Discussion Paper No. 1138

Social Learning in Recurring Games

by

Matthew Jackson*
and
Ehud Kalai*

August 1995

*Department of Managerial Economics and Decision Sciences, J. L. Kellogg Graduate School of Management, Northwestern University, 2001 Sheridan Road, Evanston, Illinois 60208. e-mail: (Jackson) jacksonm@nwu-edu; (Kalai) e-kalai@nwu.edu.

Social Learning in Recurring Games

by

Matthew Jackson and **Ehud Kalai**

This Draft: July 1995

MEDS, Kellogg Graduate School of Management,
Northwestern University, Evanston IL 60208–2009

Abstract: In a recurring game, a stage game is played sequentially by different groups of players. Each group receives publicly available information about the play of earlier groups. Players begin with initial uncertainty about the distribution of types (representing the preferences and strategic behavior) of players in the population. Later groups of players are able to learn from the history of play of earlier groups. We first study the evolution of beliefs in this uncertain recurring setting and then study how the structure of uncertainty and information determine the eventual convergence of play. We show that beliefs converge over time and, moreover, that the limit beliefs are empirically correct: their forecast of future public information matches the true distribution of future public information. Next, we provide sufficient conditions to ensure that the play of any stage game is eventually close to that of a Bayesian equilibrium where players know the true type generating distribution. We go further to identify conditions under which play converges to the play of a trembling–hand perfect (Bayesian) equilibrium.

We are grateful to the California Institute of Technology for their generous hospitality while we undertook this research, and thank the National Science Foundation for financial support. Ehud Kalai also thanks the Sherman Fairchild Foundation. We thank Ehud Lehrer, Glenn Loury, and Roger Myerson for very helpful conversations.

1. Introduction

Notions of repeated games are useful for modelling strategic dynamic interaction in economics, political, and other social sciences. They also serve as experimental tools where convergence properties of various learning and evolutionary processes can be studied.

It is worthwhile to distinguish between two types of repetition. The first type is a repeated interaction among a fixed group of players, for example firms engaged in ongoing oligopolistic competition. Formal models of *repeated games* describe such interaction and there is a large game theoretic literature in this area. (We refer the reader to Aumann and Hart (1992, 1994) for suveys and references.) A second type of repeated interaction, introduced formally as *recurring games* in this paper, involves different players at each time. As in a repeated game, in a recurring game a stage game is repeatedly played, but each stage is played by a new group of players. The repetition is important because each new group of players may observe relevant information about the play of earlier groups. Recurring games capture the evolution of social behavior in multigenerational games, as well as interaction that occurs frequently in an existing society. For example, most of the real estate transactions which occur each year involve different participants and can naturally be modeled as a recurring game. Similarly, the numerous applications of a particular auction procedure at different times and locations constitute a recurring game. Not only are recurring games obviously applicable to a wide variety of problems of practical interest, they also provide a wonderful laboratory for theories of learning and evolution.

Of particular interest here is the process of rational (Bayesian) learning in the context of recurring games. We study conditions under which play comes to approximate equilibrium play of the static stage games, and also identify interesting situations where such convergence fails. In addition, we explore how recurring game results about rational learning can be extended to include irrational players, long-lived players, and stochastic games.

Versions of repeated and recurring games have been used to study fictitious play, naive best response behavior, genetic and adaptive algorithms, Bayesian learning, and other learning and evolutionary systems. The important contributions to this literature are too numerous to list here, but examples related to our work here include: Nash (1950) and Aumann and Maschler (1967), as well as more recent work by Fudenberg and Kreps (1988), Jordan (1991), Fudenberg and Levine (1993), and Kalai and Lehrer (1993). Of this literature, the closest in spirit to the approach taken here are Jordan (1991) and Kalai and Lehrer (1993).[1] Jordan (1991) studied a semi-rational repeated game Bayesian learning process that leads players' expectations in the long run to a Nash equilibrium, and Kalai and Lehrer (1993) studied a rational learning process that leads players in a repeated game in the long run to Nash equilibrium play. Our work is built on the same mathematical foundation of Bayesian updating and convergence of beliefs, but differs substantially in the types of conclusions reached regarding convergence to various forms of equilibrium play, and the type of applications addressed.

To understand some of these differences, it is useful to note some overall differences between learning in recurring games and learning in repeated games. In a repeated game, for instance between a wife and a husband or a parent and a child, players have the time to learn to best respond to the actual players they face and their chosen strategies. On the other hand, in a recurring game, for example recurring single exchanges between pairs of buyers and sellers, such opportunities are not available. Thus, recurring game players who study the past plays of earlier players can at most learn the distribution of strategies in the population of opponents, but not the actual strategy of their realized opponents. So, in general, convergence in repeated games may lead to Nash equilibrium play while convergence in recurring games may lead to Bayesian equilibrium play. This might be considered the "bad news."[2]

There is also some "good news" as one changes from repeated games to recurring ones. First, long run learning in a recurring game does not require that players live forever, as it does in a repeated game. Instead, only sufficient cumulative social experience must survive with time. This means that one does not have to be as concerned with the speed of convergence in recurring games, as one would in repeated games. Second, the imposition on players' rationality is less severe in a recurring game. In a repeated game a player must solve an infinite horizon dynamic optimization problem in addition to Bayesian learning.

---

[1] Although the learning literature is closest in terms of the approach to modeling behavior, some evolutionary models are closer in terms of the fit into the recurring setting. For instance, Young (1993) studies the evolution of conventions in a recurring setting: a game is played consecutively by different players and some observations of history are available to current players.

[2] Of course, the news is bad only to the extent that one wants players to play Nash equilibrium. There are games for which players are better off in ignorance, i.e., for which all Bayesian equilibrium outcomes Pareto dominate the Nash equilibrium outcomes.

In a recurring game, the player's horizon can be very short, significantly simplifying the optimization problem. For instance, if the stage game requires only a single choice of action by a player, then there is no incentive to experiment and the behavior is simply myopic.

While the above myopic phenomenon makes the recurring analysis simpler, it can also create drastic welfare differences between the results of learning in recurring games versus learning in repeated games. Examples in this paper show that individual incentives not to experiment can lead to significant social welfare losses that may require social intervention, such as affirmative action. Experimentation by a player results in an externality relative to later players as it may help them learn. One can also find examples where the myopic behavior and failures to experiment lead to social gains.

Another phenomenon which is typical to recurring games is that there may be important differences in the information and monitoring between an outside observer, say the econometrician or the public, and the actual players of a stage game. Players from current or past stages of the game may know significantly more than outside observers who are only exposed to public signals from each stage. For this reason, the perceived social explanation for what is going on, may differ from the perspective of insiders which takes into account privately available information and experience concerning payoffs and actions which might not be publicly observed.

In addition to the specialization to recurring games, the current paper also develops some general results and techniques regarding rational learning. First, by allowing a sufficiently rich set of beliefs in the minds of players, we obtain convergence to trembling hand perfect (Bayesian) equilibrium. This contrasts with the previous learning literature which obtains, at best, convergence to Nash equilibrium. Second, a simple technique of representing players enables us to incorporate into the model players who do not learn at all, or who follow other methods of learning. One application of this technique is to examine the behavior of rational players in the presence of boundedly rational players. This is in the spirit of the analysis of Kreps, Milgrom, Roberts and Wilson (1982). A second application of this technique shows that we can apply the recurring game results to situations that involve long-lived players. A third application allows us to adapt the model to capture situations where the game is stochastic, i.e., changes stochastically from one stage to the next as a function of the social history.

The paper is structured as follows. In the next section, three examples motivate and illustrate the approach and results. The following sections include the formal definitions for the general model and sufficient conditions for social learning to converge to Bayesian and trembling hand perfect Bayesian equilibrium play. Several additional examples illustrate various violations of the conditions and show how convergence may fail. The paper ends with a discussion of a method by which one can incorporate some long lived players and stochastic games into the model, and some concluding remarks.

2. Motivating and Illustrating Examples.

Our first example shows a failing of learning in recurring games. We start with it because it highlights some of the differences between the recurring and repeated settings. In this example, social learning fails even with fully revealing signals because individuals do not have socially efficient incentives to experiment.

EXAMPLE 1. *Failed Social Learning: The Need for Role Models and Affirmative Action.*

Consider the members of a population facing the option of taking on a certain task, e.g., attending law school. For the purpose of this example, divide the population into two types, those who would succeed at the task and those who would fail. An individual who succeeds receives a payoff of 1, while an individual who undertakes the task and fails has a payoff of -2. Not taking on the task results in a payoff of 0 for either type.

Without knowing his or her own type[3] a member of the population decides whether to take on the task based on the type distribution in the population. For example, if the distribution of types in the population, τ, is such that .8 of the population would succeed, then the Bayesian equilibrium (optimal decision) relative to this prior τ is for every individual to take on the task ($.8 \times 1 + .2 \times (-2) = .4 > 0$).

If, however, members of the population are not sure about the distribution of types in the population, then less desirable social outcomes are possible. Suppose, for instance, that

[3] This is easily fit into a Bayesian model as follows. Introduce a second player called "nature", whose type is either "succeed" or "fail", and who has no strategic role in the game (i.e., give player 2 only one action). The first player has only 1 type, but his or her payoff depends on the realized type of player 2. This type of construction is described in detail in Harsanyi (1967-68).

members of the population believe that it is equally likely that the population is described by $\overline{\tau}$, which is just the reverse of τ. That is, under $\overline{\tau}$, .8 of the population will fail, and the Bayesian equilibrium is to not take on the task.

Given the initial beliefs of the population that τ and $\overline{\tau}$ are equally likely, the first member of the population sees his or her chance of success as .5, and given the asymmetry in payoffs (i.e., the higher cost of failure) chooses not to undertake the task. This means that the second member of the population to make this decision will have the same available information and will also choose not to undertake the task, and so forth. If τ is the true distribution, then since no one takes on the task, no one plays the optimal strategy relative to τ. In this case, social learning fails, and indeed, this example violates the sufficient conditions for social learning that are presented in Section 6.

While social policy is not the subject of this paper, notice that being aware of the failing of social learning above has policy implications. The social equilibrium has important informational externalities that are not taken advantage of. That is, if early individuals could be convinced to take on the task, against their own perceived individual incentives, with high probability their success rate will be observed by later individuals who will update the beliefs to put higher probability on the optimistic prior τ when it is the true distribution. In that case, after sufficiently long the optimal strategy will be for members of the population to take on the task.

This example illustrates a notion of positive affirmative action and role models. It would be to the benefit of the group as a whole, *even if they are not sure whether τ or $\overline{\tau}$ is the case*, to induce early members of the population to undertake the task so that in the case τ "role models" will be created for the benefit of the large population to follow.

As a further remark on affirmative action, note that the incentives must be such that they do not obscure the success of the individuals. For example, giving a population scholarships to law school, but not changing the bar exam requirements, will serve this purpose (unless the exam requirements are to be permanently changed), as changing the requirements would produce less useful signals of the true distribution in the population.

The above example also illustrates clearly the difference between a recurring game and a repeated game. If taking on the task was a repeated action available to a single patient individual, then it would pay that individual to experiment in early stages, because of the potential benefit from learning his or her ability. Moreover, in the repeated game, the individual would learn about his or her own type, rather than the distribution of types in the population.

EXAMPLE 2. *Learning to Coordinate in the Presence of Boundedly Rational Players.*

Consider a classic "Battle of the Sexes" game which is played recurringly. The payoffs are pictured below.

	A	B
A	2,1	0,0
B	0,0	1,2

In addition, consider that some players may be "boundedly rational" who follow a *naive best response* strategy, *nbr* for short. Such a player matches the last period action of the opposite sex player. For example, an *nbr* player in the role of player 1 at time $t+1$ adopts the action chosen by the player in the role of player 2 at time t. (Let first period *nbr* players randomize equally over A and B.) Thus, an *nbr* player is implicitly assuming that the other player will choose the same thing that his or her predecessor did.

Assume that in each period, when a new pair of players is randomly drawn independently of each other and past players, they may be either Bayesian rational players or an *nbr* type. Each player knows only his or her own type and the social history of the actions chosen by both players in each stage preceeding their stage.

We begin by finding the Bayesian equilibrium strategies in the recurring game where the type generating distribution τ is publicly known to all players. With that in hand we will come back to the case where τ is unknown. Suppose that under τ the probability of a player being an *nbr* type is .4 ($\tau(nbr) = .4$), and of being a rational type is .6 ($\tau(rat) = .6$). The behavior of *nbr* players is easily predicted based on the play of the previous stage. We can thus compute the following Bayesian equilibrium strategies for the stage game at time $t+1$, where $t \geq 1$, by examining strategies of rational players and requiring that they be best responses to the distribution over the strategies of other rational players and *nbr* players

Case 1. Actions were (A,A) at time t.

The *nbr* types will play A at time $t+1$ and this is also the unique equilibrium strategy for a rational player in either role. (A rational row player expects that the column player will play A with probability at least .4, which makes A a best response for a rational row player. Then given that all row players will play A, this is the unique best response for a rational column player as well.) Thus the equilibrium actions in a stage following a stage of (A,A) will be (A,A) with probability 1.

Case 2. Actions were (B,B) at time *t*.

This is similar to case 1 and so the equilibrium actions in a stage following a stage of (B,B) will be (B,B) with probability 1.

Case 3. Actions were (A,B) at time *t*.

In this case *nbr* types will play (B,A) at time $t+1$ in their respective roles. Here the only equilibrium strategies for rational players are to play (A,B), in their respective roles. Rational players are swayed by the *nbr* types and play the opposite of them in an attempt to match their counterpart. (Matching the other player if that player is an *nbr* type leads to the highest payoff, while matching the other player if that player is rational leads to the lower payoff.) Thus, the equilibrium actions in a stage following a stage of (A,B) will be: (A,A) with probability $.6\times.4$, (B,B) with probability $.6\times.4$, (A,B) with probability $.6\times.6$, and (B,A) with probability $.4\times.4$.

Case 4. Actions were (B,A) at time *t*.

In this case *nbr* types will play (A,B) at time $t+1$ in their respective roles. For rational players, however, there are three sets of equilibrium strategies. Rational players could all play A with probability one in their respective roles. Rational players could all play B with probability one in their respective roles. Rational players could also all randomize, to place probability 4/9 on their respective favorite actions and 5/9 on the other action.

In the first equilibrium, the actions in a stage following a stage of (B,A) will be: (A,A) with probability .6 and (A,B) with probability .4.

In the second equilibrium, the actions in a stage following a stage of (B,A) will be: (B,B) with probability .6 and (A,B) with probability .4.

In the third equilibrium, the actions in a stage following a stage of (B,A) will be: (A,A) with probability 2/9, (B,B) with probability 2/9, (A,B) with probability 4/9, and (B,A) with probability 1/9.

Notice that the cases of playing (A,A) and (B,B) (that is, Cases 1 and 2 above) are absorbing when rational players follow their unique equilibrium strategies. Notice also that under recurring play, cases (A,B) and (B,A) must (with probability 1) eventually lead to one of these absorbing cases. We can thus conclude that in an equilibrium play of the above recurring game, where τ is known, players will converge to coordinate forever on (A,A) or (B,B).

Now let us consider what happens in the above situation if τ is not known to the players. Different equilibrium strategies may be realized, depending on the initial beliefs of the players concerning the relative likelihood of other type generating distributions. For example, players might believe that it is possible that an alternative $\overline{\tau}$ describes the relative likelihood of rational and *nbr* types. More generally, players may allow for other type generating distributions which assign probabilities to other types of irrational players. For instance, they may believe that there are types of players who always play their favorite action, or players who best respond to what *nbr*'s would do, etc.. Clearly, there is a large set of type generating distributions of which the above τ is only one.

Our model captures this game with uncertainty about the type generating distribution τ (the uncertain recurring game) as follows. First, a type generating distribution is randomly selected according to a commonly known prior distribution Γ. No one is told any information about the realized τ. Next, the recurring game proceeds to be played, with player roles filled according to this unknown realized distribution τ. As the game progresses, however, new players can update their beliefs about the realized τ based on the observed actions of players in earlier stages.

Our theorem regarding convergence to Bayesian equilibrium (Theorem 1, below) has strong implications here. Applied to this example it states that late players of the uncertain recurring game must play close to the Bayesian equilibrium strategies of their stage game, as if they knew the realized τ. Moreover, additional conclusions of Theorem 1 are strong enough to preserve the the absorbing properties of the plays (A,A) and (B,B). So even if the equilibrium players started with a highly diffuse prior over the distribution of types in the population, allowing for many types of boundedly rational behavior, they will learn to play as if they were playing a Bayesian equilibrium relative to the realized distribution. And, in particular, if the realized τ is the one we described earlier, then play will be absorbed to a recurring play of either (A,A) or (B,B).

We will return to a proof of the above conclusions (concerning absorption) after the formal presentation of the results.

As mentioned in the introduction, some of our results provide conditions under which players will learn to play trembling hand perfect Bayesian equilibrium. In the previous examples there was no distinction between the Bayesian equilibria and the trembling hand perfect (Bayesian) equilibria of the static games. Thus, there was no room to distinguish between learning to play Nash equilibrium and learning to play a trembling hand perfect equilibrium. This distinction is made in the following example.

EXAMPLE 3. *Learning to Coordinate on a Perfect Equilibrium.*

Consider a three player game where payoffs are known and given in the following table, where player 1 chooses a row, player 2 chooses a column, and player 3 chooses a matrix.

	a_3				$\overline{a}_3$	
	a_2	$\overline{a}_2$			a_2	$\overline{a}_2$
a_1	1,1,1	1,1,1		a_1	3,3,3	0,0,0
$\overline{a}_1$	1,1,1	1,1,1		$\overline{a}_1$	0,0,0	3,3,3

In the right hand matrix players 1 and 2 play a coordination game, while in the left hand matrix the payoffs are constant. Player 3 can thus play $\overline{a}_3$ and hope that players 1 and 2 coordinate, or play a_3 and get a lower but sure payoff.

There are continua of Nash equilibria to the above game. There are 3 trembling hand perfect equilibria of the above game: $(a_1, a_2, \overline{a}_3)$ $(\overline{a}_1, \overline{a}_2, \overline{a}_3)$, and $(1/2, 1/2, \overline{a}_3)$, where the 1/2 indicates an even mixing over the two available actions. All the trembling hand perfect equilibria involve player 3 choosing $\overline{a}_3$, while there are Nash equilibria which involve player 3 playing a_3.

Consider a situation where all players are rational and they know the payoff table (and this is all common knowledge). In this situation, there are still a multitude of equilibria of the stage game from which to choose. The uncertainty that players face is thus strictly strategic. In this example, a player's type is simply an indication of what strategy they employ after each history. Players learn by observing the actions chosen in previous stages. For instance, all that matters to a first stage player 3 is his or her beliefs over the actions to be chosen by the other players. Thus, a type of the other player is what strategy they play. If player 3 has initial uncertainty which places high enough probability on coordination then 3 will choose $\overline{a}_3$, and otherwise 3 will choose a_3. Over time players will observe past plays of the game and learn to correctly predict the actions that will be chosen in subsequent stages. Thus, they will learn to coordinate on an equilibrium. The question is whether that is simply a Nash equilibrium, or a trembling hand perfect equilibrium.

Consider the strategies $(\overline{a}_1, a_2, a_3)$, which constitute a Nash equilibrium that is not trembling hand perfect. We argue that if players have "full subjective uncertainty," in that they never completely rule out the possibility that the other players will play any action combination, and if players types are with low enough correlation, then they could only learn to play a trembling hand perfect equilibrium. Suppose to the contrary, that over time players have learned to coordinate on $(\overline{a}_1, a_2, a_3)$. Thus, they play $(\overline{a}_1, a_2, a_3)$, but each player still has beliefs that place some probability on the event that the other players use any other combination of actions. Player 1 still believes that there is a positive probability (however small) that 3 will play $\overline{a}_3$. As long as player 1 does not believe that this action of player 3 is highly correlated with player 2 playing $\overline{a}_2$, then it is better for 1 to play a_1, rather than $\overline{a}_1$. This means that $\overline{a}_1$ could not be a best response to player 1's beliefs and so $(\overline{a}_1, a_2, a_3)$ could not be the result of rational learning.

It is clear that the role of the correlation in the beliefs of player 1 concerning the possible actions of players 2 and 3 is important in the above example. Given full subjective uncertainty, this correlation determines whether convergence ends up being to a trembling hand perfect equilibrium, or to an undominated Bayesian equilibrium which might not be trembling hand perfect. This is explored in more detail in Example 7, Theorem 3, and the concluding remarks.

We turn now to a formal presentation of the model and results.

3. Recurring Games

In a recurring (Bayesian) game, a *stage game* is played at each time $t \in \{1, 2, 3, \ldots\}$. New players are randomly drawn at each time t to play the stage game.

The *stage game* is a standard Bayesian game described by a list $(N, \{A_i, \Theta_i, u_i\}_{i \in N}, \tau)$ with the following interpretations. The set $N = \{1, 2, \ldots, n\}$ describes player roles with A_i describing a finite set of (pure) *actions* available to a player in role i. $A = A_1 \times \cdots \times A_n$ describes feasible action combinations. The notation $\Delta(A_i)$ denotes the *mixed actions* available to a player in role i. Each Θ_i is a countable set describing the possible *types* of a player in role i, and $\Theta = \Theta_1 \times \cdots \times \Theta_n$ describes profiles of players' types. τ is a *type generating distribution*, which is a probability distribution defined over the profiles of types in Θ. Players' payoffs depend both on the vector of types and the chosen actions. Player i's preferences are represented by the von Neumann–Morgenstern utility function $u_i : A \times \Theta \rightarrow I\!R$. Utilities are bounded: there exists a finite $M \geq 0$ such that $|u_i(a, \theta)| \leq M$ for all a and θ.

To describe the recurring game, it is necessary to define the way in which information becomes available from each stage. S is the countable set of possible *publicly observed signals*, with generic element $s \in S$. Each vector of stage game actions $a \in A$ and type $\theta \in \Theta$ results in signals according to a probability distribution $\mu_{a,\theta}$ defined over S. A (social) *history* of length t, denoted h^t, is a vector of publicly observed signals $(s^1, s^2, \ldots, s^t) \in S^t$. Players at time $t = 1$ will not have observed any signals, so we adopt the convention that $s^0 = \{\emptyset\}$. Let $H = s^0 \cup (\cup_t S^t)$ be the set of all possible social histories.

A player in the recurring Bayesian game is denoted (i, t), representing the player of role i at time t. The recurring Bayesian game is thus described as follows.

Initially (after history $\emptyset$), a vector of types θ^1 is randomly drawn according to τ. Each player $(i, 1)$ is informed of his or her respective θ_i^1. Next, each player $(i, 1)$ chooses an action $a_i^1 \in A_i$. Players are paid $u_i(a^1, \theta^1)$ and a signal s^1 is randomly drawn according to μ_{a^1,θ^1}. $h^1 = (s^1)$ becomes the publicly known social history (of length 1). In the second stage, a new vector of types, θ^2, is is randomly drawn according to τ and each new player $(i, 2)$ is informed of his or her respective type θ_i^2 and chooses an action $a_i^2 \in A_i$. Players are paid $u_i(a^2, \theta^2)$ and a signal s^2 is randomly drawn according to μ_{a^2,θ^2}. The next social history is $h^2 = (s^1, s^2)$. The recurring game is defined inductively in the manner just described.

Strategies for players in a recurring game are represented by maps $\sigma_i : H \times \Theta_i \to \Delta(A_i)$, which prescribe a mixture over the possible pure actions available to each player role as a function of the realized type in that player role and the observed history up to that stage.[4] The utility of a player in the recurring game for a given profile of strategies is defined to be the expected payoff in the obvious way.

Uncertain Recurring Games

In a recurring game, if the players know the distribution according to which types are drawn (τ), then the analysis is similar to the analysis of a static Bayesian game (as developed by Harsanyi (1967, 1968)). It differs in that the history of previous stages may serve as a correlating device. If the players do not know the distribution τ, then history also plays an interesting role in learning. This is the focus of our analysis and is captured as follows.

Let $M(\Theta)$ be a set of probability distributions over Θ. The uncertainty of players is represented by Γ, a probability distribution over $M(\Theta)$. We assume that Γ has countable support, and thus also assume that $M(\Theta)$ is countable. Prior to the start of the recurring game, a type generating distribution $\tau \in M(\Theta)$ is drawn according to the known distribution Γ. The players then proceed to play the recurring game without any information about the realized τ. However, since players in each stage know Γ and see the history of previous play, over the course of the uncertain recurring game, they update Γ according to the observed history of public signals and their own type.

Remark: It is important that we have modeled the uncertainty in the form of a prior over priors. The game is "doubly Bayesian" in the sense that players of any stage are unsure as to which Bayesian game they are playing. They learn about which Bayesian game they are playing from updating based on the social history. If one tried instead to model this by simply endowing players with a possibly incorrect prior over types, then there would be no way for players to learn from the social history.

Remark: We require that all players start with the same uncertainty Γ. This consistency assumption is common in the literature on Bayesian games. Nevertheless, this assumption

[4] Notice the convenient abuse of notation. A full description has σ as a function of i, t, since each pair (i, t) is a new player. However, since t is identifiable from a history, we economize on the notation.

is restrictive to the degree that heterogeneity in beliefs over different τ's conditional on histories cannot be incorporated into the type space. As will become obvious, the basic learning and convergence to equilibrium results are not dependent on the common prior, and extend easily to the case of diverse priors. The modifications which are necessary concern the fact that if there are an infinite number of different priors then there might be no uniform rate of convergence.[5]

The Probability Space

Before proceeding, it will be helpful to define a probability space which serves as the basis for an uncertain recurring game, given a strategy σ.

A fully described *outcome* (or *state*) is an infinite sequence $(\tau, \theta^1, a^1, s^1, \theta^2, a^2, s^2, \ldots)$ in $M(\theta) \times (\Theta \times A \times S)^\infty$. To describe the probability distribution over outcomes, it suffices to define consistent probabilities over all initial segments of outcomes, and then P_σ is the consistent extension to the set of all outcomes.[6] This is done inductively, by letting $P_\sigma(\tau) = \Gamma(\tau)$ and

$$P_\sigma(\tau, \ldots, s^t, \theta^{t+1}, a^{t+1}, s^{t+1}) = P(\tau, \ldots, s^t)\tau(\theta^{t+1})\sigma_{\theta^{t+1}, h^t}(a^{t+1}), \mu_{a^{t+1}, \theta^{t+1}}(s^{t+1}),$$

where $h^t = (s^1, \ldots, s^t)$.

The updating that players undertake is captured by $P_\sigma(\tau | h^t, \theta_i^{t+1})$, a version of the conditional probability distribution.

4. Social Learning

The first lemma states that an observer who updates the prior distribution Γ according to the observed histories will eventually stop learning, in the sense that the updated Γ ($P_\sigma(\tau | h^t)$) converges to a limit distribution. The convergence of the probability placed on any τ is a standard consequence of the martingale convergence theorem. The fact that the

[5] We would also need to add an absolute continuity condition, requiring that each player's beliefs Γ_{θ^i} be absolutely continuous with respect to the true distribution Γ. In the absence of this condition, a player might a priori rule out the realized distribution τ and would then be unable to learn.

[6] We abuse notation and let τ denote the event consisting of all outcomes with first entry τ, and $\tau, \ldots, \theta^t, a^t, s^t$ denotes the event consisting of all outcomes with this initial segment, etc.. To complete the probability space, consider the σ-field generated by the set of events which consists of all finite initial segments.

resulting limits together still constitute a probability distribution requires a bit more work, which is given in the proof.

Lemma 1: Convergence of Beliefs. *Consider an uncertain recurring game and a list of strategies σ. For almost every $h \in S^\infty$ there exists $\Gamma^\infty \in M(\Theta)$ such that $P_\sigma(\cdot|h^t)$ converges to Γ^∞.* [7]

Proof: Consider any $\tau \in M(\Theta)$. Let $X^t = P_\sigma(\tau|h^t)$. The sequence of X^t's is a martingale. By the martingale convergence theorem (for a statement see Theorem 35.4 in Billingsley (1979)), there exists a random variable X such that X^t converges to X almost surely. For each τ define $\Gamma^\infty(\tau)$ to be this X. Let Ω_τ be the outcome set (of measure 0) for which this convergence fails. Since $M(\Theta)$ is countable, it follows that $\cup_\tau \Omega_\tau$ is a set of measure zero.

We now show that $\sum_{\tau \in M(\Theta)} \Gamma^\infty(\tau) = 1$, almost surely. This implies that $\Gamma \in M(\Theta)$ (almost surely), and that $P(\tau|h^t)$ converges to $\Gamma^\infty(\tau)$ uniformly across τ (almost surely). This is accomplished in two steps. First, we will show that

$$\sum_{\tau \in M(\Theta)} \Gamma^\infty(\tau) \leq 1 \tag{1}$$

for every outcome outside of $\cup_\tau \Omega_\tau$. Next, we will show that

$$E\left[\sum_{\tau \in M(\Theta)} \Gamma^\infty(\tau)\right] = 1. \tag{2}$$

Together, (1) and (2) imply that $\sum_{\tau \in M(\Theta)} \Gamma^\infty(\tau) = 1$, almost surely.

To prove (1), suppose to the contrary that $\sum_\tau \Gamma^\infty(\tau) > 1$ for some outcome outside of $\cup_\tau \Omega_\tau$. Let h be the history of signals associated with this outcome. Then there exists a finite $B \subset M(\theta)$ such that $\sum_{\tau \in B} \Gamma^\infty(\tau) - 1 = d > 0$, for this outcome. Let $n = \#B$. Since B is finite, there exists T such that $|P_\sigma(\tau|h^t) - \Gamma^\infty(\tau)| < d/n$ for all $\tau \in B$ and $t \geq T$. This implies that $\sum_{\tau \in B} P_\sigma(\tau|h^t) > 1$ for any $t \geq T$, a contradiction.

To prove (2), fix any finite $B \subset M(\theta)$. First we show that

$$E\left[\sum_{\tau \in B} \Gamma^\infty(\tau)\right] = \sum_{\tau \in B} \Gamma(\tau). \tag{3}$$

Since B is finite it follows that $E[\sum_{\tau \in B} \Gamma^\infty(\tau)] = \sum_{\tau \in B} E[\Gamma^\infty(\tau)]$, which by definition is equal to $\sum_{\tau \in B} E[\lim_{t \to \infty} P_\sigma(\tau|h^t)]$. By Lebesgue's bounded convergence theorem (for a

[7] $P_\sigma(\cdot|h^t)$ converges to Γ^∞ at h if for any ϵ there exists T such that for all $t > T$ $\sup_{\tau \in M(\theta)} |P_\sigma(\tau|h^t) - \Gamma^\infty(\tau)| < \epsilon$.

statement see Billingsley (1979), Theorem 16.4), it follows that

$$\sum_{\tau \in B} E[\lim_{t\to\infty} P_\sigma(\tau|h^t)] = \sum_{\tau \in B} \lim_{t\to\infty} E[P_\sigma(\tau|h^t)].$$

Since, $E[P_\sigma(\tau|h^t)] = \Gamma(\tau)$ for all t, we have established (3). One easily establishes (2) using (3) and the fact that $\sum_{\tau \in M(\Theta)} \Gamma(\tau) = 1$. ∎

Although Lemma 1 establishes that an observer's beliefs will converge over time, it does not guarantee that they converge to the true distribution τ, or even that the observer will be making correct predictions concerning forthcoming public signals. This second conclusion, however, intuitively follows from the convergence of beliefs. If the predictions are still significantly incorrect at some time, then there is room for additional learning. Thus, once learning has effectively stopped, predictions should be approximately correct. This is captured in Lemma 2, below.

First, we define beliefs to be *ϵ–empirically correct* if they induce a distribution over signals which is ϵ-close to the actual distribution over signals.

Definition: *Beliefs are ϵ–empirically correct* relative to τ and h after T if $P_\sigma(s^{t+1}|h^t)$ is ϵ close[8] to $P_\sigma(s^{t+1}|\tau, h^t)$ for all $t \geq T$.

An empirically correct set of beliefs are those where an observer of social histories, h^t, has learned to predict signals as if he or she knew the realized distribution τ.

The following result can be proven directly from Theorem 3 in Kalai and Lehrer (1993a). (See also Blackwell and Dubins (1962).)

Lemma 2: Social Learning. *For any list of strategies σ, distribution τ, and $\epsilon > 0$, there exists a random time T such that the (Bayesian updated) beliefs are ϵ–empirically correct after T.* [9]

Lemma 2 tells us that over time, based on the history of the recurring game, Bayesian observers who start with uncertainty Γ will learn to predict the signal that will occur in any stage as if they knew the true distribution, τ.[10] Lemma 2 does not imply that players

[8] μ' is ϵ–close to μ if there exists a measurable set Q such that (i) $\mu(Q) > 1 - \epsilon$ and $\mu'(Q) > 1 - \epsilon$, and (ii) $(1-\epsilon)\mu'(A) \leq \mu(A) \leq (1+\epsilon)\mu'(A)$ for every measurable $A \subseteq Q$.

[9] A random time means that for almost every outcome there exists such a T.

[10] In fact, the stronger conclusion that they will correctly predict all events involving future signals is true. In a recurring game, however, players only care about their own stage.

will come to know τ, or that they will correctly predict signals conditional also on their type (see Example 5). We now turn to an examination of these issues as they pertain to the convergence of equilibrium play in uncertain recurring games.

5. Equilibrium in Uncertain Recurring Games

Given an uncertain recurring game, let us define the expected utility of player i for a given profile of strategies σ, conditional on a history of signals through time t, h^t, and player i's type θ_i^{t+1}:

$$V_i(\sigma, h^t, \theta_i^{t+1}) = \sum_{\theta^{t+1}\in\Theta, a^{t+1}\in A} P_\sigma(a^{t+1}, \theta^{t+1} | h^t, \theta_i^{t+1})\; u_i(a^{t+1}, \theta^{t+1}). \tag{4}$$

A profile of strategies σ forms an *uncertain Bayesian equilibrium* of the uncertain recurring game if, for all i, t, $h^t \in S^t$, and $\theta_i \in \Theta_i$,

$$V_i(\sigma, h^t, \theta_i) \geq V_i(\sigma/\widetilde{a}_{h^t,\theta_i}, h^t, \theta_i)$$

for all $\widetilde{a}_{h^t,\theta_i} \in A_i$, where $\sigma/\widetilde{a}_{h^t,\theta_i}$ is the profile of strategies which alters σ only by changing σ_i to $\widetilde{a}_{h^t,\theta_i}$ after history h^t for type θ_i.

6. Convergence to Bayesian Equilibrium.

As illustrated in our previous examples, the consequences of learning depend on the informativeness of signals and the structure of the uncertainty. In this section, we present conditions under which social learning leads players in the uncertain recurring game to eventually play actions which approximate a Bayesian equilibrium relative to the true (realized) type generating distribution. We then provide examples that illustrate the roles of these conditions by showing what can happen if these conditions are violated.

In the remaining sections, we restrict attention to situations where signals are deterministic: for each a and θ there exists s such that $\mu_{a,\theta}(s) = 1$. The notation $s_{a,\theta}$ replaces the distribution $\mu_{a,\theta}(s)$. Notice that this does not imply that signals are fully revealing, as different action–type pairs could map into the same signal.

Definition: A recurring game has *payoff sufficient signals* if, for all i and a, θ, $\overline{a}, \overline{\theta}$ such that $\overline{\theta}_i = \theta_i$ and $s_{a,\theta} = s_{\overline{a},\overline{\theta}}$, it follows that $u_i(\widehat{a}_i, a_{-i}, \theta) = u_i(\widehat{a}_i, \overline{a}_{-i}, \overline{\theta})$ for all $\widehat{a}_i$.

The condition of payoff sufficient signals states that the information contained in a signal, when coupled with the information of a player's type, is sufficient for that player to calculate his or her anticipated payoff as his or her action varies.

This condition is (when coupled with the next condition) sufficient for play of an uncertain Bayesian equilibrium to converge to that of a static Bayesian equilibrium relative to the true distribution. The intuition is straightforward. Players eventually learn the correct distribution over signals. If this information, together with the rest of their information (their type) allows them to correctly calculate their payoffs, then they are choosing a best response to the distribution over their payoffs induced by the strategies of the other players under the true distribution. In the absence of this condition, players are missing payoff relevant information and could choose strategies which are best responses given the information that they have, but not to the true underlying distribution.

The payoff sufficient signals is almost a necessary condition for play to converge to resemble the play of a Bayesian equilibrium where players know the type generating distribution. It is not quite necessary, because even in the absence of any information it is possible that best responses to the initial uncertainty happen to be the same as the best responses where players know the realized type generating distribution.

Let us mention two situations where it is clear that there are payoff sufficient signals. First, if players' payoffs are private valued and signals are the actions played in a stage then signals are payoff sufficient. Second, if there are common values and signals reveal types and actions then signals are payoff sufficient.

Having payoff sufficient signals alone is not quite enough to guarantee convergence to a Bayesian equilibrium relative to the true (realized) distribution, since it is possible that one has learned to correctly predict signals unconditionally, but has not learned to correctly predict them given the additional information contained in one's own type. (This is illustrated in Example 6.) It is ruled out by the following condition.

Definition: *Social learning implies private learning* relative to σ if for any $\delta > 0$ there exists $\epsilon > 0$ such that for any $h \in S^\infty$ if $P_\sigma(s^{t+1}|h^t)$ is ϵ close to $P_\sigma(s^{t+1}|\tau, h^t)$ for all t above some T, then $P_\sigma(s^{t+1}|h^t, \theta_i^t)$ is δ–close $P_\sigma(s^{t+1}|\tau, h^t, \theta_i^t)$ for all $t > T$ and any i and θ^t such that $\tau(\theta^t) > 0$.

The social learning implies private learning condition, is easily interpreted. It states that whenever an outside observer has learned to approximately predict the signals of the next stage as if he or she knew the true type generating distribution, then any inside observer (who gets an additional piece of information θ_i) will also be able to approximately predict the signals of the next stage as if he or she knew the true type generating distribution.

We mention two situations where it is clear that social learning implies private learning. (i) Each distribution τ in the support of Γ, coupled with the strategies σ leads to a different distribution over signals. In such a case, any observer who learns the distribution over signals will also learn the true type generating distribution. (ii) Types are drawn independently so that each τ is a product of τ_i's and Γ can be decomposed into a product of distributions across τ_i's, and either signals are the actions played, or signals are the types drawn, or the signals are the actions played and the types drawn. Here, one's own type does not tell one anything new about the types or actions of the other players.

Situations where social learning does not imply private learning require particular interdependencies among types, as illustrated in Example 6.

Before we state Theorem 1, it is necessary to define ideas of closeness to a Bayesian equilibrium relative to a given distribution. There are two conclusions to the theorem. First, that after a sufficient time players choose actions yielding ϵ-optimal expected utilities as captured in the definition of a tight ϵ-Bayesian equilibrium, below. Second, after a sufficient time players' strategies play close to an actual Bayesian equilibrium of the true static game. This closeness is formalized in the definition of plays ϵ-like.

Definition: A profile of strategies, σ, form a *tight ϵ-Bayesian equilibrium* of static Bayesian game $(N, A, \Theta, \tau, \{u_i\})$ if for all i, $\theta_i \in \Theta_i$, and a_i in the support of $\sigma_i(\theta_i)$

$$V_i(\sigma/a_i, \tau, \theta_i) \geq V_i(\sigma/\overline{a}_i, \tau, \theta_i) - \epsilon$$

for all $\overline{a}_i \in A_i$. [11]

Notice that a tight ϵ-equilibrium is stronger than a standard definition of ϵ-equilibrium (as in Radner (1980)). The strenghening is that every action in the support of a player's

[11] In the static Bayesian game there is no history, so $\sigma_i : \Theta_i \to \Delta(A_i)$. V_i is the usual definition of expected utility, $V_i(\sigma, \tau, \theta_i) = \sum_\theta \tau(\theta|\theta_i) \sum_a \sigma(a|\theta) u_i(a, \theta)$.

strategy must be an ϵ–best response, rather than just the strategy itself being an ϵ–best response. This strengthening rules out the possibility of placing small probabilities on actions that lead to relatively low payoffs. This stronger conclusion turns out to be important in obtaining the absorption in Example 2 (see the discussion following Theorem 1), for instance.

Definition: Given τ, a profile of strategies of the uncertain recurring game σ *plays ϵ–like* a profile of strategies of the realized static Bayesian game $\widetilde{\sigma}$ at stage $t+1$ after history h^t, if there exists $\Theta^\epsilon \subset \Theta$ with $\tau(\Theta^\epsilon) > 1-\epsilon$ such that $\sigma_i(\theta_i, h^t)$ is ϵ–close to $\widetilde{\sigma}_i(\theta_i)$ for all $\theta_i \in \Theta_i^\epsilon$.

The above definition states that a set of strategies of the uncertain recurring game plays ϵ–like a set of strategies of the true static Bayesian game if a large enough set of types choose actions under the first set of strategies which are ϵ–close to those chosen under the second set of strategies. This definition clearly implies that the overall distribution over actions will be ϵ–close under the two strategies. (For any $\delta > 0$, there exists an $\epsilon > 0$ such that if σ plays ϵ–like $\widetilde{\sigma}$, then $P_\sigma(a^{t+1}|\tau, h^t)$ is δ–close to $P_{\widetilde{\sigma}}(a|\tau)$.) It may still be that some types that occur with very low probability are playing different actions under the two strategies. This is illustrated in example 4.

Theorem 1: Learning To Play a Bayesian Equilibrium Relative to the Realized Distribution. *Consider an uncertain recurring game with payoff sufficient signals and an uncertain Bayesian equilibrium σ such that social learning implies private learning. For every $\epsilon > 0$ and almost every τ, h there is a time T such that σ, at stage $t+1$ after history h^t, is a tight ϵ–Bayesian equilibrium of the static Bayesian game $(N, A, \Theta, \tau, \{u_i\})$ for each $t \geq T$. Moreover, for each $t \geq T$ there exists a Bayesian equilibrium $\widetilde{\sigma}$ of the static Bayesian game $(N, A, \Theta, \tau, \{u_i\})$ such that σ plays ϵ–like $\widetilde{\sigma}$ at stage $t+1$ after history h^t.*

Proof: Fix any τ such that $\Gamma(\tau) > 0$. It is sufficient to show that for any $\epsilon > 0$ and almost every h (P_σ–conditional on τ) there is a time T such that for each $t \geq T$, $\sigma(\cdot, h^t)$ is a tight ϵ–Bayesian equilibrium relative to τ and that there exists an Bayesian equilibrium $\widetilde{\sigma}$ of the static Bayesian game such that σ plays ϵ–like $\widetilde{\sigma}$.

First, note that by Lemma 2, and social learning implies private learning, it follows that for almost every h (P_σ–conditional on τ) and any $\delta > 0$ there exists T such that if

$t \geq T$, then

$$\max_{s^{t+1}} |P_\sigma(s^{t+1}|h^t, \theta_i) - P_\sigma(s^{t+1}|\tau, h^t, \theta_i)| < \delta \tag{6}$$

for all i and θ such that $\tau(\theta) > 0$.

We now show the first conclusion of the theorem: for any $\epsilon > 0$ and almost every h, there exists T such that $\sigma(\cdot, h^t)$ is a tight ϵ–Bayesian equilibrium relative to τ for each $t \geq T$. Given the payoff sufficient signals, we can write the expected payoff to a player of type θ_i after history h^t of some action a_i, given that all other players follow σ, as $V_i(a_i, \sigma, h^t, \theta_i) = \sum_{s^{t+1}} P_\sigma(s^{t+1}|h^t, \theta_i)\widehat{u}_i(a_i, \theta_i, s^{t+1})$, where $\widehat{u}_i(a_i, \theta_i, s) = u_i(a_i, \overline{a}_{-i}, \overline{\theta})$, for some $\overline{a}, \overline{\theta}$ with $\overline{\theta}_i = \theta_i$, and $s = s_{\overline{a}, \overline{\theta}}$ (and $\widehat{u}_i(a_i, \theta_i, s)$ is set arbitrarily if there is no such $\overline{a}, \overline{\theta}$).[12] Similarly, define $V_i(a_i, \sigma, \tau, h^t, \theta_i) = \sum_{s^{t+1}} P_\sigma(s^{t+1}|\tau, h^t, \theta_i)\ \widehat{u}_i(a_i, \theta_i, s^{t+1})$. Then,

$$V_i(a_i, \sigma, \tau, h^t, \theta_i) - V_i(a_i, \sigma, h^t, \theta_i) =$$

$$\sum_{s^{t+1}} [P_\sigma(s^{t+1}|\tau, h^t, \theta_i) - P_\sigma(s^{t+1}|h^t, \theta_i)]\widehat{u}_i(a_i, \theta_i, s^{t+1}).$$

So, given the bounds on utility,

$$|V_i(a_i, \sigma, \tau, h^t, \theta_i) - V_i(a_i, \sigma, h^t, \theta_i)| \leq M|\sum_a [P_\sigma(s^{t+1}|\tau, h^t, \theta_i) - P_\sigma(s^{t+1}|h^t, \theta_i)]|.$$

Choose the δ preceeding (6) to be $\frac{(\#A)\epsilon}{2M}$, and find T accordingly. Thus, we know that $|V_i(a_i, \sigma, \tau, h^t, \theta_i) - V_i(a_i, \sigma, h^t, \theta_i)| \leq \epsilon/2$ for any $t \geq T$ and any a_i. Since σ is an uncertain Bayesian equilibrium, and thus each $a_i \in \sigma_i(\theta_i, h^t)$ maximizes $V_i(a_i, \sigma, h^t, \theta_i)$, the above inequalities imply that each a_i in the support of $\sigma_i(\theta_i, h^t)$ is an ϵ–best response relative to τ.

Next, we show the second conclusion of the theorem: for any $\epsilon > 0$ and almost every h (P_σ–conditional on τ) there exists T such that for each $t \geq T$ there exists an Bayesian equilibrium $\widetilde{\sigma}$ of the static Bayesian game such that σ plays ϵ–like $\widetilde{\sigma}$. Suppose to the contrary that there exists a positive measure of h such that this does not hold. Pick any h such that (6) holds and σ does not play ϵ–like any Bayesian equilibria $\widetilde{\sigma}$ of the static game, for infinitely many t's. Index the θ_i's according to the positive integers and then for the first θ_i find a subsequence of the above t's such that $\sigma_i(\theta_i, h^t)$ converges to some $\overline{\sigma}_i(\theta_i)$. Then proceed to do the same for each θ_i in the ordering, taking a further subsequence each time.

[12] Notice that in this definition, the signal is predicted as if there were no deviation by i. Given the payoff sufficient signal condition, i can couple this information with a_i and θ_i to calculate the expected utility.

Note that for some sufficiently large t along a particular subsequence, σ plays ϵ-like $\overline{\sigma}$. (Take any finite set, $\Theta^{\epsilon} \subset \Theta$ with $\tau(\Theta^{\epsilon}) > 1 - \epsilon$ and then choose the subsequence corresponding to the last indexed $\theta_i \in \Theta_i^{\epsilon}$.) Thus, $\overline{\sigma}$ cannot be a Bayesian equilibrium relative to τ. So, there exists i, θ_i and $\overline{a}_i \in A_i$ such that $V_i(\overline{a}_i, \overline{\sigma}, \tau, \theta_i) - V_i(a_i, \overline{\sigma}, \tau, \theta_i) \geq \alpha$ for some $\alpha > 0$ and any a_i in the support of $\overline{\sigma}_i(\theta_i)$, where $V_i(a_i, \overline{\sigma}, \tau, \theta_i) = \sum_s P_{\overline{\sigma}}(s|\tau, \theta_i)\widehat{u}_i(a_i, \theta_i, s)$. For any $\nu > 0$ there is some sufficiently large T along a subsequence of the subsequence defined for θ_i such that $P_{\overline{\sigma}}(s|\tau, \theta_i)$ is ν-close to $P_{\sigma}(s|\tau, h^t, \theta_i)$ for $t \geq T$ along this subsequence. [This further subsequence is found by taking a finite subset of Θ_{-i} with large enough conditional probability (according to τ) given θ_i. Intersect the subsequences corresponding to these types θ_{-i} and θ_i so that σ_{-i} converges to $\overline{\sigma}_{-i}$ for this arbitrarily large (τ-conditional on θ_i) group of types θ_{-i}.] Thus, it follows that for sufficiently large t along this subsequence, $V_i(\overline{a}_i, \sigma, \tau, h^t, \theta_i) - V_i(a_i, \sigma, \tau, h^t, \theta_i) \geq \alpha/2$ for some a_i in the support of $\sigma_i(\theta_i, h^t)$. (For large enough t on this subsequence, the support of $\overline{\sigma}_i(\theta_i)$ is a subset of the support of $\sigma_i(\theta_i, h^t)$.) By our earlier argument (that for any $\epsilon > 0$ there exists large enough T such that $|V_i(a_i, \sigma, \tau, h^t, \theta_i) - V_i(a_i, \sigma, h^t, \theta_i)| < \epsilon$ for any $t \geq T$, θ_i and a_i), it follows that $V_i(\overline{a}_i, \sigma, h^t, \theta_i) > V_i(a_i, \sigma, h^t, \theta_i)$ for some large enough t and a_i in the support of $\sigma_i(\theta_i, h^t)$. This is a contradiction, since σ is an uncertain Bayesian equilibrium. ∎

There are two conclusions to Theorem 1. First, after a sufficient time the actions that each type of any player chooses in an uncertain Bayesian equilibrium are in fact ϵ-best responses to the actions and types expected under the (unknown) realized type generating distribution. Notice that this first conclusion could allow for actions which are different from those under a Bayesian equilibrium of the static game. The second conclusion says that after a sufficient time play must, in fact, be close to that of a Bayesian equilibrium relative to the realized type generating distribution.

Notice that the static Bayesian equilibrium $\widetilde{\sigma}$ to which play is close may vary from one stage to the next. This is due to the fact that players may condition on history and thus may be playing correlated actions across time. A very simple example is one where players play one equilibrium in even periods and another in odd periods.

With Theorem 1 in hand, let us revisit Example 2 (the Battle of the Sexes game with some naive best responders). When the τ described in Example 2 is the true distribution, we have the conclusion that rational players will play close to one of the profiles of Bayesian equilibrium strategies which are outlined in Example 2.[13] Also, they must choose actions

[13] To apply Theorem 1, let naive best responders have utility 0 for any action combination.

which are ϵ–best responses to the true distribution. In stages after (A,A) or (B,B) have been played, for ϵ small enough, since best responses are strict, this requires that rational players *exactly* follow the equilibrium strategies of (A,A) or (B,B), respectively. Thus the absorbing nature of the Bayesian equilibrium is carried over to the uncertain Bayesian equilibrium.

Theorem 1 states that after some time an uncertain Bayesian equilibrium is a tight ϵ–Bayesian equilibrium relative to the realized distribution, and that at each stage after that time it plays ϵ–like some Bayesian equilibrium relative to the realized distribution. Thus, every type is almost best responding relative to the true distribution and the aggregate actions play close to a true Bayesian equilibrium. Does this also mean that each type is almost playing the same as in some Bayesian equilibrium? The answer is not necessarily if there is an infinite number of types. This is illustrated in the following example where there are always types whose strategies are not close to any of their Bayesian equilibrium strategies.

EXAMPLE 4.

There are two players, $N = \{1,2\}$, who each have two pure actions ($A_i = \{a_i, \overline{a}_i\}$) available. Player 2 has two possible types, $\Theta_2 = \{\theta_2, \overline{\theta}_2\}$, while player 1 has a countable set of types represented by $\Theta_1 = \{1,2,3,\ldots\}$. Utility is private valued and there is perfect monitoring (signals reveal the actions chosen).

In the following tables, the first entry represents u_1 and the second u_2.

θ_2

	a_2	$\overline{a}_2$
a_1	$1+(\frac{1}{2})^{\theta_1}$,1	0,0
$\overline{a}_1$	0,1	1,0

$\overline{\theta}_2$

	a_2	$\overline{a}_2$
a_1	$1+(\frac{1}{2})^{\theta_1}$,0	0,1
$\overline{a}_1$	0,0	1,1

Distributions over types are independent. There is only one distribution of player 1 types, which is described by $\tau_1(\theta_1) = (\frac{1}{2})^{\theta_1}$. There are two distributions over player 2's types, τ_2 and $\overline{\tau}_2$. These are such that $\tau_2(\theta_2) = 1/2 = \tau_2(\overline{\theta}_2)$, while $\overline{\tau}_2(\theta_2) = 1/4$ and $\overline{\tau}_2(\overline{\theta}_2) = 3/4$.

---

so that their behavior satisfies the premise of the theorem. We discuss this technique in Section 8.

Players are uncertain and $\Gamma(\tau_1, \tau_2) = 1/2 = \Gamma(\tau_1, \overline{\tau}_2)$.

The Bayesian equilibria relative to the realized distributions are unique. Player 2 always follows the dominant strategy of a_2 if θ_2 and $\overline{a}_2$ if $\overline{\theta}_2$. If the true distribution is τ, then all types of player 1 should play a_1. If the true distribution is $\overline{\tau}$, then all types of player 1 should play $\overline{a}_1$.

In any uncertain Bayesian equilibrium of this uncertain recurring game, it is clear that player 2 will play the dominant strategy of a_2 if θ_2 and $\overline{a}_2$ if $\overline{\theta}_2$. Player 1's optimal strategy depends on the preceived (updated) distribution over types of player 2. After any history the updated Γ will still place positive weight on both distributions.

Suppose that the realized distribution is τ. Then after any history, any player 1 will update, but will still place some positive weight on $\overline{\tau}$. This means that there will exist some type θ_1 whose best response to the anticipated distribution of actions of player 2 is to play $\overline{a}_1$. This is also true of all types larger than θ_1. As learning takes place, an arbitrarily large measure of of player 1's types will be choosing the correct best response of a_1, but there will always remain some types choosing $\overline{a}_1$.

Example 4 depends on an infinite number of types. It is clear that if there are only a finite number of types possible under τ, then if we take ϵ to be small enough all types would have to be playing strategies ϵ–close to the Bayesian equilibrium strategies.

For games where the payoffs are type independent, that is, uncertainty is only strategic, we obtain the stronger conclusion that play converges to that of either a correlated or a Nash equilibrium. This is made precise in the following corollary.

Corollary 1: LEARNING TO PLAY CORRELATED AND NASH EQUILIBRIA. *Consider an uncertain recurring game with payoff sufficient signals and an uncertain Bayesian equilibrium σ such that social learning implies private learning. Suppose also that u_i is type independent for each i. For every $\epsilon > 0$ and almost every τ, h there is a time T such that for each $t \geq T$ there exists a correlated equilibrium $\tilde{\sigma}$ of the full information static game $(N, \{A_i\}, \{u_i\})$ such that σ plays ϵ like $\tilde{\sigma}$ in stage $t+1$ after history h^t. Furthermore, if player types are indepedent under τ, then $\tilde{\sigma}$ is a Nash equilibrium.*

PROOF: Consider the Bayesian equilibrium from Theorem 1. Given the type independence of u_i there is a correlated equilibrium (τ with information θ_i) which plays exactly like that equilibrium. If types are independent, then this is a Nash equilibrium. ▌

In the above corollary, society is learning to "coordinate" on what to play in a game. For instance in the case where types are independent under τ and there are several Nash equilibria to a game, then they will learn to play so that all players will coordinate on a Nash equilibrium at each stage.

In some cases, the results of the corollary can be thought of as a purification. For instance, if there is a mixed strategy equilibrium to the game, then players would play so that the perceived distribution matches the mixing, even though the players themselves may be choosing pure strategies.

We now examine failures of convergence due to violations of the conditions used in Theorem 1.

Social Versus Private Equilibrium

Example 5 shows the importance of payoff sufficient signals. Example 6 illustrates the role of the social learning implies private learning condition.

EXAMPLE 5. *Payoff Sufficient Signals: Affirmative Action Revisited.*

There are two players, $N = \{1,2\}$. Player 1 has two pure actions $A_1 = \{a_1, \overline{a}_1\}$ and player 2 has only one action $A_2 = \{a_2\}$. Player 1 has one possible type, $\Theta_1 = \{\theta_1\}$, while player 2 has two possible types, $\Theta_2 = \{\theta_2, \overline{\theta}_2,\}$. All types are rational. There is perfect monitoring (signals reveal actions).

One distribution τ is such that $\tau(\theta_1, \theta_2) = .8$, while another distribution $\overline{\tau}$ is such that $\overline{\tau}(\theta_1, \theta_2) = .2$. Players are uncertain and $\Gamma(\tau) = 1/2 = \Gamma(\overline{\tau})$. Thus, player 1's initial uncertainty treats θ_2 and $\overline{\theta}_2$ as being equally likely.

In the following tables, the first entry represents u_1 and the second u_2.

θ_2	a_2
a_1	1,0
$\overline{a}_1$	0,0

$\overline{\theta}_2$	a_2
a_1	-2,0
$\overline{a}_1$	0,0

Consider constant strategies σ where player 1 always plays $\overline{a}_1$ and player 2 always plays a_2. The strategies σ form an uncertain Bayesian equilibrium of the above uncertain recurring game. Both τ and $\overline{\tau}$ lead to the same distribution over actions under σ and so Γ is never updated, even if the payoffs are revealed (but not if the type of player 2 becomes known). However, if player 1 knew that the true distribution was τ, then his or her unique best response would be $\overline{a}_1$.

In the above example there is perfect monitoring, payoffs are made known, and social learning implies private learning. The prescribed strategies form an equilibrium of the uncertain recurring game, and yet these strategies are not close to any Bayesian equilibrium of the game where the true distribution is known. The difficulty lies in the fact that player 2's type is important in determining player 1's payoffs, and can never be learned from a history of signals where player 1 only plays $\overline{a}_1$.

The following example, shows the importance of the social implies private learning condition. In this example players have private values, there is perfect monitoring, and society learns to perfectly predict future signals. However, players never learn which of two coordination games they are playing.

EXAMPLE 6. *Social versus Private Predictions.*

There are two players, $N = \{1, 2\}$, who each have two pure actions ($A_i = \{a_i, \overline{a}_i\}$) available. Each player has two possible types, $\Theta_i = \{\theta_i, \overline{\theta}_i\}$. All types of all players are rational. There is perfect monitoring. One distribution τ is such that players' types are perfectly correlated: $\tau(\theta_1, \theta_2) = 1/2 = \tau(\overline{\theta}_1, \overline{\theta}_2)$. Another distribution, $\overline{\tau}$, has the opposite correlation structure: $\overline{\tau}(\theta_1, \overline{\theta}_2) = \overline{\tau}(\overline{\theta}_1, \theta_2) = 1/2$. Players are uncertain and $\Gamma(\tau) = 1/2 = \Gamma(\overline{\tau})$.

In the following tables, the first entry represents u_1 and the second u_2.

θ_1, θ_2

	a_2	$\overline{a}_2$
a_1	1,1	0,0
$\overline{a}_1$	0,1	1,0

$\overline{\theta}_1, \overline{\theta}_2$

	a_2	$\overline{a}_2$
a_1	1,0	0,1
$\overline{a}_1$	0,0	1,1

$\theta_1, \overline{\theta}_2$

	a_2	$\overline{a}_2$
a_1	1,0	0,1
$\overline{a}_1$	0,0	1,1

$\overline{\theta}_1, \theta_2$

	a_2	$\overline{a}_2$
a_1	1,1	0,0
$\overline{a}_1$	0,1	1,0

Consider constant strategies σ which are an even mixing over a_1 and $\overline{a}_1$, for each type of player 1 and every history. Let each type of player 2 play his or her strictly dominant action (θ_2 plays a_2, while $\overline{\theta}_2$ plays $\overline{a}_2$). The strategies σ form an uncertain Bayesian equilibrium of the described recurring game. Both τ and $\overline{\tau}$ lead to the same distribution over actions under σ and so Γ is never updated. However, if player 1 knew which of τ or $\overline{\tau}$ was realized, then conditional on his or her type player 1 would also know the type of player 2 and thus what action player 2 would take. Thus, an even mixing is not part of a Bayesian equilibrium strategy for player 1 when τ is known.

In the above example, there is perfect monitoring (signals could even show the mixture chosen), and players correctly predict public signals and strategies. Players also have private values. However, social learning does not imply private learning and player 1's actions conditional on his or her type are not part of a Bayesian equilibrium under the true distribution. The prescribed strategies form an equilibrium of the uncertain recurring game, and yet these strategies are not close to any Bayesian equilibrium of the game where the true distribution is known.

Examples 5 and 6 present us with an interesting phenomenon. There are situations where the predictions of an outside observer converge and are completely accurate concerning the actions which will be chosen in a game. Yet the players in the game have not yet learned all that they might like to concerning the uncertainty in the game and are not necessarily playing an equilibrium of the "true" game.

7. Learning to Play Perfectly

In situations where the sufficient conditions for convergence to Bayesian equilibrium are met, it will actually be quite natural to expect convergence to a refinement of Bayesian equilibrium. For instance, if a player is never completely certain of the actions that the other players might choose, then that player should never play a weakly dominated action. This means that convergence will actually be to an undominated Bayesian equilibrium of the static game. If, in addition, players types are drawn independently, then convergence will be to that of a tembling hand perfect equilibrium.

We will first present a result concerning convergence to trembling hand perfect equilibrium, and then discuss convergence to an undominated Bayesian equilibrium that may not be trembling hand perfect. To make these statements precise consider the following definitions.

Definition: A distribution Γ satisfies *full subjective uncertainty* relative to a profile of strategies σ if $P_\sigma(a_{-i}^{t+1}|h^t,\theta_i^t) > 0$ for all i, t, a_{-i}^{t+1}, and almost every h^t, θ_i^t.

Definition: A distribution Γ satisfies *full independence* if there exist $M_1(\Theta_1), \ldots M_N(\Theta_N)$ (sets of distributions over each Θ_i) and corresponding distributions Γ_i over each $M_i(\Theta_i)$ such that $\{\tau|\Gamma(\tau) > 0\} = \Pi_i\ M_i(\Theta_i)$ and $\Gamma(\tau) = \Pi_i\ \Gamma_i(\tau_i)$ for any τ such that $\Gamma(\tau) > 0$.

We define trembling hand perfect Bayesian equilibrium according to Selten (1975).[14] Let $\Delta^{Int}(A_i)$ denote the interior of $\Delta(A_i)$.

Definition: $\overline{\sigma}$ is a *trembling hand perfect Bayesian equilibrium* of the static Bayesian game $(N, S, \Theta, \tau, \{u_i\})$ if there exists a sequence of $\tilde{\sigma}^k$ with $\tilde{\sigma}_i^k(\theta_i) \in \Delta^{Int}(A_i)$ and for each i and θ_i there exists a subsequence of the k's such that $\tilde{\sigma}_i^k(\theta_i) \to \overline{\sigma}_i(\theta_i)$ and $\overline{\sigma}_i(\theta_i)$ is a best response to $\tilde{\sigma}_{-i}^k$ along the subsequence.

If there are only a finite number of types, then the mention of subsequences in the definition is inconsequential as we can intersect them to find a single sequence, and the definition is equivalent to Selten's. Here we need to apply the definition to infinite numbers

[14] A trembling hand perfect Bayesian equilibrium is thus distinct from the notion of perfect Bayesian equilibrium defined by Fudenberg and Tirole (1991) for extensive form games.

of players (viewing each player as an agent, taking the agent normal form corresponding to the Bayesian game) which accounts for the need to make explicit use of subsequences. This definition allows us to capture situations such as Example 4, where players behavior converges at different rates.

Theorem 2: LEARNING TO PLAY TREMBLING HAND PERFECT BAYESIAN EQUILIBRIUM. *Consider an uncertain recurring game such that u_i is private valued, there is perfect monitoring (signals reveal actions), and Γ is fully independent. If σ is an uncertain Bayesian equilibrium such that Γ has full subjective uncertainty relative to σ, then for every $\epsilon > 0$ and almost every τ, h there is a time T such that for each $t \geq T$ there exists a trembling hand perfect Bayesian equilibrium $\tilde{\sigma}$ of the static Bayesian game $(N, S, \Theta, \tau, \{U_i\})$ such that σ plays ϵ like $\tilde{\sigma}$ at stage $t+1$ after history h^t.*

PROOF: Fix any $\epsilon > 0$ and τ such that $\Gamma(\tau) > 0$. It is sufficient to show that for almost every h (P–conditional on τ) there is a time T such that for each $t \geq T$ there exists a trembling hand perfect Bayesian equilibrium $\tilde{\sigma}$ of the static Bayesian game $(N, A, \Theta, \tau, \{u_i\})$ such that σ plays ϵ–like $\tilde{\sigma}$

Suppose that there exists a positive measure of h such that σ does not play ϵ-like any trembling hand perfect equilibrium of the stage game for an infinite subsequence of t's. By Lemma 2 and perfect monitoring, we can find such an h such that for any δ, $\max_a |P_\sigma(a|h^t) - P_\sigma(a|\tau, h^t)| < \delta$ for all t sufficiently large.

Index the θ_i's (according to the positive integers) and then for the first θ_i find a subsequence of the above t's such that $\sigma_i(\theta_i, h^t)$ converges to some $\overline{\sigma}_i(\theta_i)$. Then proceed to do the same for each θ_i in the ordering, taking a further subsequence each time. For any $\nu > 0$ and sufficiently large t along a particular subsequence (similar to the one defined in the proof of Theorem 1), σ plays ν–like $\overline{\sigma}$. Thus, $\overline{\sigma}$ cannot be a trembling hand perfect Bayesian equilibrium relative to τ.

We now show that for each i and θ such that $\tau(\theta) > 0$, there exists a subsequence of the t's such that each a_i in the support of $\overline{\sigma}_i(\theta_i)$ is a best response to $P_\sigma(a_{-i}|h^t)$ along that subsequence. For any i and θ such that $\tau(\theta) > 0$, $\sigma_i(\theta_i, h^t)$ converges to $\overline{\sigma}_i(\theta_i)$ along the subsequence of t's used in the definition of $\overline{\sigma}$. Thus, given the finite action space, the support of $\sigma_i(\theta_i, h^t)$ contains the support of $\overline{\sigma}_i(\theta_i)$ far enough along the subsequence. By full independence, private values, and the definition of uncertain Bayesian equilibrium we know that $\sigma_i(\theta_i, h^t)$ is a best response to $P_\sigma(a_{-i}|h^t)$. Thus, far enough along the subsequence, every a_i in the support of $\overline{\sigma}_i(\theta_i)$ is a best response to $P_\sigma(a_{-i}|h^t)$, which establishes our desired conclusion.

To complete the proof of the theorem, we construct a sequence $\widetilde{\sigma}^t$ such that $\widetilde{\sigma}^t(\theta_i) \in \Delta^{Int}(A_i)$ and $\widetilde{\sigma}^t(\theta_i) \rightarrow \overline{\sigma}_i(\theta_i)$ for each θ_i, and $P_{\widetilde{\sigma}^t}(a|\tau) = P_\sigma(a|h^t)$. The construction of this sequence completes the proof of the theorem since it follows from above (and full independence) that every a_i in the support of $\overline{\sigma}_i(\theta_i)$ is a best response to $\widetilde{\sigma}^t$ far enough along a subsequence. This is a contradiction since our supposition implied that $\overline{\sigma}$ is not a trembling hand perfect Bayesian equilibrium.

To construct the sequence $\widetilde{\sigma}^t$, we first construct an auxillary sequence $\widehat{\sigma}_i^t$ for each i as follows. Order the set A_i. Consider the first $a_i \in A_i$ such that $P_{\overline{\sigma}}(a_i|\tau) > P_\sigma(a_i|h^t)$ and identify the set $\overline{\Theta}_{a_i}$ such that $\overline{\sigma}(a_i|\theta_i) > P_\sigma(a|h^t)$ iff $\theta_i \in \overline{\Theta}_{a_i}$. Find $x_{a_i}^t$ such that

$$\sum_{\theta_i \notin \overline{\Theta}_{a_i}} \tau(\theta_i)\overline{\sigma}_i(a_i|\theta_i) + \sum_{\theta_i \in \overline{\Theta}_{a_i}} \tau(\theta_i)(\max[\overline{\sigma}_i(a_i|\theta_i) - x_{a_i}^t, 0]) = P_\sigma(a_i|h^t).$$

For $\theta_i \notin \overline{\Theta}_{a_i}$ let $\widehat{\sigma}^t(a_i|\theta_i) = \overline{\sigma}_i(a_i|\theta_i)$ and for $\theta_i \in \overline{\Theta}_{a_i}$ let $\widehat{\sigma}^t(a_i|\theta_i) = \max[\overline{\sigma}_i(a_i|\theta_i) - x_{a_i}^t, 0]$. (It is clear from its definition that $x_{a_i}^t \in [0, 1]$.) Now proceed to the next a_i' in the ordering over A_i, except instead of using $\overline{\sigma}(a_i'|\theta_i)$ as the base, we use $\overline{\sigma}(a_i'|\theta_i) + [\overline{\sigma}_i(a_i|\theta_i) - \widehat{\sigma}_i^t(a_i|\theta_i)]$ (thus ensuring that the base strategies still add to one). This process is iterated.[15]

Define $\widetilde{\sigma}^t(\theta_i)$ by

$$\widetilde{\sigma}_i^t(\theta_i) = (1 - \lambda^t)\widehat{\sigma}^t(\theta_i) + \lambda^t P_\sigma(a_i|h^t),$$

where $\lambda^t > 0$ is a sequence converging to 0.

By full subjective uncertainty, we know that $\widetilde{\sigma}^t(a_i|\theta_i) \in \Delta^{Int}(A_i)$ for each θ_i. We also know that by construction $P_{\widehat{\sigma}^t}(a|\tau) = P_\sigma(a|h^t)$ and so $P_{\widetilde{\sigma}^t}(a|\tau) = P_\sigma(a|h^t)$. The last thing to show is that in a subsequence (of the previously defined subsequence) for each θ_i, $\widetilde{\sigma}^t(\theta_i) \rightarrow \overline{\sigma}_i(\theta_i)$. Given that $\lambda^t \rightarrow 0$, we need to show that $\widehat{\sigma}^t(\theta_i) \rightarrow \overline{\sigma}_i(\theta_i)$. Without loss of generality, assume that $P_\sigma(a|h^t)$ converges to $\overline{P}(a)$ along the sequence of t's. (Given the finiteness of A we can choose an appropriate subsequence where this is true). Since $P_\sigma(a|h^t)$ converges to $\overline{P}(a)$ along the sequence of t's, and since for any $\nu > 0$, far enough along particular subsequences σ plays ν-like $\overline{\sigma}$, it follows that far enough along these particular subsequences we can make $P_\sigma(a|h^t)$ arbitrarily close to $P_{\overline{\sigma}}(a|\tau)$. (See the comment after the definition of plays ϵ like). Thus $\overline{P}(a) = P_{\overline{\sigma}}(a|\tau)$ and so $P_\sigma(a|h^t)$ converges to $P_{\overline{\sigma}}(a|\tau)$. This implies that $x_{a_i}^t \rightarrow 0$ for each a_i in the construction of $\widehat{\sigma}^t$, and so $\widehat{\sigma}^t(\theta_i) \rightarrow \overline{\sigma}_i(\theta_i)$. ∎

[15] Once the distribution over an a_i has been adjusted downward so that $P_{\widehat{\sigma}^t}(a_i|\tau) = P_\sigma(a_i|h^t)$, it is not adjusted any further. If the last a_i' is the sequence is adjusted down, it must be that some previous a_i'' was such that $P_{\widehat{\sigma}^t}(a_i''|\tau) < P_\sigma(a_i''|h^t)$, and the weight taken off action a_i' can be transferred to a_i''. in such a process, each action is adjusted downward at most once and when there are no actions to adjust downward the distributions must be equal.

Again, in situations where preferences are type independent, we obtain a stronger conclusion.

Corollary 2: Learning To Play Trembling Hand Perfect Equilibrium. *Consider an uncertain recurring game such that there is perfect monitoring and an equilibrium σ. Suppose also that u_i is type independent. If Γ is fully independent and has full subjective uncertainty relative to σ, then for every $\epsilon > 0$ and almost every τ, h there is a time T such that for each $t \geq T$ there exists a trembling hand perfect equilibrium $\tilde{s}$ of the full information static game $(N, A, \{u_i\})$ such that σ plays ϵ–like $\tilde{s}$ at stage $t+1$ after history h^t.*

Proof: This follows from Theorem 2 and Corollary 1. ∎

The Role of Full Independence in Convergence to Perfect Equilibria

If the full independence condition in Theorem 2 (and corollary 2) is not satisfied, but there is full subjective uncertainty, then play still converges to a refinement of Bayesian Nash equilibrium. Notice that if a player has uncertainty which allows for every action combination of the other players, then any best response to that player's beliefs about actions of other players has to be an undominated action. Thus, in situations where play converges to a Bayesian equilibrium play and there is full subjective uncertainty, play will actually converge to an undominated Bayesian equilibrium. The following theorem is then an obvious consequence of Theorem 1.

Theorem 3: Learning To Play Undominated Bayesian Equilibrium. *Consider an uncertain recurring game with payoff sufficient signals and an equilibrium σ such that social learning implies private learning. If Γ has full subjective uncertainty relative to σ, then for every $\epsilon > 0$ and almost every τ, h there is a time T, such that for each $t \geq T$ there exists an undominated Bayesian equilibrium[16] $\tilde{\sigma}$ of the static Bayesian game $(N, S, \Theta, \tau, \{U_i\})$ such that σ plays ϵ like $\tilde{\sigma}$ at stage $t+1$ after history h^t.*

In the case of $N = 2$, the set of undominated Nash equilibria is the same as the set of trembling hand perfect equilibria (see van Damme (1987)). It is not clear to what extent

[16] Undominated is defined in the weak sense, where a_i dominates a_i' for θ_i if it yields at least as high a payoff (and sometimes higher) no matter the actions and the types of the other players (considering only those types which receives positive probability conditional on a player's type).

this result generalizes to Bayesian equilibrium. When $N \geq 3$, however, this equivalence is broken and play does not necessarily converge to trembling hand perfect equilibrium play. This is illustrated in Example 7, below. Different players end up with different beliefs after observing the same history because conditioning on their types may lead them to different updating. Thus there is no common set of trembles justifying their actions.[17]

EXAMPLE 7.

There are three players whose payoffs are type independent. Actions and payoffs represented below. Player 1 chooses a row, player 2 chooses a column, and player 3 chooses a matrix.

	l			r		
	L	M	R	L	M	R
t	1,2,1	2,0,1	0,0,0	1,2,1	2,0,0	0,0,2
b	1,1,1	0,0,1	1,0,0	1,2,1	0,0,0	1,0,2

Here all full information trembling hand perfect equilibria result in payoffs of 1,2,1.[18] The set of undominated Nash equilibria is larger, consisting of those where player 1 chooses any mixture of t and b, player 2 plays L, player 3 chooses any mixture of l and r.

Consider, the undominated Nash equilibrium (b, L, l), which is not trembling hand perfect. Let us describe initial uncertainty which leads play to converge to b, L, l.

L is strictly dominant for 2, so any equilibrium will have 2 play L. To justify b for player 1, she must believe that 2 will play R with at least twice as high a probability as 2 will play M. To justify l for player 3, he must believe that 2 will play R with at most half as high a probability as 2 will play M.[19] Such beliefs can arise in the learning enviroment, as players update based on their own types, before choosing an action. This may lead them to different updating for players 1 and 3.

[17] Similar considerations for different types of the same player are what may break the equivalence between undominated Bayesian equilibrium and trembling hand perfect Bayesian equilibrium for $N = 2$.

[18] The perfect equilibria are any mixture between t and b coupled with L, r; and t, L coupled with any mixture of l and r.

[19] This is not a perfect equilibrium since there is no single set of trembles for player 2 which would justify both 1's and 3's actions simultaneously.

For instance, let $\Theta_1 = \{\theta_1^r, \theta_1^m\}$, $\Theta_2 = \{\theta_2^r, \theta_2^M, \theta_2^R\}$, and $\Theta_3 = \{\theta_3^r, \theta_3^m\}$. θ_i^r is interpreted to be a rational type who has payoffs given in the above table. The other types are irrational and have flat preferences, so that their payoff from any action combinations is 0. Define strategies as follows. $\sigma_i(\theta_i^r)$ is the prescribed part of (b, L, l). $\sigma_i(\theta_i^m)$ is a uniform mixing across i's actions. $\sigma_2(\theta_2^M)$ plays M with probability 3/4 and R with probability 1/4. $\sigma_2(\theta_2^R)$ plays M with probability 1/4 and R with probability 3/4.

Consider the following distributions: $\tau(\theta^r) = 1$, $\overline{\tau}(\theta_1^m, \theta_2^M, \theta_3^r) = 1$, $\widetilde{\tau}(\theta_1^r, \theta_2^R, \theta_3^m) = 1$, and $\widehat{\tau}(\theta_1^m, \theta_2^r, \theta_3^m) = 1$. Initial uncertainty, Γ, has weight $\frac{3}{4}$ on τ and $\frac{1}{12}$ on each of the remaining distributions.

When τ is the true (realized) distribution only (b, L, l) will ever be played. Conditioning on histories, players will put weight increasingly on τ and decreasingly on the others, since no other distribution can lead to b, L, l. The weight on these other distributions will become arbitrarily small, but never disappear. Once player 1 sees that she is θ_1^r, she will update ruling out $\overline{\tau}$ and $\widehat{\tau}$, but retaining $\widetilde{\tau}$. This justifies playing b. Similarly, once player 3 sees that he is θ_3^r, he will update ruling out $\widetilde{\tau}$ and $\widehat{\tau}$, but retaining $\overline{\tau}$. This justifies playing l.

8. Bounded Rationality, Long Lived Players, and Stochastic Games.

Several easy technical modifications enable one to extend the previous results on uncertain recurring games with rational players to significantly richer models. Here, we briefly describe such modifications.

First, we describe the incorporation of "irrational" players. A player type following some irrational strategy can be made game theoretically rational (Bayesian, expected utility maximizing) by modifying their utility function, as done for example in Kreps, Milgrom, Roberts, and Wilson (1982) and Aumann (1992). We use an easier method (cheaper trick?) by endowing such irrational players with flat utility functions, i.e., utility functions that are independent of types and actions.[20] Clearly, any behavior is a best response given such preferences. This approach was used following Theorem 1 in discussing the Battle of the sexes game Example 2, which had rational types, as well as naive best responding types who mimicked the action of the previous period opposite player. This example fully fits the uncertain Bayesian recurring game setup once we modify the payoff to be uniformly 0 for

[20] For some "irrational" type $\widehat{\theta}_i$ let $u_i(a, \theta) = 0$ for all a and any θ such that $\theta_i = \widehat{\theta}_i$.

any player whose realized type is nbr. We then consider an uncertain Bayesian equilibrium where nbr types follow their mimicking strategy. Our results can then be applied and the asymptotic results stated in Example 2 follow. In Jackson and Kalai (1995a) we used this technique to explore Kreps–Milgrom–Roberts–Wilson results concerning chain stores in a recurring setting.

Second, the uncertain recurring model is easily modified to accomodate a recurring version of a stochastic game. Following Shapley (1953), a stochastic n person game consists of the usual individual sets of actions (A_i), but with individual utility functions $u_{i,q}$ which depend on a publicly known state q selected from a finite set of states Q. A fixed, known collection of stochastic transition rules (η_a), with $\eta_a \in \Delta(Q)$, describe the movement among states according to the vector of actions chosen in the last stage. Thus, the game starts in an initial known state q^1 and players choose actions a^1 and receive payoffs $u_{i,q^1}(a^1)$. Based on these actions, nature randomly chooses a new state q^2 according to $\eta_{a^1}(q^2)$, which is publicly announced before the second stage is played. This process is repeated.

A stochastic recurring game has only one modification. For each stage a new set of players is selected. To represent such a stochastic recurring game as a special case of the uncertain recurring games presented in this paper, model nature as an extra player, labeled 0, who has a flat utility function. Formally, nature's actions are the set of states Q, and the utility of each player $i \neq 0$, when nature chooses q and the remaining players choose a, $u_i(a,q)$, is simply defined to be the corresponding stochastic game utility $u_{i,q}(a)$. To endow the $t+1$ stage players with the knowledge of which state nature will choose in stage $t+1$, model the public signal coming out of stage t to be $s^t = (a^t, q)$ where q is randomly chosen according to η_{a^t}, and then nature follows a strategy that plays q (from s^t) at time $t+1$. Notice that this construction can be modified so that the signals only partly reveal q, and nature randomizes after observing the signal allows for substantially richer types of stochastic games with incomplete and imperfect information.

Finally, our recurring game results can be useful in modeling situations where some players are long lived. For instance, player 1 may be infinitely lived with a type drawn during the first period and fixed forever, while the remaining players are new each period and have their types drawn in each period. Such a game can be modeled by replacing the infinitely lived player 1 with a series of players $(1,t)$ who may be thought of as player 1's agents and who have flat utility functions. Further, let the type generating distributions τ be such that a single type of player one is chosen with probability one, and incorporate into Γ the original prior probability that the long lived player 1 is of the type described by τ.[21] The type of player 1 is then drawn at time 0 (when τ is drawn) and remains fixed thereafter. To such a game we can apply our results to conclude that players 2 through n will learn to best respond, each in his or her own period t, to the strategy of player $(1, t)$. Then, we can conclude that whatever the type of the long lived player 1, late recurring short lived players will be approximately best responding to player 1's actual strategy.[22]

9. Concluding Remarks

We close with comments on two assumptions that we have maintained in our analysis, and a comment on existence.

First, we have assumed that the initial uncertainty, Γ, is the same for all players. The model is easily adapted to allow the uncertainty, Γ, to depend on a player's type. The learning results (Lemmas 1 and 2) will still hold for each type separately, provided that each type's beliefs are absolutely continuous with respect to the true uncertainty. The change in the results due to type dependent uncertainty is that there may be no uniform rate of convergence when the set of types is infinite. This affects the conclusions of Theorems 1 and 3 only in that some arbitrarily small measure of types may not have converged to be ϵ-best responding to the true distribution at any time. The results of playing ϵ-like a true Bayesian equilibrium (or undominated Bayesian equilibrium) would remain unchanged since it only required convergence for a finite number of types to begin with. The convergence to trembling hand perfect Bayesian equilibrium is affected more substantially (except for 2 player games). For instance, two players might have different beliefs about the distribution of actions that a third player might choose. This could allow convergence to an undominated Bayesian equilibrium that is not trembling hand perfect, along similar lines to Example 7.

Second, we have limited attention to a countable set of types and a countable set of distributions over those types. The method of proof that we have used relies on these

[21] For each τ there exists θ_1 such that if $\tau(\overline{\theta}) > 0$, then $\overline{\theta}_1 = \theta_1$.

[22] Player 1's best response to the short lived players, may involve long term goals which are not captured in the period by period maximization of the recurring setup. Thus, a hybrid analysis of rational learning by players with lives of varying lengths is needed to fully address situations which are not strictly repeated, nor strictly recurring.

assumptions. Moreover, results by Jordan (1993) in the context of boundedly rational learning suggests that in general there may be difficulties in learning to play mixed strategy equilibria when these assumptions are relaxed. At this time, we do not have much of an idea of how the results extend when these assumptions are violated, but view it as an important issue for future research.[23]

Finally, let us say something about the existence of an uncertain Bayesian equilibrium.[24] If the set of types receiving positive probability given the initial uncertainty Γ is finite (even if the support of Γ over possible τ's is not), then existence can be established using standard results. The first stage is simply a finite Bayesian game (with a possibly inconsistent prior) for which an equilibrium exists. Using the first stage equilibrium strategies for each type, the second period updating is then clearly defined and the existence of equilibrium strategies for the second period, conditional on any of the finite possible histories, can be established, and so on. If there is an infinite set of possible types under the initial uncertainty, then the type-agent representation of the first stage game will have an infinite number of players, as possibly will subsequent stages. In such a case the existence of an uncertain Bayesian equilibrium will depend accordingly on the existence of an equilibrium in these corresponding static games.

[23] We refer the reader to Lehrer and Smorodinsky (1994) for some recent work which has implications for this issue.

[24] Nachbar (1995) raises interesting issues related to existence. He shows that in a class of infinitely repeated 2 by 2 games, if the sets of strategies that each player considers as plausible are sufficiently diverse in a specific sense, then it is not possible to have beliefs over the plausible strategies such that players learn and have plausible strategies that best respond to those beliefs. Although, the incompatibility depends on the class of games considered and appropriateness of the diversity condition (which is not easy to gauge), Nachbar's results suggest that there is still much to understand about the beliefs that are compatible with both learning and the existence of an equilibrium. Also, although Nachbar considers repeated games, there may be some analog of his results for recurring games. Such an analog is not obvious, however, since in a recurring game players care only about their own stages, thus substantially weakening the best response requirement, and are learning a distribution over types rather than a specific type.

References

AUMANN, R. J. [1992], "Irrationality in Game Theory." In P. Dasgupta, D. Gale, O. Hart, and E. Maskin, *Economic Analysis of Markets and Games: Essays in Honor of Frank Hahn*, MIT Press, Cambridge MA, pp 214–227.

AUMANN, R. J. AND S. HART (EDITORS) [1992, 1994], *Handbook of Game Theory, with Economic Applications, (Vol. 1, 2)*, North Holland: Amsterdam.

AUMANN, R. J. AND M. MASCHLER [1967], "Repeated Games with Incomplete Information: A Survey of Recent Results," *Mathematica*, Vol. 116, pp. 287–403.

BATTIGALLI, P., M. GILLI, AND M.C. MOLINARI [1992], "Learning and Convergence to Equilibrium in Repeated Strategic Interactions: An Introductory Survey," *Richerche Economiche*, Vol. 96, pp. 335–378.

BLACKWELL, D. AND L. DUBINS [1962], "Merging of Opinions with Increasing Information," *Annals of Mathematical Statistics*, Vol. 38, pp. 882–886.

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

FUDENBERG, D. AND D. KREPS [1994], "Learning in Extensive Form Games, I: Self Confirming Equilibria," *Games and Economic Behavior*, Vol. 8, pp. 20–55.

FUDENBERG, D. AND D. LEVINE [1993a], "Self-Confirming Equilibrium," *Econometrica*, Vol. 61, pp. 523–545.

FUDENBERG, D. AND D. LEVINE [1993b], "Steady State Learning and Nash Equilibrium," *Econometrica*, Vol. 61, pp. 547–573.

FUDENBERG, D. AND J. TIROLE [1991], "Perfect Bayesian Equilibrium and Sequential Equilibrium," *Journal of Economic Theory*, Vol. 53, pp. 236–260.

HARSANYI, J.C. [1967–1968], "Games of Incomplete Information Played by Bayesian Players, Parts I–III," *Managment Science*, Vol. 14, pp. 159–182, 320–334, 486–502.

JACKSON, M. AND E. KALAI [1995a], "Recurring Bullies, Trembling Challengers, and Learning," mimeo.

JACKSON, M. AND E. KALAI [1995b], "Learning to Play Perfectly in Recurring Extensive Form Games," mimeo.

JORDAN, J. [1991], "Bayesian Learning in Normal Form Games," *Games and Economic Behavior*, Vol. 3, pp. 60–81.

JORDAN, J. [1992], "The Exponential Convergence of Bayesian Learning in Normal Form Games," *Games and Economic Behavior*, Vol. , pp. 202–217.

JORDAN, J. [1993], "Three Problems in Learning Mixed-Strategy Equilibria," *Games and Economic Behavior*, Vol. 5, pp. 368–386.

KALAI, E. AND E. LEHRER [1993a], "Rational Learning Leads to Nash Equilibrium," *Econometrica*, Vol. 61, pp. 1019–1045.

KALAI, E. AND E. LEHRER [1993b], "Subjective Equilibrium in Repeated Games," *Econometrica*, Vol. 61, pp. 1231–1240.

KREPS, D., P. MILGROM, D.J. ROBERTS, AND R. WILSON [1982], "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma," *Journal of Economic Theory*, Vol. 27, pp. 245–252.

LEHRER, E. AND R. SMORODINSKY [1994], "Repeated Large Games with Incomplete Information," mimeo Tel-Aviv University.

MYERSON, R. [1991], *Game Theory: Analysis of Conflict*, Harvard University Press, Cambridge MA.

NACHBAR, J. [1995], "Prediction, Optimization, and Rational Learning in Games," mimeo: Washington University in St. Louis.

NASH, J. [1950], "Non-Cooperative Games," PhD dissertation, Princeton University Department of Mathematics.

RADNER, R. [1980], "Collusive Behavior in Noncooperative Epsilon Equilibrium of Oligopolies with Long but Finite Lives," *Journal of Economic Theory*, Vol. 22, pp. 136–154.

RUBINSTEIN, A. AND A. WOLINSKY [1990], "Rationalizable Conjectural Equilibrium: Between Nash and Rationalizability," forthcoming: *Games and Economic Behavior*.

SELTEN, R. [1975], "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games," *International Journal of Game Theory*, Vol. 4, pp. 25–55.

SHAPLEY, L.S. [1953], "Stochastic Games," *Proceedings of the National Academy of Science U.S.A.*, Vol. 39, pp. 327–332.

VAN DAMME, E. [1987], *Stability and Perfection of Nash Equilbria*, Springer Verlag.

YOUNG, P. [1993], "The Evolution of Conventions," *Econometrica*, Vol. 61, pp. 57–84.