and possible extensions.

2. EXAMPLES AND ELABORATION

In the two person games that follow, we will sometimes refer to player 1, PI, as he, and to player 2, PII, as she.

Example 2.1: *Infinitely Repeated Prisoners' Dilemma Games*

As usual for these games, we will denote the possible actions for each of the two players in each stage of the game by A—to describe aggressive behavior, and by C—to describe cooperative behavior. The following matrix represents the stage payoffs to PI as a function of pairs of action choices:

		PII: A	PII: C
PI	A	c	a
	C	d	b

As usual, we assume that $a > b > c > d$. PI uses a discount parameter λ_1 $(0 < \lambda_1 < 1)$ to evaluate infinite streams of payoffs. We use the convention that λ_1 close to 0 describes an impatient (myopic) player. PII has a similar payoff matrix and discount parameter but not necessarily with the same numerical values as the ones of PI. We assume here, as we do throughout this paper, that each player knows his own parameters, and that the game is played with perfect monitoring. That is, prior to making the choice in every stage, a player is informed of all the choices made by both players in all previous stages.

Departing from the traditional game theoretic approach, we do not explicitly model a player's knowledge about the parameters (payoff matrices, discount parameters, etc.) of his opponent. Instead, a player starts with prior subjective beliefs, described by a probability distribution, over the strategies his opponent will use. We assume that a player uses any specific knowledge he has about his opponent in creating these beliefs.

To illustrate such beliefs we consider a countable set of (pure) strategies g_t for $t = 0, 1, 2, \ldots, \infty$, defined as follows. g_∞ is the well-known (grim) trigger strategy. This strategy prescribes cooperation initially and after fully cooperative histories, but "triggers" to the aggressive action after every history that contains any aggression by either of the two players. For $t < \infty$, g_t coincides with g_∞ at all histories shorter than t but prescribes the aggressive action A after all histories of length t or more. In other words, if not triggered earlier, g_t will prescribe unprovoked aggression starting from time t on. With this convention, g_0 is the constant aggressive strategy.

Suppose PI believes that PII is likely to cooperate by playing her grim trigger strategy; but he also believes there are positive probabilities that she will stop cooperating earlier for other reasons. More precisely, he will assign her strategies $g_0, g_1, \ldots, g_\infty$ probabilities $\beta = (\beta_0, \beta_1, \ldots, \beta_\infty)$ that sum to 1 and with each $\beta_t > 0$. Depending on his own parameters he chooses a best response strategy of the form g_{T_1} for some $T_1 = 0, 1, \ldots$ or ∞. PII holds similar beliefs, represented by a vector α, about PI's strategy, and chooses a strategy g_{T_2} as her best response. Now the game will really be played according to the two strategies (g_{T_1}, g_{T_2}).

It is easy to see that the beliefs are compatible with the chosen strategies. All positive probability events in the game, e.g., cooperation up to time $t < \min(T_1, T_2)$, aggression in all times exceeding the $\min(T_1, T_2)$, are assigned positive probability by the original beliefs of both players. Thus, the results described earlier must hold.

Indeed, learning to predict the future play must occur. If, for instance, $T_1 < T_2$, then from time $T_1 + 1$ on, PII Bayesian updated beliefs regarding PI's choice will assign probability 1 to his choice of g_{T_1} and she will predict correctly the future noncooperative play. PI, on the other hand, will never fully know her strategy since he would only know that $T_2 > T_1$. But he will still be able to infer the forthcoming noncooperative play. This should clarify the point that players do not learn the strategy of their opponent *off* the play path; they only learn to predict actual future play against the strategy they themselves use. Also notice that accuracy of the above predictions did not rely on T_1 and T_2 being optimal choices. It only relied on correct updating of the truth-compatible subjective beliefs.

A second point to emphasize is that players' beliefs will not necessarily coincide with the truth after a finite time; beliefs may only converge to the truth as time goes by without ever coinciding with it. Suppose, for example, that $T_1 = T_2 = \infty$. Now the only resulting play path is the totally cooperative one. After playing it for t periods, PI, for example, will infer that she did not choose $g_0, g_1, \ldots, g_t$ and his Bayesian updated belief will assign probabilities $(\beta_{t+1}, \ldots, \beta_\infty)/\sum_{t+1}^{\infty}\beta_r$ to her remaining strategies: $(g_{t+1}, \ldots, g_\infty)$. Since $\beta_\infty > 0$, after sufficiently long time, his posterior probability $\beta_\infty/\sum_{t+1}^{\infty}\beta_r$ will be arbitrarily close to one and he will be almost certain that she chose $T_2 = \infty$.

The second main result, regarding convergence to Nash equilibrium play, is also easily seen in this example. The only two play paths that can survive after a long play are those generated by Nash equilibrium (and hence also by ε-Nash equilibrium). The totally aggressive path results from Nash equilibria regardless of the parameters of the game. But if the discount parameters are "generous," then the totally cooperative play path can also be obtained at a Nash equilibrium. Notice, however, that the overall play of these subjectively rational players can be of a type not generated by any Nash equilibrium, or even ε-Nash equilibrium for small ε. For example, the path that is fully cooperative up to time 3 and not cooperative afterwards cannot be the outcome of any Nash, or small ε-Nash, equilibrium. But such a path will be generated by the players if

their subjective individual beliefs assign high probability to the opponent not cooperating at time 4. Nevertheless, as follows from our second main result, from a certain time on these players will follow an ε-Nash equilibrium path. Thus, if the true parameters of the game allow only the totally aggressive Nash equilibrium (and hence the only path compatible with arbitrarily small ε-Nash equilibrium is the totally aggressive one), then at least one of the T_i's must be finite, and eventually constant mutual aggression must emerge. If, on the other hand, the game's parameters permit trigger strategies as a Nash equilibrium, then it is also possible that both T_i's are infinite and the play path follows cooperation throughout.

It is easy to observe in this example that players who hold optimistic prior probabilities (high α_∞ and β_∞) will follow a cooperative path while pessimistic players must eventually follow a noncooperative path. Thus, in the case of multiple equilibria, initial prior beliefs determine the final choice.

Example 2.2: *Absolute Continuity and Grain of Truth Assumptions*

In Example 2.1, each player's private beliefs assigned a strictly positive probability to the strategy actually chosen by the opponent. This condition, that beliefs regarding opponent's strategies contain a grain of truth, is stronger than needed.

Suppose, for instance, that in Example 2.1, PII's beliefs, given by the vector $\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_\infty)$, had sufficiently low values of all α_t's with $t < \infty$ to allow the trigger strategy g_∞ as her best response. Then the well-known tit-for-tat (tft) strategy (where she starts by cooperation and then proceeds to mimic her opponent's last move) can also be chosen as a best response. If she actually chooses tft as her strategy, then PI's beliefs about her strategy do not contain a grain of truth, given that his beliefs assign probability zero to nontrigger strategies. Yet his beliefs regarding future play paths will contain a grain of truth. Consider the play paths $z_0, z_1, \ldots, z_\infty$ with z_t describing the path in which both players cooperate up to time t, he cooperates and she aggresses at time t, and both aggress from time $t+1$. If PI's beliefs about PII's strategy are described by the vector β as in Example 2.1, and in response he chooses g_∞ for himself, then his induced beliefs on the future play paths are given by a distribution $\tilde{\mu}_1$ which assigns probability β_t to each of the paths z_t. Given both of their choices, the true distribution on future play paths, μ, will assign probability one to the path z_∞. We now can write $\tilde{\mu}_1 = \varepsilon\mu + (1-\varepsilon)\hat{\mu}$ for some probability distribution $\hat{\mu}$ and with $\varepsilon = \beta_\infty > 0$. When this is the case, i.e., when the belief distribution on future play paths assigns positive weight to the true distribution, we say that *the beliefs on play paths contain a grain of truth*. This last condition, which is weaker than the belief on strategies containing a grain of truth, is also sufficient for our main result.

The sufficient condition we end up using is weaker yet. We require that each player's belief distribution on play paths, $\tilde{\mu}_i$, not rule out positive probability events according to the real probability distribution, μ. That is, there should be

no event in the play of the infinite game which can occur, i.e., has a μ positive probability, yet be ruled out by the beliefs of an individual player, i.e., has a zero probability according to $\tilde{\mu}_i$. In mathematical language, we require that μ be *absolutely continuous* with respect to each $\tilde{\mu}_i$ ($\mu \ll \tilde{\mu}_i$).

To understand the difference between the grain of truth conditions and absolute continuity it is useful to consider behavior and mixed strategies, i.e., ones that allow for randomization in the choice of actions and strategies (the next section contains a more detailed discussion of Kuhn's theorem and these notions). Suppose PI's beliefs about PII's strategy are as in Example 2.1 with $\beta_t = (1/3)^{t+1}$ for $t < \infty$ and $\beta_\infty = 1/2$. Suppose that his choice in response to these beliefs is to play g_∞. His induced beliefs on the future play paths, $\tilde{\mu}_1$, assign probability $(1/3)^{t+1}$ to each of the paths z_t with $t < \infty$ and $1/2$ to z_∞. If, unlike his beliefs, she chooses to randomize over the choices of g_t with probabilities $(1/2)^{t+1}$ (zero probability on g_∞), then the real distribution on future play paths, μ, assigns probability $(1/2)^{t+1}$ to each of the paths z_t and zero to z_∞. It is easy to check that $\tilde{\mu}_1$ cannot be written as $\varepsilon\mu + (1-\varepsilon)\hat{\mu}$ with a positive ε for any probability distribution $\hat{\mu}$. Thus, even his beliefs about play paths do not contain a grain of truth. Yet the absolute continuity condition holds. Every event in the play of the game that has μ-positive probability, i.e., contains some paths z_t with finite t, is assigned a positive probability by the belief distribution $\tilde{\mu}_1$. Thus, the results of this paper regarding learning and convergence to Nash equilibrium must hold.

In the above example, however, if PI assigned probability one to her playing g_∞, yet she randomized on g_t with probability $(1/2)^{t+1}$ and probability zero on g_∞, then the absolute continuity condition would fail. And, indeed, learning and convergence to Nash equilibrium would fail, too.

Example 2.3: *On the Limitation of the Absolute Continuity Assumption*

Consider a repeated game with PI having to choose between l and r in every stage. Suppose PII believes that PI flips a coin to choose l with probability $1/2$ and r with probability $1/2$ after every history. This means that PII believes that future plays of PI are independent of his past actions and learning from the past is hopeless. Even if PI played the constant strategy L, always playing l, PII will not learn it since, given her initial beliefs, she will always dismiss long chains of l's as random outcomes. Notice that this is a situation where the absolute continuity assumption is violated. The event "l will be played forever" has probability one but is assigned probability zero in the beliefs of PII.

The discussion above shows that, without the absolute continuity assumption, or other assumptions that connect the future to the past, learning in general cannot take place. We know, however, that weaker assumptions suffice for approximate learning and for convergence of measures in a weak sense. Assume, for example, that player one plays a constant *behavior* strategy by which he randomizes with probability λ on l and $(1-\lambda)$ on r after every history of the

game. PII knows that this is the type of strategy PI uses but does not know the value of λ. She assumes that PI chose λ according to a uniform distribution on the interval $[0,1]$. Now, PII's beliefs do not satisfy the absolute continuity assumption, which can be seen by considering the event that the long run average of l's is λ (it has probability one but is assigned probability zero by the diffused beliefs of PII). However, after long enough play PII will be able to approximate the true λ and have a fairly accurate prediction of PI's near future play. Section 7 contains discussion on weak learning and the possibility of weakening the absolute continuity assumption.

Example 2.4: *Learning and Teaching*

While this paper presents a theory of learning, it does not put the players in a passive state of learning. The following example shows that optimizers, who believe their opponents are open to learning, may find it in their own interest to act as teachers.

We consider a two person infinite symmetric version of a "*chicken game*" described as follows. Simultaneously, at the beginning and with perfect monitoring after every history, each player chooses to "yield" (Y) or "insist" (I). However, once a player yields (chooses Y) he has to continue yielding forever. The stage game payoffs are the following:

		PII: Y	PII: I
PI	Y	0,0	1,2
	I	2,1	−1,−1

Infinite payoff streams are evaluated with discounting. We denote the individual pure strategies of this game by $s_0, s_1, \ldots, s_\infty$ with s_t indicating the one that prescribes the initial yielding at time t.

Notice that this game differs from the prisoners' dilemma example in some important ways. First, the stage game has no dominant strategies and it contains two symmetric pure strategy equilibria. Also, the repeated game contains exactly two pure strategy Nash equilibria, the one where he yields immediately and she insists forever (s_0, s_∞) and the symmetrically reversed one (s_∞, s_0). (In addition, subgames following mutual simultaneous yield actions contain the mutual yield forever equilibrium of these subgames.) While technically this game is not an infinitely repeated one, due to the absorbing nature of the action Y, the results of this paper still hold and offer interesting insights.

We assume as in the prisoners' dilemma example that PI's beliefs, about PII's strategy, are given by a vector $\beta = (\beta_0, \beta_1, \ldots, \beta_\infty)$ and, also, that PII's beliefs about PI are given by a similar vector, α. Putting himself partially in her shoes, PI may think that she is equally likely to wait any number of the first n periods before yielding, to see if he would yield first; or that she may insist forever with

probability ε because, unlike him, she assigns a very large loss to ever yielding. Such thinking will lead him to a prior beliefs vector of the type $\beta = ((1-\varepsilon)/n, \ldots, (1-\varepsilon)/n, 0, 0, \ldots, \varepsilon)$. If the future is important enough to PI, his best response to β would be to wait n periods in case she yields first, but if she does not, then yield himself at time $n+1$. Interpreted according to the thought process that led him to the choice of vector β, he reasons that as long as she is willing to find out about him, he will try to convince her by his actions that he is indeed tough.

If both players adopt such reasoning, a pair of strategies (s_{T_1}, s_{T_2}) will be chosen. In cases of the type $T_1 = 0$ and $T_2 > 0$, there was no attempt to teach on the part of PI and the resulting play is as in some Nash equilibrium of the infinite game. But in cases of the form $0 < T_1 < T_2$, PI failed in his attempt to teach her. The resulting play paths, with real initial fighting segments and continuing with his yielding, could not be justified by any Nash equilibrium or ε-Nash equilibrium with small ε. Similarly, when $T_1 > T_2 > 0$, we obtain a non-Nash equilibrium path with him winning. In any of the cases, however, as the main results of this paper state, both players will learn to predict the future play and end up playing a Nash equilibrium in sufficiently late subgames.

3. THE MODEL AND ASSUMPTIONS

3.1. *The Repeated Game*

A group of n players are about to play an infinitely repeated game. The stage game is described by the following components.

1. n finite sets $\Sigma_1, \Sigma_2, \ldots, \Sigma_n$ of *actions* with $\Sigma = \times_{i=1}^{n} \Sigma_i$ denoting the set of *action combinations*.

2. *n payoff functions* u_i: $\Sigma \to \mathbb{R}$.

We let H_t denote the *set of histories of length* t, $t = 0, 1, 2, \ldots$ (i.e., $H_t = \Sigma^t$, with Σ^0 being a singleton consisting of the null history). Denote by $\bar{H} = \bigcup_t H_t$ the set of all (*finite*) *histories*. A (behavior) *strategy* of player i is a function f_i: $\bar{H} \to \Delta(\Sigma_i)$ 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.

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.

3.2. *Infinite Play Paths*

Let $f = (f_1, \ldots, f_n)$ be a vector of behavior strategies. At the first stage player i plays $f_i(h^0)$, where h^0 stands for the null history. Notice that $f_i(h^0)$ is a probability distribution over his set of actions. Denote by $z_i^1 (= z_i^1(f_i))$, the realization of $f_i(h^0)$ and by z^1 the realized action combination, i.e., $z^1 = (z_1^1, \ldots, z_n^1)$. Player i is paid $x_i^1 = u_i(z^1)$ and receives the datum z^1 (he is

informed of the realized action combination). At the second stage player i randomizes over his actions according to $f_i(z^1)$. Denote by z_i^2 and by z^2 the realized action of player i at the second stage and the realized action combination, respectively. The payoff of player i is $x_i^2 = u_i(z^2)$ and he is informed of z^2. The game proceeds in this fashion infinitely many times. The infinite vector $(z^1, z^2, \dots)$ of action combinations is the *realized play path*.

The procedure described above defines a probability distribution, μ_f, induced by the strategy vector f, on the set of infinite play paths. First (with some abuse of notation), μ_f is defined inductively for finite histories $h \in \bar{H}$. μ_f of the empty history is 1 and $\mu_f(ha) = \mu_f(h) \times_i f_i(h)(a_i)$. In other words, the probability of the history h followed by an action vector a being played is the probability of h times the product of the a_i's being selected by the individual $f_i(h)$'s.

In the set of *infinite play paths*, Σ^∞, the event history h being played is described by the cylinder set $C(h)$, consisting of all paths with initial segment h. Thus f induces a probability $\mu_f(C(h))$ (the *probability of the history* h) to all such cylinder sets. Following the standard construction of probability theory, we let $\mathcal{F}_t$ denote the σ-algebra generated by the cylinder sets of the histories of length t, and $\mathcal{F}$, the σ-algebra used for Σ^∞, is the smallest one containing all $\mathcal{F}_t$'s. The probability distribution μ_f, defined on $(\Sigma^\infty, \mathcal{F})$, is the unique extension of μ_f from the $\mathcal{F}_t$'s to $\mathcal{F}$.

3.3. *The Payoffs*

Let λ_i, $0 < \lambda_i < 1$, be the discount factor of player i. Recall that x_i^t denotes player i's payoff at stage t. If the strategy vector f is played, then the payoff of player i in the repeated game is defined by

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

where E_f denotes the expected value calculated with respect to the probability measure, μ_f, induced by $f = (f_1, \dots, f_n)$.

Notice that $U_i(f)$ can be written also as $(1 - \lambda_i) \int [\Sigma x_i^{t+1} \lambda_i^t] \, d\mu_f$.

3.4. *Behavior and Beliefs*

In order to play the infinite game, each player i chooses a behavior strategy f_i. In addition, player i has a joint strategy $f^i = (f_1^i, f_2^i, \dots, f_n^i)$ describing his beliefs about the strategies adopted by his opponents. Thus, f_j^i denotes the behavior strategy that player i thinks player j will follow. We will assume throughout this paper that players know their own choice of strategies, i.e., $f_i^i = f_i$.

As usual, we say that a strategy of player i, f_i, is a *best response* to $f_{-i}^i = (f_1^i, \dots, f_{i-1}^i, f_{i+1}^i, \dots, f_n^i)$ if $U_i(f_{-i}^i, \bar{f}_i) - U_i(f_{-i}^i, f_i) \leqslant 0$ for all strategies $\bar{f}_i$ of player i. We say that f_i is an ε-best response ($\varepsilon \geqslant 0$) if the same inequality holds with ε replacing 0.

Suppose that f and g are two vectors of individual behavior strategies in the repeated game, with μ_f and μ_g denoting the distributions over infinite play paths induced by f and g, respectively. The measure μ_f is said to be *absolutely continuous* with respect to (w.r.t.) μ_g (denoted by $\mu_f \ll \mu_g$) if every event having a positive measure according to μ_f also has a positive measure according to μ_g. Formally, $\mu_f(A) > 0$ implies $\mu_g(A) > 0$ for every measurable set $A \subseteq \Sigma^\infty$. If $\mu_f \ll \mu_g$ we also say that f is *absolutely continuous* w.r.t. g.

It is important to expand on the assumption that the beliefs player i holds regarding player j's strategy are described by a single behavior strategy f_j^i. This represents no serious restriction because the well-known Kuhn's (1953) theorem (see also Selten (1975)) assures us that if player i's beliefs were given by a probability distribution over behavior strategies of player j, then these beliefs could be replaced by an equivalent single behavior strategy. Since beliefs are a central topic of this paper, and since Kuhn's equivalent behavior strategies use in their construction Bayesian updating, another central topic to this paper, we briefly review this construction.

Suppose player i believes that player j will play the behavior strategy $f_{j,r}$ with probability λ_r, $r = 1, \ldots, l$. A Kuhn's equivalent behavior strategy f_j^i will choose the action a after the history h with probability

$$f_j^i(h)(a) = \sum (\lambda_r | h) f_{j,r}(h)(a)$$

with $\lambda_r | h$ being the posterior probability of $f_{j,r}$ having been chosen given the observed history h, i.e.,

$$\lambda_r | h = \lambda_r \eta_{f_{j,r}}(h) \Big/ \sum_w \lambda_w \eta_{f_{j,w}}(h),$$

where $\eta_{f_{j,r}}(h)$ denotes the probability of h being reached when all players other than j take the actions leading to h with probability one, and player j mixes according to $f_{j,r}$. (In the case that $\eta_{f_{j,w}}(h) = 0$ for $w = 1, \ldots, l$, $f_j^i(h)$ can be chosen arbitrarily.)

Kuhn's equivalence is strong. Playing against the strategies $(f_{j,r})_r$ with the probabilities $(\lambda_r)_r$ and playing against an equivalently constructed behavior strategy f_j^i generate identical probability distributions on the future play paths of the game (and hence also all positive probability subgames). We refer the reader to Aumann (1964) for the infinite version of Kuhn's theorem.

It is important to emphasize here an assumed restriction on the nature of the beliefs. Player i believes that different opponents, say, j and k, are described by individual strategies f_j^i and f_k^i. In evaluating the probabilities of potential histories, he uses the product of the probabilities induced by such strategies. In other words, he believes that players j and k choose their actions independently. This rules out important cases where player i believes that j and k's strategy choices are correlated—for example, they both depend on the same random event whose outcome he himself does not know (e.g., they both went to school A or both went to school B and their strategies depend on the school

they went to). Our results regarding future play prediction can be extended to these cases, but convergence to ε-Nash equilibrium may fail.

4. STATEMENT OF THE MAIN RESULTS

Recall that by a *path* we mean an infinite sequence of action combinations, i.e., an element of Σ^{∞}. For any path z and time $t \in \mathbb{N}$ we denote by $z(t)$ the t-prefix of z (the element in H_t consisting of the first t action combinations of z).

DEFINITION 1: Let $\varepsilon > 0$ and let μ and $\tilde{\mu}$ be two probability measures defined on the same space. We say that μ is *ε-close to $\tilde{\mu}$* if there is a measurable set Q satisfying:

(i) $\mu(Q)$ and $\tilde{\mu}(Q)$ are greater than $1-\varepsilon$; and

(ii) for every measurable set $A \subseteq Q$

$$(1-\varepsilon)\tilde{\mu}(A) \leqslant \mu(A) \leqslant (1+\varepsilon)\tilde{\mu}(A).$$

Notice that this notion of ε-closeness is strong. Unlike closeness measures that depend on differences (e.g., $|\mu(A)-\tilde{\mu}(A)| \leqslant \varepsilon$, where $\tilde{\mu}(A)$ can equal $2\mu(A)$ without violating the closeness requirement for small probability A), our definition requires that any two events in Q can only differ by a small percentage. It also implies closeness of conditional probabilities. If $A, B \subseteq Q$ then μ being ε-close to $\tilde{\mu}$ in the above sense implies that

$$\tilde{\mu}(A|B)(1-\varepsilon)/(1+\varepsilon) \leqslant \mu(A|B) \leqslant \tilde{\mu}(A|B)(1+\varepsilon)/(1-\varepsilon).$$

Thus, in the sequel where μ represents true probabilities of events in the game and $\tilde{\mu}$ represents beliefs of a player, being ε-close would mean that not only does the player assess the future correctly, he even assesses developments following small probability histories correctly, provided that he considers paths in the large set Q. This is important since it implies no cumulative buildup of an error in his assessment of the future no matter how far.

Being close in our sense, on a large set Q, and being close in the sense of differences, as mentioned above but without a restriction to a large set Q, turn out, however, to be asymptotically equivalent notions, as we discuss in Section 5.

Let f and g be two joint strategies.

DEFINITION 2: Let $\varepsilon \geqslant 0$. We say that *f plays ε-like g* if μ_f is ε-close to μ_g.

DEFINITION 3: Let f be a strategy, $t \in \mathbb{N}$ and $h \in H_t$. The *induced strategy* f_h is defined by

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

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'. If $f = (f_1, \ldots, f_n)$ is a joint strategy, f_h denotes the joint strategy consisting of all the induced individual strategies.

The following theorem states that if the vector of strategies actually chosen is absolutely continuous w.r.t. the beliefs of a player, then the player will learn to accurately predict the future play of the game.

THEOREM 1: *Let f and f^i be two n-vectors of strategies, representing the ones actually chosen and the beliefs of player i, respectively. Assume that f is absolutely continuous w.r.t. f^i. Then for every $\varepsilon > 0$ and for almost every play path z (according to the measure induced by f) there is a time T $(= T(z, \varepsilon))$ such that for all $t \geq T$, $f_{z(t)}$ plays ε-like $f^i_{z(t)}$.*

In other words, after the history $z(t)$, the real probability distribution over the future play of the game is ε-close to what player i believes the distribution is. It implies that the real probability of any future history cannot differ from the beliefs of player i by more than ε. But, as discussed earlier, it is substantially stronger. It implies closeness of probabilities for small events and for conditional probabilities.

Notice that, in Theorem 1, other than absolute continuity, no assumptions were made on f and f^i. Thus, it is applicable to any strategies of interest and not just to those maximizing expected utility. For instance, if a player were following a minmax strategy and still conducting a Bayesian update, he would also learn to predict the future play. The theorem essentially states that Bayesian updating by itself will lead to a correct prediction of the important parts (those that determine the actual play) of other players' strategies. It does not state that a player would learn to predict other players' future randomization in response to actions that will not be taken.

Theorem 1, by itself, has immediate implications for theories dealing with payoff maximizing players. Suppose, as Theorem 1 implies, that $f_{z(t)}$ plays ε-like $f^i_{z(t)}$ for all $i = 1, \ldots, n$. Furthermore, assume that f_i is a best response to f^i. Then, after sufficiently long time: (i) each player maximizes his payoff against his subjective beliefs and, moreover, (ii) these beliefs are almost (up to ε) realization equivalent to the real strategies played. Thus, each player is maximizing relative to (possibly false) subjective beliefs which will never be contradicted by the play of the game (even statistically). The following solution concept captures these two elements.

DEFINITION 4: An n-vector of strategies, g, is a *subjective ε-equilibrium* if there is a matrix of strategies $(g^i_j)_{1 \leq i, j \leq n}$ with $g^i_i = g_i$ such that

(i) g_i is a best response to g^i_{-i}, $i = 1, \ldots, n$; and

(ii) g plays ε-like g^i, $i = 1, \ldots, n$.

COROLLARY 1: *Let f and $f^1, f^2, \ldots, f^n$ be vectors of strategies representing the actual choice and the beliefs of the players. Suppose that for every player i:*

(i) *f_i is a best response to f^i_{-i}; and*

(ii) *f is absolutely continuous w.r.t. f^i.*

Then for every $\varepsilon > 0$ and for almost every (w.r.t. μ_f) path z there is a time T

($= T(z, \varepsilon)$) such that for all $t \geq T$ $f_{z(t)}$ with $f^1_{z(t)}, \ldots, f^n_{z(t)}$ is a subjective ε-equilibrium.

PROOF: The corollary follows immediately from Theorem 1 when we recognize that maximizing expected discounted utility implies maximizing expected utility after every positive probability history relative to the posterior distribution induced by the history.

Notice that if g is a subjective 0-equilibrium (or just *subjective equilibrium*), then μ_g, the distribution induced by g, is identical to μ_{g^i}. Thus, g and g^i are realization equivalent. Despite the equivalence, a subjective 0-equilibrium does not necessarily induce the same behavior as a Nash equilibrium (the one person multi-arm bandit game is a well-known example).

However, under the assumptions of knowing own payoff matrices and perfect monitoring, or "observed-deviators" in the language of Fudenberg and Levine (1993b), it is easy to see that identical behavior is induced (see also Battigalli et al. (1992), for earlier versions of this observation). Clearly, every Nash equilibrium, being a subjective equilibrium, induces a subjective equilibrium behavior. Conversely, starting with a subjective equilibrium, one can modify the strategies used as follows. After histories that are in the support of all players' strategies leave the actions of all players unchanged. In subgames that follow a one person deviation from his support, have all the players switch their actions to the ones attributed to them by the beliefs of the deviator. In subgames that follow a multiperson deviation assign the players any actions. It is easy to observe that this modification yields a Nash equilibrium which is realization equivalent to the original subjective equilibrium.

When perturbations are introduced to the accuracy of the beliefs in a subjective ε-equilibrium and to the accuracy of optimization in an ε-Nash equilibrium, the discrepancy between the two concepts is greater and the equivalence of behavior sketched above fails for obvious reasons (see Kalai and Lehrer (1993a) for discussion and elaborations). Yet, for the family of games studied here, the two notions induce asymptotically identical behavior.

PROPOSITION 1: *For every $\varepsilon > 0$ there is $\eta > 0$ such that if g is a subjective η-equilibrium then there exists $\bar{f}$ such that*

(i) *g plays ε-like $\bar{f}$, and*

(ii) *$\bar{f}$ is an ε-Nash equilibrium.*

Theorem 1 will be proven in the next section. We refer the reader to Kalai and Lehrer (1993a) for the proof of Proposition 1 and for a general discussion on subjective equilibrium. Together, however, Corollary 1 and Proposition 1 imply our main result.

THEOREM 2: *Let f and $f^1, f^2, \ldots, f^n$ be strategy vectors representing respectively the one actually played and the beliefs of the players. Suppose that for every*

player i:

(i) f_i *is a best response to* f^i_{-i}; *and*

(ii) f *is absolutely continuous with respect to* f^i.

Then for every $\varepsilon > 0$ *and for almost all* (*with respect to* μ_f) *play paths z there is a time* $T = T(z, \varepsilon)$ *such that for every* $t \geq T$ *there exists an* ε*-equilibrium* $\bar{f}$ *of the repeated game satisfying* $f_{z(t)}$ *plays* ε*-like* $\bar{f}$.

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 utility maximizing players start with individual subjective beliefs, with respect to which the true strategies are absolutely continuous, then in the long run, their behavior must be essentially the same as a behavior described by an ε-Nash equilibrium. In the last section of the paper, we show that by using a weaker version of closeness of behavior one can replace the ε-Nash equilibrium in Theorem 2 by the usual Nash equilibrium.

5. BAYESIAN LEARNING

Our main result, Theorem 2, combines two issues: (i) Bayesian updating and (ii) payoff maximization. In this section, we concentrate on the first one and prove Theorem 1. In fact, the treatment of Bayesian updating, given here, is applicable to issues that lie beyond the scope of this paper. The reader is referred to Kalai and Lehrer (1990a, b and 1993b) and Monderer and Samet (1990).

Suppose that $(\Omega, \mathscr{F})$ is a measure space interpreted as the set of states of the world. Let $\{\mathscr{P}_t\}_t$ be an increasing sequence of finite or countable partitions of Ω (i.e., $\mathscr{P}_{t+1}$ refines $\mathscr{P}_t$). $\mathscr{P}_t$ is interpreted as the information available at time t. In other words, at time t the agent is informed of the part $P_t(\omega) \in \mathscr{P}_t$ that contains the prevailing state $\omega \in \Omega$.

We assume that the σ-field $\mathscr{F}$ is the smallest one that contains all the elements of all the $\mathscr{P}_t$'s.

The agent's initial belief about the distribution of states of nature is denoted by $\tilde{\mu}$ (a σ-additive measure defined on $(\Omega, \mathscr{F})$). However, the real distribution is given by a measure μ. Our task in this section is to show that the subjective probability (the belief) converges to the real one as information increases.

Denote the field generated by $\mathscr{P}_n$ by $\mathscr{F}_n$. The next theorem is a restatement of Theorem 1 but in the language of partitions. It is essentially equivalent to the Blackwell and Dubins (1962) theorem discussed later.

THEOREM 3: *Let* $\mu \ll \tilde{\mu}$. *With* μ*-probability* 1, *for every* $\varepsilon > 0$ *there is a random time* $r(\varepsilon)$ *such that for all* $r \geq r(\varepsilon)$, $\mu(\cdot \mid P_r(\omega))$ *is* ε*-close to* $\tilde{\mu}(\cdot \mid P_r(\omega))$.

PROOF: Theorem 3 is a consequence of Proposition 2 and Lemma 1 (see also Monderer and Samet (1990)), which follow.

PROPOSITION 2: *Suppose that* $\mu \ll \tilde{\mu}$ (*i.e.*, $\mu(A) > 0$ *implies* $\tilde{\mu}(A) > 0$ *for every* $A \in \mathcal{F}$). *With* μ*-probability* 1 *for every* $\varepsilon > 0$ *there is a random variable* $t(\varepsilon)$ *such that for every* $s \geq t \geq t(\varepsilon)$:

$$(1) \qquad 1 - \varepsilon \leq \frac{\mu(P_s(\omega)|P_t(\omega))}{\tilde{\mu}(P_s(\omega)|P_t(\omega))} \leq 1 + \varepsilon.$$

PROOF: Since $\mu \ll \tilde{\mu}$, by the Radon-Nikodym theorem, there is an $\mathcal{F}$-measurable function ϕ satisfying

$$(2) \qquad \int_A \phi \, d\tilde{\mu} = \mu(A) \qquad \text{for every } A \in \mathcal{F}.$$

By Levy's theorem (see Shiryayev (1984)), $E_{\tilde{\mu}}(\phi|\mathcal{F}_t) \to E_{\tilde{\mu}}(\phi|\mathcal{F}) = \phi$ $\tilde{\mu}$ almost surely (and therefore, μ-a.s.). However, for $\tilde{\mu}$ almost all ω

$$(3) \qquad E_{\tilde{\mu}}(\phi|\mathcal{F}_t)(\omega) = \big(1/\tilde{\mu}(P_t(\omega))\big)\int_{P_t(\omega)} \phi \, d\tilde{\mu} = \mu(P_t(\omega))/\tilde{\mu}(P_t(\omega)).$$

Moreover, by (2), $\phi > 0$ μ-a.s. Thus, the right side of (3) tends μ-a.s. to a positive number. In other words, there is a $t(\varepsilon)$ such that for μ-a.e. ω the following holds:

$$(4) \qquad 1 - \varepsilon \leq \frac{\mu(P_t(\omega))}{\tilde{\mu}(P_t(\omega))} \Big/ \frac{\mu(P_s(\omega))}{\tilde{\mu}(P_s(\omega))} \leq 1 + \varepsilon \qquad \text{for all } s \geq t \geq t(\varepsilon).$$

The middle term of (4) is equal to the middle one in (1). Since (1) holds for every $\varepsilon > 0$ with μ-probability 1, the proposition follows. *Q.E.D.*

LEMMA 1: *Let* $\{W_t\}$ *be an increasing sequence of events satisfying* $\mu(W_t) \uparrow 1$. *For every* $\varepsilon > 0$ *there is a random time* $t(\varepsilon)$ *such that any random* $t \geq t(\varepsilon)$ *satisfies*

$$\mu\{\omega; \mu(W_t|P_t(\omega)) \geq 1 - \varepsilon\} = 1.$$

PROOF: $\mu(W_t) \to_{t \to \infty} 1$. Thus, $\mu(C_t) \to_{t \to \infty} 0$, where $C_t = \Omega \setminus W_t$.

Suppose, to the contrary, that the lemma does not hold. Then there is a μ-positive set A and $\varepsilon > 0$ such that for all $\omega \in A$, $\mu(W_t|P_t(\omega)) < 1 - \varepsilon$ for infinitely many t's.

Fix $s \in \mathbb{N}$ and define

$$B_r = \{\omega \in A; r = \min\{t | t \geq s \quad \text{and} \quad \mu(W_t|P_t(\omega)) < 1 - \varepsilon\}\}.$$

Observe that $\{B_r\}$ are pairwise disjoint and, moreover, $\{\bigcup_{\omega \in B_r} P_r(\omega)\}_r$ are also pairwise disjoint. By the definition, $A = \bigcup_{r \geq s} B_r$.

Since $C_s \supseteq C_t$ when $t \geq s$, for all $\omega \in A$, $\mu(C_s|P_t(\omega)) > \varepsilon$ for infinitely many t. Thus, $\mu(C_s|\bigcup_{\omega \in B_t} P_t(\omega)) > \varepsilon$. Therefore, $\mu(C_s) > \varepsilon\mu(\bigcup_{t \geq s} \bigcup_{\omega \in B_t} P_t(\omega)) \geq \varepsilon\mu(\bigcup_{t \geq s} B_t) = \varepsilon\mu(A)$.

Hence, the sequence $\{\mu(C_s)\}$ is bounded away from zero, which is a contradiction. This concludes the proof of the lemma. *Q.E.D.*

In order to apply the lemma set

$$W_t = \{\omega; |E(\phi|\mathcal{F}_s)(\omega)/E(\phi|\mathcal{F}_t)(\omega) - 1| < \varepsilon \quad \text{for} \quad \forall s \geq t\}.$$

An immediate corollary is a version of the main result of Blackwell and Dubins (1962).

COROLLARY 2 (see Blackwell and Dubins (1962)): *For μ-a.e. ω there is time $t = t(\varepsilon)$ such that for $A \in \mathcal{F}$ and $s \geq t$ $|\mu(A|P_s(\omega)) - \tilde{\mu}(A|P_s(\omega))| < \varepsilon$.*

The converse statement, that the Blackwell-Dubins' result implies Theorem 3, is also true but not obvious. One can actually show that the topology generated by our notion of closeness is equivalent to the one generated by Blackwell and Dubins. That is a sequence of measures $\mu_s \to \mu$ in one sense if and only if it does so in the other. That our topology is stronger is immediate. However, since the notion used by Blackwell and Dubins applies to all events, not just in large set Q, it turns out to be as strong. See Kalai and Lehrer (1993b) for details.

Example: One biased coin with parameter p_i is selected with probability $\alpha_i > 0$ from a countable set of such coins. The chosen coin is tossed infinitely many times. An agent believes that the coin p_i is drawn with probability β_i. Define Ω to be the set of infinite sequences of 0's and 1's generated by the tosses of the coin. The probability measure on Ω, induced by $\{\alpha_i\}$, say μ, is absolutely continuous with respect to the one induced by $\{\beta_i\}$, say $\tilde{\mu}$, if $\beta_i > 0$ for all i. Theorem 3 states that, after sufficiently long time, the posterior probability of μ will be arbitrarily close to the posterior one of $\tilde{\mu}$.

REMARK: For general probability measures, we say that $\tilde{\mu}$ contains a "grain of truth" of μ, if $\tilde{\mu} = \lambda\mu + (1-\lambda)\overline{\mu}$ for some probability measure $\overline{\mu}$ and $\lambda > 0$. It is equivalent to requiring that the Radon-Nikodym derivative, $\phi = d\mu/d\tilde{\mu}$, is bounded.

Notice that in the previous example $\tilde{\mu}$ contains a grain of truth if and only if α_i/β_i are uniformly bounded.

6. REPEATED GAMES WITH INCOMPLETE INFORMATION AND JORDAN'S RESULTS

In this paper uncertainties regarding other players are captured by the individual beliefs a player holds about others' strategies. This is unlike traditional game theory where uncertainties are expressed by a commonly known prior distribution over the unknown parameters of the game (payoffs, discount parameters, feasible actions, etc.) with a commonly known signaling mechanism that gives different players different additional information. We proceed to show by example how the traditional approach can be viewed as a special case of the current paper. In particular, the equilibria of a large class of repeated games with incomplete information will satisfy the assumptions of our main theorems, and the conclusions will yield interesting new insight.

Consider two players about to play an infinitely repeated game of the type described earlier, but with a randomly generated fixed size pair of payoff matrices $(A_i, B_j)_{(i,j)\in I\times J}$. We assume that both I and J are finite or countable, and that the selection of the pair (i, j) will be done according to a commonly known prior probability distribution π on $I\times J$. After the selection, PI will be told the realized value $\tilde{i}$ and PII will be told the realized value $\tilde{j}$.

In order to play the game, PI chooses a vector $(f_i)_{i\in I}$ with each f_i being the infinite game strategy that he would follow if he is told that his realized payoff matrix is A_i. PII chooses a similar vector of strategies $(g_j)_{j\in J}$. A pair of such vectors is a Harsanyi-Nash equilibrium if each f_i is a best response (in long term discounted utility) against the strategies $(g_j)_{j\in J}$ when mixed according to the conditional distribution on J given the realized value $i, \pi(j|i)$, and with the symmetric property holding for each g_j (see also Hart (1985)).

To relate such an equilibrium to the current paper, assume that the random drawing of the payoff matrices has been performed and that $\tilde{i}$ and $\tilde{j}$ were selected. Thus, the real play of the game will follow the pair of strategies $(f_{\tilde{i}}, g_{\tilde{j}})$. Given his information, PI believes that PII will play $(g_j)_{j\in J}$ with probabilities $\pi(j|\tilde{i})$ and being at a Nash equilibrium his $f_{\tilde{i}}$ is actually a best response to this belief. Moreover, given the finiteness of J, PI's belief contains a grain of truth (i.e., assigns positive probability to $g_{\tilde{j}}$). Similarly, PII's choice of $g_{\tilde{j}}$ is a best response to the distribution $\pi(i|\tilde{j})$ on $(f_i)_{i\in I}$ and it also contains a grain of truth.

In the set-up above, let $f_{\tilde{i}}$ and $g_{\tilde{j}}$ be the strategies realized and let $\tilde{g}$ and $\tilde{f}$ be the induced beliefs over opponent's strategies, e.g., $\tilde{g}$ is the behavior strategy obtained by mixing the vector $(g_j)_{j\in J}$ with probabilities $\pi(j|\tilde{i})$.

The analogies of Theorem 1, Proposition 1 and Theorem 2, when applied to the Harsanyi-Nash equilibrium, follow as immediate corollaries.

THEOREM 1.1: *For every $\varepsilon > 0$ and almost every play path z (relative to the distribution induced by $f_{\tilde{i}}, g_{\tilde{j}}$) there is a time T such that for all $t \geqslant T$ $(f_{\tilde{i}}, g_{\tilde{j}})_{z(t)}$ plays ε-like $(f_{\tilde{i}}, \tilde{g})_{z(t)}$.*

In other words, at such a Harsanyi-Nash equilibrium the players eventually predict the future play of the game accurately even if they do not necessarily learn the payoff matrices of their opponents.

THEOREM 2.1: *For every $\varepsilon > 0$ and almost every play path z we can find a time T such that for all $t \geqslant T$ there is an ε-Nash equilibrium of the realized repeated game $(A_{\tilde{i}}, B_{\tilde{j}})$, $(\hat{f}, \hat{g})$, with $(f_{\tilde{i}}, g_{\tilde{j}})_{z(t)}$ plays ε-like $(\hat{f}, \hat{g})$.*

So even if the players do not learn the identity of the payoff matrices actually played, they eventually play almost as ε-Nash players who do know the identity of the payoff matrices.

Theorem 2.1 is related to an earlier result of Jordan (1991). His players faced the same uncertainty about opponents' payoff matrices. His prior distribution about such matrices, however, was more general since he did not restrict himself to a discrete set of possible matrices. On the other hand, his players played myopically. In each period they played a Harsanyi-Nash equilibrium, updated on all the information obtained earlier, as if each current period were the last one. He then studied the limit beliefs about opponents' next period actions as the number of periods became large. His main result was that all cluster points of the expectation sequence are Nash equilibria of the underlying realized stage game.

Our model can be made nearly myopic by letting the discount parameter approach zero. In general, when one totally discounts the future, Nash equilibria of the repeated game consist of repeated plays of Nash equilibria of the stage game. Thus, as a limit case when we let the discounted parameters approach zero, our result regarding convergence to Nash equilibria of the repeated game confirms Jordan's result of convergence to Nash equilibria of the stage game. (Of course, if the stage game had a multiplicity of equilibria, then one could see oscillation among them.) Notice, though, that Jordan obtains convergence of the beliefs to Nash equilibrium, while we obtain convergence of the beliefs and the actual play to ε-Nash equilibrium.

When considering generalizations of Theorems 1.1 and 2.1 above to the n-player case with $n > 2$ we observe the following. Theorem 1.1 generalizes. Theorem 2.1 does not unless we impose an additional independence assumption on the prior distribution over payoff matrices. The condition needed is that the prior distribution, $\pi(i_1, i_2, \ldots, i_n)$, over payoff matrices should be independent over opponents for every realization of every player i_j. For example, for player 1, $\pi(i_2, \ldots, i_n | i_1)$ should be independent over i_2 through i_n.

The need for the above independence condition arises in the application of Proposition 1. As assumed in the definition of a subjective ε-equilibrium, each player assigns independent beliefs to the strategies of his opponents. However, if the prior distribution π did not satisfy the independence condition, one would not be able to replace the mixed combination of opponent's strategies by an equivalent product of behavior strategies, so the proposition would not hold. Indeed, convergence to ε-Nash equilibrium will fail.

In order to correct for such dependencies, we would have to generalize the concept of subjective ε-equilibrium to allow for correlated beliefs. This will yield a concept closer to the notion of self-confirming equilibrium developed by Fudenberg and Kreps (1988) and Fudenberg and Levine (1993b). The new concept will have to be defined for infinite games and will have to include a suitable notion of perturbation. The convergence of Theorem 2.1 to ε-Nash equilibrium is then likely to be replaced by convergence to a correlated ε-equilibrium.

7. REMARKS

This section includes some additional remarks about assumptions of the model, the scope of the results, and possible extensions.

7.1. *An Alternative Notion of Closeness*

As discussed in this paper, the notions of one measure being close to another and of one strategy vector playing like another are strong. They guarantee that if f plays ε-like g, then with probability $1-\varepsilon$, f and g will assign close probabilities to future events and will continue to do so regardless of how long the game has been played. The result about learning to play ε-like ε-Nash gives a strong notion of proximity to the ε-Nash for the rest of the infinite game. We do not know, at this time, if the same theorem can be proven with only one ε, i.e., learning to play ε-like a full Nash equilibrium. However, with a less demanding notion of being close, the players will learn to play a full, rather than ε, Nash equilibrium.

For an $\varepsilon > 0$ and a positive integer l we say that μ is (ε, l)-close to $\tilde{\mu}$ if for every history h of length l or less $|\mu(h) - \tilde{\mu}(h)| \leqslant \varepsilon$. Similarly, f plays (ε, l)-like g if the induced measure μ_f is (ε, l)-close to μ_g. Thus, playing (ε, l)-like means playing ε the same up to a horizon of l periods. It was shown in Kalai and Lehrer (1993a) that for a given ε and l, if g is a subjective η-equilibrium, with sufficiently small η, then it must play (ε, l)-like some Nash (rather than ε-Nash) equilibrium of the repeated game. So, taking this less ambitious notion of approximating behavior, we can obtain ε-closeness in finite time to a full Nash equilibrium. Therefore, Theorem 2 can be restated as follows.

THEOREM 2*: *Let f and $f^1, f^2, \ldots, f^n$ be strategy vectors representing the one actually played and the beliefs of the players. Suppose that for every player i:*

(i) *f_i is best response to f^i_{-i}; and*

(ii) *f is absolutely continuous w.r.t. f^i.*

Then for every $\varepsilon > 0$ and a positive integer l there is a time $T = T(z, \varepsilon, l)$ such that for every $t \geqslant T$ there is a Nash equilibrium $\bar{f}$ of the repeated game satisfying $f_{z(t)}$ plays (ε, l)-like $\bar{f}$.

The obvious identical modification can be applied to Theorem 2.1 as well.

7.2. *Dispersed Beliefs and Weak Learning*

Example 2.3 suggests that if the belief assigns a positive probability to any "neighborhood" of the real strategy, then a "weak" learning may take place. The following example shows that this is not an easy task, and that careful studies of the topologies and notions of learning involved have to be conducted.

The notion of "neighborhood" called for by Example 2.3 is the following. We say that a behavior strategy f' is in the ε-neighborhood of another behavior strategy f if the probabilities they assign to any action after every history are close to each other up to an ε. In the following example we show a case where every neighborhood around the strategy denoted below by c_∞ is given positive probability and nevertheless no learning (even in a weak sense) takes place.

As before, PI has two actions: l and r. Let c_n be the stationary strategy that plays with probability $1-1/n$ the action l and with probability $1/n$ the action r, $n=1,2,\ldots,\infty$. Let d_n be the strategy which plays n times l and r thereafter. PII believes that c_n is played with probability $\alpha_n>0$ ($\Sigma\alpha_n\leqslant 1$) and d_n with β_n, while in fact PI plays constantly l, i.e., c_∞, which is assigned zero probability by PII. It is clear then that PII ascribes positive probability to any neighborhood around c_∞. One might expect that c_n's with large n will be assigned growing probabilities in the course of the infinite game. But we will show that whether or not this occurs depends on the sequences $\{\alpha_n\}$ and $\{\beta_n\}$. If the sequence $\{\alpha_n\}$ tends to zero much faster than $\{\beta_n\}$ even "weak" learning fails.

After observing t times l, the posterior probability of c_m being played is $\alpha_m(1-1/m)^t/(A_t+B_t)$ where $A_t=\sum_{m=1}^{\infty}\alpha_m(1-1/m)^t$ and $B_t=\sum_{m\geqslant t}\beta_m$. For sufficiently large t, A_t can be bounded from above as follows:

$$A_t=\sum_{m=1}^{[t^{1/2}]}\alpha_m(1-1/m)^t+\sum_{[t^{1/2}]+1}^{\infty}\alpha_m(1-1/m)^t.$$

The first term,

$$\begin{aligned}\sum_{m=1}^{[t^{1/2}]}\alpha_m(1-1/m)^t &\leqslant\left(1-1/[t^{1/2}]\right)^t\\ &=\left(1-1/[t^{1/2}]\right)^{t^{1/2}t^{1/2}}\\ &\leqslant e^{-(t^{1/2})/2}.\end{aligned}$$

The last inequality is obtained by the fact that $(1-1/[t^{1/2}])^{t^{1/2}}\to e^{-1}$ for large t's. The second term,

$$\sum_{[t^{1/2}]+1}^{\infty}\alpha_m(1-1/m)^t\leqslant\sum_{[t^{1/2}]+1}^{\infty}\alpha_m.$$

If $\alpha_m=a2^{-m}$ and $a<1$, then

$$\sum_{[t^{1/2}]+1}^{\infty}\alpha_m<2^{-[t^{1/2}]}<2^{-(t^{1/2})/2}$$

so that $A_t\leqslant 2\cdot 2^{-(t^{1/2})/2}$. Suppose now that $\beta_m=a/m^2$, where the constant a is chosen in such a way that $\Sigma_m(\alpha_m+\beta_m)=1$ ($a<1$). In this case, B_t behaves asymptotically like a/t. Thus, A_t/B_t approaches zero as t goes to infinity. We conclude that the probability assigned to the event that a c_n is played approaches zero as more observations of l arrive. Thus, the future event which consists of the infinite run of l's is given smaller and smaller probability as time goes by. Since the strategy played by PI gives this infinite run of l's probability 1, we do not have here learning in the sense discussed above. PII does not learn the future behavior of PI.

In this example, however, the updated belief assigns probability converging to 1 to the event that the next outcome will be l. Thus, PII learns to predict *near*

future behavior of PI. In other words, PII learns in a weak sense the strategy of PI.

By defining the β_n's a bit differently, we can construct an example in which every once in a while PII will expect the next outcome to be r with probability close to $1/2$, while the outcome will be always l. On an infinite set of integers, say, M, we set $\hat{\beta}_m = \Sigma_{k=m}^{\infty}\beta_k$, while if $m \notin M$ we define $\hat{\beta}_m = \beta_m$. Thus, if M is very sparse, the series $\hat{\beta}_m$ converges and, moreover, $\hat{\beta}_m/\Sigma_{k=m+1}^{\infty}\hat{\beta}_k$ tends to 1 as $m \to \infty$ if attention is restricted to $m \in M$. Now we define $\alpha_k = a2^{-n}$ and $\beta'_m = a\hat{\beta}_m$. Once again, a is chosen in a way that $\Sigma(\alpha_m + \beta'_m) = 1$. The calculation of A_t in this case is the same as the calculation above. Defining $B'_t = \Sigma_{m \geq t}\beta'_m$ one gets $A_t/B'_t \leq A_t/B_t$. Therefore, as the game evolves the set of strategies $\{c_n\}$ is getting diminishing weight. After a long time, when B_t is very close to one and when $t = m$ for some $m \in M$, out of the probability B_t (assigned to all d's) β'_m is given to d_m, which plays r at the $(m+1)$th stage. But β'_m/B'_m is close to $1/2$. We conclude that after m observations of l, the prediction of PII regarding the immediate play of PI is approximately $1/2$ on l and $1/2$ on r, while in fact PI plays l with probability 1. It means that PII does not learn, even in the weak sense, the behavior of PI. In other words, Bayesian updating will not lead to accurate prediction even of events that take place in the *near* future.

7.3. *The Necessity of Knowing Your Own Preferences*

The assumption that each player knows his own preferences is crucial. For example, Blume and Easley (1992) 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 Theorem 2.1. The difference lies in the fact that, in their example, players do not know their own payoff matrix.

More familiar, perhaps, are results regarding the multi-arm bandit problem (see, for example, Rothschild (1974)) where an optimal Bayesian learning strategy does not lead to an optimal strategy of the full information case. Since we can view the multi-arm bandit problem as a special case of the one person version of Corollary 1, optimal play corresponds to a Nash equilibrium. The discrepancy between the optimal Bayesian play and the optimal full information play contradicts the conclusion of Theorem 2.1, which states that the player should eventually behave ε-optimally as if the uncertainty were not present. The cause for this discrepancy lies in our assumption that the players know their own payoff matrix. In the multi-arm bandit problem this assumption requires that the player know the payoffs associated with the different actions, which is not true for that model.

The contrast with the multi-arm bandit problem illustrates an important point. The uncertainty in our model is regarding strategies of the opponent. Unlike nature's uncertainties, opponents' actions will continue to be observed as long as the game lasts, and thus, perfect learning of them will take place.

7.4. *The Need for Perfect Monitoring*

Consider again the multi-arm bandit problem, but now view it as a two person game with the original player being player I and nature being player II. We let PII have a flat utility function, and his action set consists of choosing one of a few possible payoff distributions for each one of PI's activities. We assume that PII made his choices randomly at the beginning according to some fixed probability distribution π, and then kept the same realized choices throughout the infinite game. We also assume that PI knows that PII followed the strategy described above but PI is not told the realized payoff choices.

It is easy to see that now we have modeled the multi-arm bandit problem as a two person infinite game. However, this game has imperfect monitoring. Nash equilibria with PII playing a strategy of the type discussed above exist, and they require that PI plays optimally against the realized choices of PII. Again, examples of optimal long term strategy in the multi-arm bandit problem violate the conclusion of our Theorem 2.1, but under this formulation, the discrepancy is explained by the failure of the perfect monitoring condition.

7.5. *Extensions*

Stochastic Games: The results of this paper should extend to the more general model of stochastic games (see Shapley (1953)) under the informational assumptions that each player knows his own payoff matrices as well as the transition probabilities and the state realizations of the stochastic game. To what extent our results generalize when the realized states are not told to the players seems to be an interesting problem.

General Continuous Payoff Functions: Since the proof of Proposition 1 (in Kalai and Lehrer (1993a)) relies only on the continuity of U_i, the proof of Theorem 1 applies also to repeated games with general payoff functions that are continuous with respect to the product topology: for example, when the discount factor changes with time.

Dept. of Managerial Economics and Decision Sciences, J. L. Kellogg Graduate School of Management, Northwestern University, Evanston, Il. 60208, U.S.A.

Manuscript received August, 1990; final revision received January, 1993.

REFERENCES

ARROW, K. J. (1962): "The Economic Implications of Learning by Doing," *Review of Economic Studies*, 29, 155–173.

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

——— (1981): "Survey of Repeated Games," in *Essays in Game Theory and Mathematical Economics in Honor of Oskar Morgenstern*. Mannheim/Wien/Zurich: Bibliographisches Institut, 11–42.

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.

BATTIGALLI, P., M. GILLI, AND M. C. MOLINARI (1992): "Learning and Convergence to Equilibrium in Repeated Strategic Interactions: An Introductory Survey," *Ricerche Economiche*, forthcoming.

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

BLACKWELL, D., AND L. DUBINS (1962): "Merging of Opinions with Increasing Information," *Annals of Mathematical Statistics*, 38, 882–886.

BLUME, L., M. BRAY, AND D. EASLEY (1982): "Introduction to the Stability of Rational Expectations Equilibrium," *Journal of Economic Theory*, 26, 313–317.

BLUME, L., AND D. EASLEY (1992): "Rational Expectations and Rational Learning," Cornell University.

BRAY, M., AND D. M. KREPS (1987): "Rational Learning and Rational Expectations," in *Arrow and the Ascent of Modern Economics Theory*, ed. by G. R. Feiwel. New York: NYU Press, pp. 597–625.

BROCK, W., R. MARIMON, J. RUST, AND T. SARGENT (1988): "Informationally Decentralized Learning Algorithms for Finite-Player, Finite-Action Games of Incomplete Information," University of Wisconsin.

CANNING, D. (1992): "Average Behavior in Learning Models," *Journal of Economic Theory*, 57, 442–472.

CRAWFORD, V. (1989): "Learning and Mixed Strategy Equilibria in Evolutionary Games," *Journal of Theoretical Biology*, 140, 537–550.

EASELY, D., AND N. KIEFER (1988): "Controlling a Stochastic Process with Unknown Parameters," *Econometrica*, 56, 1045–1064.

FUDENBERG, D., AND D. KREPS (1988): "A Theory of Learning, Experimentation and Equilibrium in Games," Stanford University.

FUDENBERG, D., AND D. LEVINE (1993a): "Steady State Learning and Nash Equilibrium," *Econometrica*, 61, 547–573.

——— (1993b): "Self Confirming Equilibrium," *Econometrica*, 61, 523–545.

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. (1985): "Learning Rational Expectations: The Finite State Case," *Journal of Economic Theory*, 36, 257–276.

——— (1991): "Bayesian Learning in Normal Form Games," *Games and Economic Behavior*, 3, 60–81.

——— (1992): "The Exponential Convergence of Bayesian Learning in Normal Form Games," *Games and Economic Behavior*, 4, 202–217.

KALAI, E., AND E. LEHRER (1990a): "Bayesian Learning and Nash Equilibrium," Northwestern University.

——— (1990b): "Merging Economic Forecasts," Northwestern University.

——— (1992): "Bounded Learning Leads to Correlated Equilibrium," Northwestern University.

——— (1993a): "Subjective Equilibrium in Repeated Games," *Econometrica*, 61, 1231–1240.

——— (1993b): "Weak and Strong Merging of Opinion," *Journal of Mathematical Economics*, forthcoming.

KIRMAN, A. P. (1983): "On Mistaken Beliefs and Resultant Equilibria," in *Individual Forecasting and Aggregate Outcomes*, ed. by R. H. Day and T. Groves. New York: Academic Press, pp. 137–156.

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

LINHART, 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 (1991): "Lakates and Experimental Economics," in *Appraising Economic Theories*, ed. by N. de Marchi and M. Balug. London: Edward Elgar, pp. 197–226.

McLennan, A. (1987): "Incomplete Learning in a Repeated Statistical Decision Problem," University of Minnesota.
Mertens, J. F. (1987): "Repeated Games," in *Proceedings of the International Congress of Mathematicians* (Berkeley, 1986). American Mathematical Society, pp. 1528–1577.
Mertens, J. F., S. Sorin, and S. Zamir (1990): *Repeated Games*, to be published.
Milgrom, P., and J. Roberts (1991): "Adaptive and Sophisticated Learning in Normal Form Games," *Games and Economic Behavior*, 3, 82–100.
Miyasawa, K. (1961): "On the Convergence of the Learning Process in a 2×2 Non-Zero-Sum Game," Research Memorandum No. 33, Princeton University.
Monderer, D., and D. Samet (1990): "Stochastic Common Learning," *Games and Economic Behavior*, forthcoming.
Nash, J. F. (1950): "Equilibrium Points in *n*-person Games," *Proceedings of the National Academy of Sciences USA*, 36, pp. 48–49.
Nyarko, Y. (1991): "Learning in Mis-Specified Models and the Possibility of Cycles," *Journal of Economic Theory*, 55, 416–427.
Pearce, D. (1984): "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica*, 52, 1029–1050.
Prasnikar, V., and A. E. Roth (1992): "Considerations of Fairness and Strategy: Experimental Data from Sequential Games," *Quarterly Journal of Economics*, 107, 865–888.
Robinson, J. (1951): "An Iterative Method of Solving a Game," *Annals of Mathematics*, 54, 296–301.
Roth, A. E., V. Prasnikar, M. Okuno-Fujiwara, and S. Zamir (1991): "Bargaining Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study," *American Economic Review*, 81, 1068–1095.
Rothschild, M. (1974): "A Two-Armed Bandit Theory of Market Pricing," *Journal of Economic Theory*, 9, 195–202.
Selten, R. (1975): "Reexamination of the Perfectness Concept for Equilibrium Points ir Extensive Games," *International Journal of Game Theory*, 4, 25–55.
——— (1988): "Adaptive Learning in Two Person Games," University of Bonn.
Shapley, L. S. (1953): "Stochastic Games," *Proceedings of the National Academy of Sciences of the USA*, 39, 1095–1100.
——— (1964): "Some Topics in Two-Person Games," in *Advances in Game Theory*, *Annals of Mathematical Studies*, 5, 1–28.
Shiryayev, A. N. (1984): *Probability*. New York: Springer-Verlag.
Smith, V. L. (1990): "Experimental Economics: Behavioral Lessons for Theory and Microeconomic Policy," Nancy L. Schwartz Memorial Lecture, Northwestern University, Evanston, Illinois.
Stanford, W. G. (1991): "Pre-Stable Strategies in Discounted Duopoly Games," *Games and Economic Behavior*, 3, 129–144.
Woodford, M. (1990): "Learning to Believe in Sunspots," *Econometrica*, 58, 277–307.