Learning to Play Equilibria: The Bayesian, Repeated Games Approach[*]

John Nachbar[†]

March 22, 2004

1 Introduction.

This is a survey of the last 15 years or so of research into an important class of models of how players might come to play equilibria in games: Bayesian models of learning in repeated strategic form games. There have been two main objectives for this research. One has been to tell stories for how equilibrium might arise, with the ultimate aim of developing a predictive theory. The other has been, as a normative benchmark, to understand how a rational player might learn.

The talk has three main sections.

1. *What IS a model of Bayesian learning model in repeated games?* The basic observation is that any belief learning model with a deterministic prediction rule is "as if" Bayesian. In particular, many models that do not look, and arguably are not, remotely rational are Bayesian.

2. *Within Equilibrium Learning.* Learning takes place within the equilibrium of a type space game in which players know their own stage game payoffs but not those of their opponents. The theory yields a nice interpretation of mixed strategy equilibrium as a kind of Harsanyi purification with endogenous noise.

3. *Out of Equilibrium Learning.* There are conceptual problems with out of equilibrium Bayesian learning; the problems take the form of theorems stating that there are no models that simultaneously exhibit certain, otherwise desirable, properties. For the Bayesian models that have been studied, convergence obtains for some classes of games but not for others. I also briefly discuss the literature on belief learning models with stochastic prediction rules.

[*] Notes for talks at Tokyo, March 19 and March 22, 2004.

[†] Visiting Foreign Scholar, Institute for Social and Economic Research, Osaka University. Permanent affiliation: Department of Economics, Washington University, St. Louis, MO 63130. `nachbar@wueconc.wustl.edu`.

Although these notes will be fairly comprehensive, they will not be encyclopedic, and, reflecting my own interests, I will focus more on some issues more than others.[1]

2 Bayesian Learning in Repeated Games.

2.1 An example.

I start with an example. Suppose that the stage game, the game being repeated, is Matching Pennies.

	Heads	*Tails*
Heads	$1, -1$	$-1, 1$
Tails	$-1, 1$	$1, -1$

Players play this game with each other, over and over, forever. Throughout, I will refer to *actions* in the stage game and *strategies* in the repeated game.

Suppose each player believes that the other will play an i.i.d. strategy, that is, a strategy of the form, in each period, regardless of the past history of play, play *Heads* with probability q. Each player's belief can then be represented as a measure over $[0, 1]$.

Each player then best responds to his belief, by maximizing his expected discounted payoff. Although I assume throughout that players use a common discount factor δ, in the special case of i.i.d. beliefs, discounting is irrelevant: optimization is equivalent to *myopic* (period-by-period) optimization (because players in the i.i.d. model *assume* that their actions today do not affect their opponent's future behavior).

Now, any belief over repeated game strategies will generate, via integration and Bayesian updating, a *prediction rule*, which is a function from histories to next period forecasts. Suppose that each player's belief takes the form of a Beta distribution over $[0, 1]$. Then one can show that the prediction rule takes the form

$$\phi(h)(\textit{Heads}) = \frac{k}{n+k} q^\circ + \frac{n}{n+k}(\text{Empirical Frequency of action } \textit{Heads})$$

where n is the number of periods so far and k and q° are parameters that depend on the Beta distribution. In the particular case that the distribution is uniform, $q^\circ = 1/2$ and $k = 2$. This class of prediction rule is, in the language of Fudenberg and Kreps (1988), *asymptotically empirical*, since the forecast is asymptotically equal to the empirical frequency.

This Bayesian learning model with Beta priors is essentially equivalent to *fictitious play.* Fictitious play was originally proposed as an algorithm for calculating equilibria. It was also thought of as a possible model of player deliberation: players

[1] Other surveys of Bayesian learning models can be found in the relevant sections of Marimon (1997), Fudenberg and Levine (1998), and Vega-Redondo (2003).

would run through the fictitious play dynamical system in their heads (hence the "fictitious" in fictitious play) before actually the game, which they then played only once. The classic cite is Brown (1951); see also Luce and Raiffa (1957). Fudenberg and Kreps (1988) was the first paper to propose using fictitious play as a model of behavior in a repeated game, and also the first paper to point out the link between fictitious play and Bayesian models. There is now a considerable modern literature exploring fictitious play and variations thereof; see Fudenberg and Levine (1998).

2.2 The basic model.

I focus on finite two-player repeated strategic form games with perfect monitoring. By focusing on two-player games, I ignore some issues with belief correlation that are not central to the main discussion. Learning with continuum action spaces, with imperfect monitoring, and with asynchronous moves, are largely open topics. For issues that are particular to learning in extensive form games, see Fudenberg and Levine (1993b), Fudenberg and Levine (1993a), and Fudenberg and Levine (1998).

A repeated game *pure strategy* takes the form

$$s_i : \mathcal{H} \to A_i$$

where A_i is the set of stage game *actions* and $\mathcal{H}$ is the set of finite histories. (A history is a record of what each player has done in each period to date; $\mathcal{H}$ includes, to simplify notation, a null history h° representing information prior to the start of the game.) The set of pure strategies is S_i.

A *behavior strategy* is

$$\sigma_i : \mathcal{H} \to \Delta(A_i),$$

where the notation $\Delta(X)$ gives the set of measures over some (measurable) set X. The set of behavior strategies is Σ_i.

A *belief* about player i is a probability measure $\beta_i \in \Delta(\Sigma_i)$.

In a Bayesian model, each player chooses his strategy to maximize, or at least ε maximize, his expected discounted payoff, where the expectation is with respect to the measure over play paths (infinite histories) generated by his strategy and his belief as to his opponent's strategy.

Remark 1. A *mixed strategy* is a measure over S_i. By Kuhn's Theorem, every mixed strategy is *outcome equivalent* to a behavior strategy (not necessarily unique), and vice versa.[2] By "outcome equivalent" I mean that the mixed strategy and the behavior strategy induce the same distribution over play paths (infinite histories), for every possible opposing strategy. I model players as choosing behavior strategies rather than mixed strategies; by Kuhn's Theorem, this is purely a matter of interpretation. □

[2] Kuhn's Theorem, Kuhn (1964), was originally stated for finite extensive form games. Aumann (1964) contains a generalization that covers repeated games as a special case.

Remark 2. Standard ε optimization allows players to play continuation strategies that are wildly suboptimal in continuation games that are either (subjectively) unlikely or far in the future. I assume, instead, that ε optimization is *uniform*, by which I mean that player 1 chooses a strategy that generates continuation strategies that are ε optimal from the vantage point of the start of that continuation game. In the special case in which the discount factor δ is zero, uniform $\varepsilon = 0$ optimization is equivalent to myopic optimization. □

2.3 Equivalent beliefs and reduced forms.

For any belief β_2 there are typically many other beliefs that are outcome equivalent. In particular, for any β_2 there is a behavior strategy σ_2 that is outcome equivalent to β_2 (this is a minor variation on Kuhn's Theorem). I call such a σ_i a *reduced form* of β_i. The reduced form will be unique if all histories are reachable under β_i; otherwise, there will be an infinity of outcome equivalent reduced forms, differing at histories.

Conversely, given any σ_i there is a belief β_i for which σ_i is the reduced form. Indeed, simply give σ_i probability 1. There are typically more interesting equivalent beliefs; see Jackson, Kalai, and Smorodinsky (1999).

Example 1. Consider a 2×2 game like Matching Pennies and consider the belief that gives probability 1 to *sequence strategies*, strategies that depend on time but are otherwise independent of history: strategies of the form, play a in period 1, play a in period 2, play b in period 3 and so on. The set of such strategies can be identified in a natural way with the interval $[0, 1]$ and so a belief can be viewed as a distribution over $[0, 1]$. The uniform distribution over $[0, 1]$ has as its reduced form the i.i.d. behavior strategy, randomize 50:50 in every period. □

2.4 *As if* Bayesian models.

Given a belief β_2, any reduced form σ_2 can be interpreted as a *prediction rule* ϕ, giving one period ahead forecasts as a function of history, and vice versa. This means that any "belief learning" model in which players best respond to a deterministic prediction rule is *as if* Bayesian.

Fictitious play was one example. I think that the Bayesian interpretation of fictitious play is natural. The fictitious play prediction rule seems to reflect some underlying view that the opponent's play is i.i.d. The Bayesian interpretation makes this explicit.

Another example is the best response dynamics of Cournot (1838), in which players best respond to the forecast that the opponent will do again next period whatever he did this period. The Cournot prediction rule is consistent with a large number of Bayesian beliefs, including the belief that puts probability 1 on the strategy, "play a in every period, regardless of history" (if a is the prediction of the first period). This Bayesian belief "works," in the sense of yielding the desired reduced form, because most histories are unreachable, allowing the reduced form to be pretty much anything. This Bayesian interpretation is weak, obviously, but this weakness accurately reflects the weakness of the Cournot prediction rule, which jumps capriciously from one extreme forecast to the next.

Because the set of prediction rules is so large, virtually any pattern of observed actions is consistent with *some* form of Bayesian learning model. Belief learning rules out only strictly dominated behavior, and even then, ε optimization can let some strictly dominated behavior back in. One can still argue as to whether Bayesian learning describe the thought process of actual players. Perhaps actual players follow some alternative, like reinforcement learning Roth and Erev (1995) or regret minimization, Hart and Mas-Colell (2000). For a pro-Bayesian perspective on this question, based on experimental data that attempts to elicit beliefs, see Nyarko and Schotter (2002).[3] The issue is in one sense moot: if observed play is "as if" Bayesian then the relevant question is less whether Bayesian models are correct than whether Bayesian models are useful. My own view is that they are but that the issue is far from settled.

3 Within Equilibrium Learning

3.1 Type space models.

For within equilibrium learning to be interesting, one needs more structure, one needs something to learn.

Rather than assume that stage game payoffs are common knowledge, assume instead that each player is privately informed, prior to the start of repeated play, about his own payoff function. In the 2×2 case, player i's payoff type is a four-tuple, an element of $\mathbb{R}^4$. The type profile is an element of $\mathbb{R}^8$. Normalize payoffs to the unit sphere (ruling out the zero game). Let $\Theta = \Theta_1 \times \Theta_2$ denote the (normalized) space of types and let ρ be the joint distribution over types, with marginal distribution ρ_i.

For simplicity, my discussion focuses on the special case in which ρ_i has a strictly positive density.

A *meta strategy* for player i is a measure μ_i on $\Theta_i \times \Sigma_i$, with the property that ρ_i is the marginal on Θ_i. What I am calling a meta strategy is more commonly called a *distributional strategy* – Milgrom and Weber (1985). I use the term "meta" here to draw attention to the fact that μ_i is a strategy (in the type space game) governing strategies (in the repeated game).

Any μ_i induces a reduced form (taking expectations over θ_i) behavior strategy σ_i: μ_i and σ_i induce the same distribution over play paths, for any opposing meta

[3] This is also as good a time as any to note that many non-Bayesian learning models exhibit behavior that is qualitatively similar to some version of fictitious play. See, for example, Hopkins (2002).

strategy. Let $\sigma = (\sigma_1, \sigma_2)$. σ represents the belief of an outside observer who (a) knows μ but (b) does not know the type realization.

I am interested in the posterior of the μ_i, conditional on the realized history h. A convenient fact is that, because the reduced form builds in Bayesian updating, the posterior of μ_i conditional on the history h has as its reduced form the continuation strategy, which I denote σ_{ih}. σ_h represents the posterior of an outside observer who knows μ, which gives intended play as a function of type, and the realized history, but not the type realization.

3.2 The basic result.

A Nash equilibrium in meta strategies exists, Jordan (1995), and has the following convergence property.

> If μ is a Nash equilibrium of the type-space game then, for μ almost every realization of the type profile and path of play, σ_h eventually plays like a Nash equilibrium of the realized repeated game.

By "plays like," I mean that σ_h induces a distribution over continuation play paths that is asymptotically like that of a Nash equilibrium, at least over finite continuation histories.

The fundamental cites in this literature are Jordan (1991) and Jordan (1995). The particular version I've presented here follows Jackson and Kalai (1999) (which was, however, for recurring, rather than repeated, games). Kalai and Lehrer (1993a) covers within equilibrium learning for the case in which ρ has a countable support, rather than a density. Important additional cites include Nyarko (1994) and Nyarko (1998).

The intuition for the convergence result is as follows. Arbitrarily fix a reference type profile (θ_1^*, θ_2^*). For each player i, suppose that i's actual type is drawn from a neighborhood N_i of θ_i^* and consider the reduced form of i's meta strategy, conditional on that neighborhood, call it σ_i^N. Player 1's belief puts probability 1 on player 2's actual meta strategy (the players are in equilibrium). Therefore, player 1's belief puts, in effect, positive probability on σ_2^N. Since the players' actual repeated game strategy profile is a realization of σ^N, it follows from results in Kalai and Lehrer (1993a) that player 1's forecast along the path of play will be, for σ^N almost all realizations of the play path, asymptotically as accurate as if player 1 had assigned probability 1 to σ_2^N. That is, player 1's forecast along the path of play will be asymptotically as accurate as if he had known that player 2's type was drawn from N. Moreover, since μ is an equilibrium, σ_1^θ, the reduced form conditional on player 1's realized type, is a best response for a type θ_1 player 1. Thus, σ_1^N is an average of best responses over the neighborhood N_1. Continuity then implies that σ_1^N is an approximate best response for any type realization in N_1. It follows that the continuation strategy profile $\left(\sigma_{1h}^N, \sigma_{2h}^N\right)$ is asymptotically an approximate subjective equilibrium of the continuation game.[4] This in turn implies that the induced distribution over finite continuation play paths is asymptotically that of an approximate Nash equilibrium; see Kalai and Lehrer (1993a) and Kalai and Lehrer (1993b). The approximation can be made better by taking tighter neighborhoods N. The actual argument, of course, requires some care.

3.3 What happens to actual play?

The convergence result says that the play generated by σ_h is close to that of a Nash equilibrium of the realized repeated game, even though σ_h does not condition on types. What does this mean in terms of actual play in the game? This question is easiest to address in the special case of myopic optimization, which eliminates certain repeated game effects.

Under myopia, a Nash equilibrium of a realized repeated game plays like a sequence of stage game equilibria.

If the play predicted by σ_h converges to a sequence of strict equilibria then realized play must likewise converge to a sequence of strict equilibria, possibly different equilibria in different periods.

The more interesting case concerns mixed strategies. Suppose that the realized stage game is Matching Pennies. Then play under σ_h asymptotically looks like both players are randomizing 50:50. For an outside observer, this means that actual play is, asymptotically, empirically indistinguishable from mutual play of the i.i.d. 50:50 strategy.

But realized strategies are *not* random. As was noted by Fudenberg and Kreps (1988), and then more generally by Jordan (1993), for ρ almost every type profile, and any history, every player will have a strict, hence pure, myopic best response; see also Foster and Young (2001) for a related result for the general discounting case. Thus realized strategies are pure and in particular do not form an equilibrium of the repeated game. What happens in equilibrium is that, as play proceeds, each player gets an increasingly accurate fix on the other player's type, but the best response strategies depend so delicately on the precise realized type that players remain uncertain about their opponent's actions in the continuation game.

This is a form of *purification*, in the style of Harsanyi (1973). In Harsanyi (1973), if we consider a type space version of Matching Pennies, in which payoffs are bumped by small, privately observed, noise parameters, the distribution over stage game actions induced by the equilibrium of the one-shot type space game is approximately the same as the distribution induced by the Nash equilibrium of Matching Pennies, even though actual play in the type space game is pure. Moreover, the approximation is better the tighter the distribution of the noise parameters. In the

[4] In a subjective equilibrium, players best respond to their beliefs, and their beliefs are correct along the path of play, but players may be in error about how their opponent would respond off the path of play.

Bayesian learning model, the noise distribution is, in a sense, endogenous: the noise distribution tightens as the game unfolds.

Remark 3. I have assumed that ρ has a density. If ρ has an atom at, say, θ, and θ is realized, then for large h, σ^{θ}_{h}, the continuation reduced form conditional on the realized type profile, will play approximately like a Nash equilibrium; the basic cite is Kalai and Lehrer (1993a). If θ is like Matching Pennies, this implies that, unlike in the case where ρ has a density, players will actually be randomizing. □

4 Out of Equilibrium Learning

The results thus far are for within equilibrium: in effect, each player knows the other player's meta strategy. They answer the question, "how does equilibrium arise in the realized repeated game," by assuming that players are already in an equilibrium of the type space game. But how did *that* equilibrium arise?

For simplicity, I focus on the base model, with known stage game payoffs. Most of the issues with out of equilibrium learning, both positive and negative, can be illustrated with the base model, and can be extended to the type space model.

4.1 An impossibility result.

I start with a negative result that illustrates some general properties of out of equilibrium Bayesian models. I am following Nachbar (2004), which is a generalization of Nachbar (1997).

Consider once again the i.i.d. model and either Matching Pennies or a repeated coordination game. For such games, beliefs in the i.i.d. model are *inconsistent*: although each player is certain that the other will play an i.i.d. strategy, neither actually plays an i.i.d. strategy because neither has a best response, or even an ε best response, that is i.i.d. It is as though each player thinks the other is less sophisticated than he himself is.

The point of Nachbar (2004) is that this inconsistency is a general feature of Bayesian models. For any beliefs β_i and strategy subsets $\hat{\Sigma}_i \subset \Sigma_i$, consider the following properties.

1. *Learnability.* For any $\sigma = (\sigma_1, \sigma_2) \in \hat{\Sigma}$, for σ almost every path of play, each player's one-period ahead forecast is asymptotically as good as if he knew his opponent's strategy. For example, if player 2's actual behavior is i.i.d. 50:50 then the requirement is that player 1's next period forecast is asymptotically 50:50.

2. *Consistency.* Each player i has a uniform ε best response in $\hat{\Sigma}_i$, for any $\varepsilon > 0$.

3. *Caution and Symmetry* (CS).

(a) There is a $\xi > 0$ such that for any $\sigma_i \in \hat{\Sigma}_i$ there is a $s_i \in \hat{\Sigma}_i$ such that, for any h, if $s_1(h) = a$ then $\sigma_i(h)(a) > \xi$.

(b) For any pure strategy $s_2 \in \hat{\Sigma}_2$ and any function $\gamma_{21} : A_2 \to A_1$ there is a pure strategy $s_1 \in \hat{\Sigma}_1$ with

$$s_1(h) = \gamma_{12}(s_2(h))$$

The result is the following.

> For any repeated game from a large class, there is no belief profile and no $\hat{\Sigma}$ for which $\hat{\Sigma}$ simultaneously satisfies learnability, CS, and consistency.

The "large class" comprises all repeated games in which neither player has a weakly dominant action in the stage game, provided δ is small enough, and all repeated games in which each player's stage game pure action maxmin payoff is less than his minmax payoff (both stage games above have this property), for any δ. The latter class of games includes repeated Matching Pennies and many repeated coordination games.

Example 2. If the $\hat{\Sigma}_i$ comprise all strategies in some complexity class, according to any standard definition of complexity, the same complexity class for both players, then CS holds. Thus, CS holds for the set of all strategies that have memory k, the set of all strategies that are automaton implementable, the set of strategies that are Turing implementable, the set of i.i.d., strategies, and, of course, the full strategy set $\hat{\Sigma} = \Sigma$. It follows that for any such $\hat{\Sigma}_i$, for a large class of games, either consistency or learnability fails, for any beliefs. In the i.i.d. example, learnability holds and hence, as already noted, consistency fails.

Informally, what is going on here is that if, given beliefs, learnability holds then, for a large class of games, a player's best response must be more complicated than any of the strategies a player anticipates his opponent might play. Thus, for example, if the opponent's $\hat{\Sigma}_i$ comprises all strategies of memory m or less, the player's own best response must have a memory of *more* than m.

This is a game theoretic cousin of an observation about Bayesian forecasting made by Dawid (1985), building on Oakes (1985). Roughly put, Dawid (1985) points out that if a Bayesian, looking at the data generated by an unknown stochastic process, thinks that the set of possible stochastic processes satisfies a diversity condition similar to CS then the Bayesian's prediction rule, which is itself a stochastic process, will be, loosely speaking, more complicated than any of the processes that he thinks are possible.

Finally, although my discussion has assumed that beliefs are such that learnability holds, if $\hat{\Sigma}_i$ is too large then this is impossible. In particular, if $\hat{\Sigma}_i = \Sigma_i$ then both CS and consistency hold so, for any beliefs, learnability fails (for all games with at least two stage game actions for each player): the set of all strategies is too big for learnability. This can also be seen by a more direct diagonalization argument along the lines of Oakes (1985). □

Example 3. Consider repeated Matching Pennies and suppose that each player's belief gives probability 1 to the 50:50 i.i.d. strategy. Let $\hat{\Sigma}_i$ be the singleton set consisting of this one strategy. These beliefs form a Nash equilibrium, so consistency holds. Moreover learnability holds trivially (there is nothing to learn). But CS fails.

Alternatively, let $\hat{\Sigma}_i$ be the set of sequence strategies and let beliefs be uniform over this set, as in Example 1. Consistency holds (for these beliefs, *any* strategy is a best response) and CS holds. But learnability fails: learning the sequence would be like learning the realization of the 50:50 i.i.d. strategy.

As noted in Example 1, these two beliefs are outcome equivalent. As this illustrates, the impossibility result is robust to the particular belief representation chosen. Regardless of what representation one chooses, and regardless of what $\hat{\Sigma}_i$ one considers, one (or more) of the three conditions will fail. *Which* condition fails will depend on the $\hat{\Sigma}_i$. □

The basic intuition for the result is as follows. For any game in the class considered, for any σ_1, there is a σ_2 that is an *evil twin* of σ_1: σ_1 is not an ε best response, for all ε sufficiently small, to any belief for which the path of play generated by (σ_1, σ_2) is learnable.

It follows that if $\hat{\Sigma}$ is learnable and if σ_1's evil twin is in $\hat{\Sigma}_2$ then σ_1 is not ε optimal for ε sufficiently small. Say that beliefs have the *evil twin property* if *every* strategy in the support of player 2's belief about player 1 has an evil twin in the support of player 1's belief about player 2, and vice versa. If learnability is satisfied and if the evil twin property holds then consistency fails. The proof then follows by showing that the diversity condition CS implies the evil twin property.

If one takes a normative view of the result — what properties *should* a learning model have — then the result has been criticized in two main ways.

First, some have objected to the learnability condition. Why should one, in effect, focus on belief representations in which the belief is a distribution over learnable strategies? Note that there is a tension in these models. Because players discount, they care only about the near future. But learnability is about the distant future. Why should a player, when formulating his beliefs, care about the distant future? My view on this is that although a player may not care much about date 1,001 from the perspective of date 0, he cares a great deal about date 1,001 from the perspective of date 1,000. And in a Bayesian model, players cannot step outside the model at date 1,000 and change their forecasts; good forecasting must be built into the prior. In a Bayesian model, belief formation must be modeled as if players care about the distant future. To me, dropping learnability looks like a non-starter.[5]

[5] Note also that the learnability requirement is stronger than needed for the impossibility result. A weaker sufficient condition is the following. Suppose that there is a sequence $\{\sigma_1^{\varepsilon}\}$ of strategies, possibly not distinct, in $\hat{\Sigma}_1$ such that, for each $\varepsilon > 0$, σ_1^{ε} is a uniform ε best response to player 1's belief. Then for at least one such sequence, for any strategy σ_2, player 1 learns the path of play generated by $(\sigma_1^{\varepsilon}, \sigma_2)$. Thus player 1 is required to learn the path of play only if his own strategy makes sense.

Second, some have objected to CS because CS is stated in a form that is independent of payoffs. CS requires, for example, that a strategy be in $\hat{\Sigma}_i$ even if it is not rationalizable, even if it is strictly dominated, even if it is downright idiotic. This is justified, I think, by an appeal to player caution, the same sort of caution that motivates weak dominance arguments. I think that the payoff independent version of CS is a natural benchmark. But I also think that CS may well fail in an articulated model of *boundedly* rational belief formation. What is less clear to me is *how* it will fail.

Remark 4. The impossibility result extends to type space models; see Nachbar (2001). In type space models, the $\hat{\Sigma}$ are sets of meta strategies, rather than sets of repeated game strategies. Thus, players learn to forecast meta strategies. As I have already noted, in a type space model, players may never learn to forecast the realized repeated game strategies. □

Remark 5. The analog of the impossibility result holds even if players are subject to small trembles. But the definitions are more cumbersome. □

4.2 Convergence to Nash equilibrium.

Say that beliefs $\beta = (\beta_1, \beta_2)$ satisfies the *learnable best response* (LBR) property if there is a strategy profile $\sigma = (\sigma_1, \sigma_2)$ such that (a) each σ_i is a best response and (b) if players play σ then they will learn to forecast the path of play. Thus, if the players play these best responses then their continuation strategies are best responses to beliefs that are asymptotically correct along the path of play. This implies that the continuation strategies are asymptotically in subjective equilibrium, which implies that the path of play generated by the continuation strategies is asymptotically that of a Nash equilibrium of the continuation game. The fundamental cites are Kalai and Lehrer (1993a) and Kalai and Lehrer (1993b).

Remark 6. Kalai and Lehrer (1993a) provides a characterization of learnability in terms of an absolute continuity condition on distributions over play paths. A strong form of learnability will hold if, whenever the true distribution governing play assigns positive probability to some event in the space of play paths, beliefs do so as well. See also Kalai and Lehrer (1994). This absolute continuity condition generalizes the condition that one's belief put positive probability on the opponent's actual strategy, a condition that Kalai and Lehrer (1993a) call *grain of truth.* The absolute continuity/grain of truth condition is useful in application; I alluded to it in my sketch of the proof of the main convergence result for within equilibrium learning.

Absolute continuity is *violated* in the i.i.d. example. For concreteness, suppose the prior over *i.i.d.* strategies is uniform. If player 2 is actually randomizing $q^* = \pi/4$ then player 1 will learn to predict the path of play in the sense that player 1 thinks that player 2 is i.i.d. $q = q^* + \eta$, and η goes to zero asymptotically. By the Strong Law of Large Numbers, the true distribution assigns probability 1 to the event that the empirical frequency of player 2's future play is q^*. But, also by the Strong Law of Large Numbers, player 1 assigns probability 1 to the event that the empirical frequency of player 2's future play is q, and hence probability 0 to the event that the empirical frequency of her opponent's future play is q^*. And player 1 continues to assign probability 0 to this event throughout the entire history of the game (under the uniform prior, because q^* is irrational, the posterior $q \neq q^*$ for any finite history).

If instead player 1's prior had assigned an atom of probability to q^*, even if that probability were extremely small, then player 1 would have assigned positive probability to the event that the empirical frequency of her opponent's future play is q^*, and the posterior on this event would (with probability 1 under the true distribution) have converged to 1. □

Suppose one had a Bayesian model in which beliefs could be represented as a distribution over sets $\hat{\Sigma}_i$ that are learnable. If $\hat{\Sigma}$ is consistent then LBR holds. But why should consistency hold? One conjecture would be that consistency would hold provided the $\hat{\Sigma}_i$ were big, satisfying some form of diversity condition, a condition like CS. But the impossibility theorem says that consistency and learnability are fundamentally incompatible with CS. In this sense, the impossibility theorem is extremely bad for Bayesian explanations of why players might learn to play an equilibrium.

There are three important qualifications. The first is that convergence to equilibrium play can obtain even if consistency fails. The most widely studied "as if" Bayesian model is fictitious play and its variants. For the version of fictitious play known as smooth fictitious play, convergence to equilibrium play obtains for all repeated games in which the stage game falls into one of the following categories (not mutually exclusive):

- 2×2,
- zero sum,
- games with an interior ESS (i.e., games with a fully mixed Nash equilibrium that satisfies an additional, dynamic stability-like criterion),
- potential games (games in which all players receive the same payoffs, and some related games),
- supermodular games (games in which actions are linearly ordered and in which incentives to choose a higher action are increasing in the other player's action).

See Hofbauer and Sandholm (2002). See also Ellison and Fudenberg (2000) for related analysis of 3×3 games. Smooth fictitious play is the same as standard fictitious play except that players follow a smooth selection from their stage game ε best response correspondence rather than strict best response. In particular, players randomize when they are close to indifferent. One motivation is some form of bounded rationality; perhaps a smooth best response is easier to implement.

A second motivation, introduced by Fudenberg and Kreps (1988), is that stage game payoffs are subject to small i.i.d. shocks. A third motivation is that smooth fictitious play, in contrast to standard fictitious play, satisfies a property called *safety*, meaning that players do not earn less than minmax, on average, no matter what their opponents do; Fudenberg and Levine (1995) and Fudenberg and Levine (1998). Note that this last motivation is not Bayesian. Bayesian players do not *expect* to earn less than minmax, but they may nevertheless earn less than minmax if their beliefs are wrong. It is not hard, however, to imagine evolutionary stories in which an implicit concern for safety might become built into behavior.

There are, it is reasonable to speculate, other interesting classes of games in which smooth fictitious play converges as well. But there are also classes of games where no variant of fictitious play converges to equilibrium play. The following non-convergence example originates in Shapley (1962).

	L	C	R
T	$1,0$	$0,0$	$0,1$
M	$0,1$	$1,0$	$0,0$
B	$0,0$	$0,1$	$1,0$

Under any fictitious play variant, play in this game cycles endlessly through the stage game payoff matrix, never settling down to the unique Nash equilibrium of the game (namely, probability 1/3 on all actions). See also Jordan (1993).

The second qualification is that consistency may hold if CS fails. As noted earlier, CS may be too strong as a diversity condition and it may be, at least for some classes of games, that there are compelling diversity conditions that do not rule out consistency. I do not, however, have any examples to offer.[6]

Finally, there may be convergence to Nash equilibrium in some weaker sense than I have been considering. For example, in Matching Pennies, the empirical frequency of play under standard (rather than smooth) fictitious play converges to 50:50. That is, the path of play passes *some* statistical checks for Nash equilibrium play. But actual play is *not* random; it exhibits patterns. For example, if ROW and COLUMN both played *Heads* last period then ROW will always play *Heads* again next period. For a recent discussion of play that passes some, but not all, tests for Nash equilibrium, see Sandroni and Smorodinsky (2004). Note that in the Shapley example, fictitious play does not converge to Nash equilibrium even in this weaker sense.

I do not know of any non-trivial out of equilibrium Bayesian learning model that exhibits universal convergence to Nash equilibrium for all stage games. This is sometimes viewed as a defect of the Bayesian approach but I do not subscribe

[6] I should, however, take note of Sandroni (2000). That paper establishes convergence to cooperation in coordination games (games within the class considered by the impossibility result) by imposing conditions directly on the prediction rule. These conditions, while appealing, are difficult to interpret from a Bayesian perspective.

to this view. To the degree that the aim is to get a predictive theory, then not getting convergence is the right answer in games in which real people fail to play a Nash equilibrium. Whether there exists a tractable Bayesian learning model that will provide good predictions is another matter.

4.3 Non-Bayesian belief learning models.

I conclude by briefly discussing a class of *non-Bayesian* belief learning models. In these models, players best respond to prediction rules, but their prediction rules are *stochastic*. For example, in Matching Pennies, following some history, there might by a 60% chance that player 1 will think his opponent will play *Heads* and a 40% chance that she will think her opponent will play *Tails*. This implies that, from the opponent's perspective, there is a 60% chance that player 1 will play *Heads* and a 40% chance that she will play *Tails*: from the opponent's perspective, the randomness in player 1's belief is inducing randomness in player 1's action. In contrast, if player 1 held the deterministic belief that player 2 plays *Heads* with probability 60% and *Tails* with probability 40% then player 1 would play *Heads* for certain.

A number of papers have developed learning theories in which prediction rules are stochastic functions of history; see Young (1993), Hurkens (1995), Sanchirico (1996), Sonsino (1997), Jehiel (1998) (in a slightly different context — alternating move games), Fudenberg and Levine (1999) (for what they call the endogenous case of categorical smooth fictitious play), and Foster and Young (2003).

There are two threads to this literature that I would like to highlight. First the randomness that comes from random beliefs is useful for knocking players out of non-equilibrium behavior, as in the cycling that occurs under fictitious play in the Shapley (1962) example discussed above. In particular, the randomness plays a critical roll in Foster and Young (2003), which gets convergence to Nash equilibrium for all finite stage games.

Second, the randomness plays a key roll in getting prediction rules that satisfy certain statistical properties regardless of the opponents' actual strategy. In particular, there has been great interest in prediction rules that satisfy calibration, a property introduced by Dawid (1982). A forecast is calibrated if, for example, the empirical frequency of *Heads* is asymptotically 1/2 for those dates on which the forecasted probability of *Heads* was 1/2. Thus, a Bayesian *expects* that her forecasts will be calibrated, but a Bayesian may be wrong. It follows from Oakes (1985) that no Bayesian will be calibrated, even approximately, for all possible opposing strategies. But Foster and Vohra (1998) showed that it was possible to get approximate universal calibration using a stochastic prediction rule; see also Fudenberg and Levine (1996). And if players best respond to prediction rules that are calibrated then there is a sense in which play converges to correlated equilibrium; Foster and Vohra (1997). By using stochastic prediction rules, one can do even better, and make predictions that pass large sets of statistical checks for prediction accuracy, although one cannot pass *all* possible statistical checks (of which there are a continuum) for all possible opposing strategies; see Sandroni, Smorodinsky, and Vohra (2003).

The non-Bayesian belief learning models, while possessing many attractive properties, have been difficult to motivate on decision theoretic grounds. One story that I find intriguing, a story very much in the spirit of Foster and Young (2003), among other papers, is that a Bayesian recognizes that it is mathematically impossible for his beliefs to take into account every possible eventuality, in the sense of learning to make accurate predictions no matter what his opponent does, and therefore a Bayesian is willing to reconsider his beliefs, from time to time, and possibly form a new belief. Of course, this is what happens, in effect, under the Cournot best response prediction rule. The proposal, then, is for belief formation that is like a much more sensible version of the Cournot model, and the conjecture is that such a model would induce a stochastic prediction rule. The basic problem with constructing such a model is that, since we don't have a compelling story for how beliefs are formed in the first place, we don't have a compelling story for how beliefs should be changed.

References

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

Brown, G. W. (1951): "Iterative Solutions of Games By Fictitious Play," in *Activity Analysis of Production and Allocation*, ed. by T. J. Koopmans, pp. 374–376. John Wiley, New York.

Cournot, A. (1838): *Researches into the Mathematical Principles of the Theory of Wealth*. Kelley, New York, Translation from the French by Nathaniel T. Bacon. Translation publication date: 1960.

Dawid, A. P. (1982): "The well calibrated Bayesian," *Journal of the American Statistical Association*, 77(379), 605613.

——— (1985): "The Impossibility of Inductive Inference," *Journal of the American Statistical Association*, 80(390), 340–341.

Ellison, G., and D. Fudenberg (2000): "Learning Purified Mixed Equilibria," *Journal of Economic Theory*, 90(1), 84–115.

Foster, D., and R. Vohra (1997): "Calibrated Learning and Correlated Equilibrium," *Games and Economic Behavior*, 21, 40–55.

——— (1998): "Asymptotic Calibration," *Biometrika*, 85, 379–390.

Foster, D., and P. Young (2001): “On the Impossibility of Predicting the Behavior of Rational Agents,” *Proceedings of the National Academy of Sciences*, 98, 12848–12853.

——— (2003): “Learning, Hypothesis Testing, and Nash Equilibrium,” *Games and Economic Behavior*, 45, 73–96.

Fudenberg, D., and D. Kreps (1988): “A Theory of Learning, Experimentation, and Equilibrium in Games,” Stanford University.

Fudenberg, D., and D. Levine (1993a): “Self-Confirming Equilibrium,” *Econometrica*, 61(3), 523–545.

——— (1993b): “Steady State Learning and Nash Equilibrium,” *Econometrica*, 61(3), 547–574.

——— (1995): “Universal Consistency and Cautious Fictitious Play,” *Journal of Economic Dynamics and Control*, 19, 1065–1089.

——— (1996): “An Easier Way to Calibrate,” *Games and Economic Behavior*, pp. 131–137.

——— (1998): *Theory of Learning in Games*. MIT Press, Cambridge, MA.

——— (1999): “Conditional Universal Consistency,” *Games and Economic Behavior*, 29, 104–130.

Harsanyi, J. (1973): “Games with Randomly Disturbed Payoffs: A New Rationale for Mixed-Strategy Eequilibrium Points,” *International Journal of Game Theory*, 2, 1–23.

Hart, S., and A. Mas-Colell (2000): “A Simple Adaptive Procedure Leading to Correlated Equilibrium,” *Econometrica*, 68(5), 1127–1150.

Hofbauer, J., and W. Sandholm (2002): “On the Global Convergence of Stochastic Fictitious Play,” *Econometrica*, 70(6), 2265–2294.

Hopkins, E. (2002): “Two Competing Models of How People Learn in Games,” *Econometrica*, 70(6), 2141–2166.

Hurkens, S. (1995): “Learning by Forgetful Players,” *Games and Economic Behavior*, 11, 304–329.

Jackson, M., and E. Kalai (1999): “False Reputation in a Society of Players,” *Journal of Economic Theory*, 88(1), 40–59.

Jackson, M., E. Kalai, and R. Smorodinsky (1999): “Bayesian Representation of Stochastic Processes under Learning: de Finetti Revisited,” *Econometrica*, 67(4), 875–893.

Jehiel, P. (1998): “Learning to Play Limited Forecast Equilibria,” *Games and Economic Behavior*, pp. 274–298.

Jordan, J. S. (1991): “Bayesian Learning in Normal Form Games,” *Games and Economic Behavior*, 3, 60–81.

——— (1993): “Three Problems in Learning Mixed-Strategy Nash Equilibria,” *Games and Economic Behavior*, 5(3), 368–386.

——— (1995): “Bayesian Learning in Repeated Games,” *Games and Economic Behavior*, 9, 8–20.

Kalai, E., and E. Lehrer (1993a): “Rational Learning Leads to Nash Equilibrium,” *Econometrica*, 61(5), 1019–1045.

——— (1993b): “Subjective Equilibrium in Repeated Games,” *Econometrica*, 61(5), 1231–1240.

——— (1994): “Weak and Strong Merging of Opinions,” *Journal of Mathematical Economics*, 23, 73–86.

Kuhn, H. W. (1964): “Extensive Games and the Problem of Information,” in *Contributions to the Theory of Games, Volume II*, ed. by M. Dresher, L. S. Shapley, and A. W. Tucker, pp. 193–216. Princeton University Press, Annals of Mathematics Studies, 28.

Luce, R. D., and H. Raiffa (1957): *Games and Decisions.* John Wiley and Sons, New York.

Marimon, R. (1997): “Learning from Learning in Economics,” in *Advances in Economics And Econometrics*, ed. by D. Kreps, and K. Wallis, vol. I, chap. 9. Cambridge University Press, Cambridge, UK.

Milgrom, P., and R. Weber (1985): “Distributional Strategies for Games with Incomplete Information,” *Mathematics of Operations Research*, 10, 619–632.

Nachbar, J. H. (1997): “Prediction, Optimization, and Learning in Repeated Games,” *Econometrica*, 65, 275–309.

——— (2001): “Bayesian Learning in Repeated Games of Incomplete Information,” *Social Choice and Welfare*, 18(2), 303–326.

——— (2004): “Beliefs in Repeated Games,” Washington University, St. Louis.

Nyarko, Y. (1994): “Bayesian Learning Leads to Correlated Equilibria in Normal Form Games,” *Economic Theory*, 4, 821–841.

——— (1998): "Bayesian Learning and Convergence to Nash Equilibria Without Common Priors," *Economic Theory*, 11(3), 643–655.

Nyarko, Y., and A. Schotter (2002): "An experimental study of belief learning using elicited bleiefs," *Econometrica*, 70(3), 971–1006.

Oakes, D. (1985): "Self-Calibrating Priors Do Not Exist," *Journal of the American Statistical Association*, 80(390), 339.

Roth, A., and I. Erev (1995): "Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term," *Games and Economic Behavior*, 8, 164–212.

Sanchirico, C. W. (1996): "A Probabilisitc Model of Learning in Games," *Econometrica*, pp. 1375–1393.

Sandroni, A. (2000): "Reciprocity and Cooperation in Repeated Coordination Games: The Principled-Player Approach," *Games and Economic Behavior*, 32(2), 157–182.

Sandroni, A., and R. Smorodinsky (2004): "Belief-Based Equilibrium," *Games and Economic Behavior*, 47, 157–171.

Sandroni, A., R. Smorodinsky, and R. Vohra (2003): "Calibration with Many Checking Rules," *Mathematics of Operations Research*, 47(1), 141–153.

Shapley, L. (1962): "On the Nonconvergence of Fictitious Play," Discussion Paper RM-3026, RAND.

Sonsino, D. (1997): "Learning to Learn, Pattern Recognition, and Nash Equilibrium," *Games and Economic Behavior*, 18, 286–331.

Vega-Redondo, F. (2003): *Economics and the Theory of Games*. Cambridge University Press, Cambridge, UK.

Young, H. P. (1993): "The Evolution of Conventions," *Econometrica*, 61(1), 57–84.