# Learning in Games

Jonathan Levin

May 2006

Most economic theory relies on an equilibrium analysis, making use of either Nash equilibrium or one of its refinements. As we discussed earlier, one defense of this is to argue that Nash equilibrium might arise as a result of learning and adaptation. Some laboratory evidence — such as Nagel's beauty contest experiment — seems consistent with this. In these notes, we investigate theoretical models of learning in games.

A variety of learning models have been proposed, with different motivations. Some models are explicit attempts to define dynamic processes that lead to Nash equilibrium play. Other learning models, such as stimulus-response or re-inforcement models, were introduced to capture laboratory behavior. These models differ widely in terms of what prompts players to make decisions and how sophisticated players are assumed to be. In the simplest models, players are just machines who use strategies that have worked in the past. They may not even realize they're in a game. In other models, players explicitly maximize payoffs given beliefs; these beliefs may involve varying levels of sophistication. Thus we will look at several approaches.

## 1 Fictitious Play

One of the earliest learning rules to be studied is fictitious play. It is a "belief-based" learning rule, meaning that players form beliefs about opponent play and behave rationally with respect to these beliefs.

Fictitious play works as follows. Two players, $i = 1, 2$, play the game $G$ at times $t = 0, 1, 2, ....$ Define $\eta_i^t : S_{-i} \to \mathbb{N}$ to be the number of times $i$ has observed $s_{-i}$ in the past, and let $\eta_i^0(s_{-i})$ represent a starting point (or fictitious past). For example, if $\eta_1^0(U) = 3$ and $\eta_1^0(D) = 5$, and player two plays $U, U, D$ in the first three periods, then $\eta_1^3(U) = 5$ and $\eta_1^3(D) = 6$.

Each player *assumes* that his opponent is using a stationary mixed strategy. So beliefs in the model are given by a distribution $\nu_i^t$ on $\Delta(S_j)$. The

standard assumption is that $\nu_i^0$ has a Dirichlet distribution, so

$$\nu_i^0(\sigma_{-i}) = k \prod_{s_{-i} \in S_{-i}} \sigma_{-i}(s_{-i})^{\eta_i^0(s_{-i})}$$

Expected play can then be defined as:

$$\mu_i^t(s_{-i}) = \mathbb{E}_{\nu_i^t} \sigma_{-i}(s_{-i})$$

The Dirichlet distribution has particularly nice updating properties, so that Bayesian udating implies that

$$\mu_i^t(s_{-i}) = \frac{\eta_i^t(s_{-i})}{\sum_{s_{-i} \in S_{-i}} \eta_i^t(s_{-i})}. \tag{1}$$

In other words, this says that $i$ forecasts $j$'s strategy at time $t$ to be the empirical frequency distribution of past play.[1]

Note that even though updating is done correctly, forecasting is not fully rational. The reason is that $i$ assumes (incorrectly) that $j$ is playing a stationary mixed strategy. One way to think about this is that $i$'s *prior* belief about $j$'s strategy is wrong, even though he updates correctly from this prior.

Given $i$'s forecast rule, he chooses his action at time $t$ to maximize his payoffs, so

$$s_i^t \in \arg \max_{s_i \in S_{-i}} g_i(s_i, \mu_i^t)$$

This choice is myopic. However, note that myopia is consistent with the assumption that opponents are using stationary mixed strategies. Under this assumption, there is no reason to do anything else.

**Example** Consider fictitious play of the following game:

| | $L$ | $R$ |
|---|---|---|
| $U$ | $3,3$ | $0,0$ |
| $D$ | $4,0$ | $1,1$ |

Period 0: Suppose $\eta_1^0 = (3,0)$ and $\eta_2^0 = (1, 2.5)$. Then $\mu_1^0 = L$ with probability 1, and $\mu_2^0 = \frac{1}{3.5}U + \frac{2.5}{3.5}D$, so play follows $s_1^0 = D$ and $s_2^0 = L$.

[1] The whole updating story can also be dropped in favor of the direct assumption that players just forecast today's play using the naive forecast rule (1).

Period 1: $\eta_1^1 = (4,0)$ and $\eta_2^1 = (1,3.5)$, so $\mu_1^1 = L$ and $\mu_2^1 = \frac{1}{4.5}U + \frac{3.5}{4.5}D$. Play follows $s_1^1 = D$ and $s_2^1 = R$.

Period 2: $\eta_1^2 = (4,1)$ and $\eta_2^2 = (1,4.5)$, so $\mu_1^2 = \frac{4}{5}L + \frac{1}{5}R$ and $\mu_2^2 = \frac{1}{5.5}U + \frac{4.5}{5.5}D$. Play follows $s_1^1 = D$ and $s_2^1 = R$.

Basically, $D$ is a dominant strategy for player 1, so he *always* plays $D$, and eventually $\mu_2^t \to D$ with probability 1. At this point, player will end up playing $R$.

**Remark 1** *One striking feature of fictitious play is that players don't have to know anything at all about their opponent's payoffs. All they form beliefs about is how their opponents will play.*

An important question about fictitious play is what happens to the sequence of play $s^0, s^1, s^2, ....$ Does it converge? And to what?

**Definition 1** *The sequence $\{s^t\}$ converges to $s$ if there exists $T$ such that $s^t = s$ for all $t \geq T$.*

**Definition 2** *The sequence $\{s^t\}$ converges to $\sigma$ in the time-average sense if for all $i, s_i$ :*

$$\lim_{T\to\infty} \frac{1}{T+1} \left[\#\ times\ s_i^t = s_i\ in\ \{0,1,...,T\}\right] = \sigma_i(s_i)$$

Note that the former notion of convergence only applies to pure strategies. The latter is somewhat more standard, though it doesn't mean that players will actually every play a Nash equilibrium in any given period.

**Proposition 1** *Suppose a fictitious play sequence $\{s^t\}$ converges to $\sigma$ in the time-average sense. Then $\sigma$ is a Nash equilibrium of $G$.*

**Proof.** Suppose $s^t \to \sigma$ in the time-average sense and $\sigma$ is not a NE. Then there is some $i, s_i, s_i'$ such that $\sigma_i(s_i) > 0$ and $g_i(s_i', \sigma_{-i}) > g_i(s_i, \sigma_{-i})$. Pick $\varepsilon > 0$ such that $\varepsilon < \frac{1}{2}\left[g_i(s_i', \sigma_{-i}) - g_i(s_i, \sigma_{-i})\right]$ and choose $T$ such that whenever $t \geq T$,

$$\left|\mu_i^t(s_{-i}) - \sigma_{-i}(s_{-i})\right| < \frac{\varepsilon}{2N}$$

where $N$ is the number of pure strategies. We can find such a $T$ since $\mu_i^t \to \sigma_{-i}$. But then for any $t \geq T$,

$$\begin{aligned} g_i(s_i, \mu_i^t) &= \sum g_i(s_i, s_{-i})\mu_i^t(s_{-i}) \\ &\leq \sum g_i(s_i, s_{-i})\sigma_{-i}(s_{-i}) + \varepsilon \\ &< \sum g_i(s_i', s_{-i})\sigma_{-i}(s_{-i}) - \varepsilon \\ &\leq \sum g_i(s_i', s_{-i})\mu_i^t(s_{-i}) = g_i(s_i', \mu_i^t) \end{aligned}$$

So after $t$, $s_i$ is never played, which implies that as $T \to \infty$, $\mu_j^t(s_i) \to 0$ for all $j \neq i$. But then it can't be that $\sigma_i(s_i) > 0$, so we have a contradiction. *Q.E.D.*

**Remark 2** *The proposition is intuitive if one thinks about it in the following way. Recall that Nash equilibrium requires that (i) playes optimize given their beliefs about opponents play, and (ii) beliefs are correct. Under fictitious play, if play converges, then beliefs do as well, and in the limit they must be correct.*

It is important to realize that convergence in the time-average sense is not necessarily a natural convergence notion, as the following example demonstrates.

**Example** (Matching Pennies)

| | $H$ | $T$ |
|---|---|---|
| $H$ | $1, -1$ | $-1, 1$ |
| $T$ | $-1, 1$ | $1, -1$ |

Consider the following sequence of play:

| | $\eta_1^t$ | $\eta_2^t$ | Play |
|---|---|---|---|
| 0 | $(0,0)$ | $(0,2)$ | $(H,H)$ |
| 1 | $(1,0)$ | $(1,2)$ | $(H,H)$ |
| 2 | $(2,0)$ | $(2,2)$ | $(H,T)$ |
| 3 | $(2,1)$ | $(3,2)$ | $(H,T)$ |
| 4 | $(2,2)$ | $(4,2)$ | $(T,T)$ |
| 5 | $(2,3)$ | $(4,3)$ | $(T,T)$ |
| 6 | ... | ... | $(T,H)$ |

Play continues as $(T,H),(H,H),(H,H)$ — a deterministic cycle. The time average converges to $(\frac{1}{2}H + \frac{1}{2}T, \frac{1}{2}H + \frac{1}{2}T)$, but players never actually use mixed strategies, so players never end up playing Nash.

Here's another example, where fictitious play leads to really perverse behavior!

**Example** (Mis-coordination)

| | $A$ | $B$ |
|---|---|---|
| $A$ | $1,1$ | $0,0$ |
| $B$ | $0,0$ | $1,1$ |

Consider the following sequence of play:

| | $\eta_1^t$ | $\eta_2^t$ | Play |
|---|---|---|---|
| 0 | $(1/2,0)$ | $(0,1/2)$ | $(A,B)$ |
| 1 | $(1/2,1)$ | $(1,1/2)$ | $(B,A)$ |
| 2 | $(3/2,1)$ | $(1,3/2)$ | $(A,B)$ |
| 3 | ... | ... | $(B,A)$ |

Play continues as $(A,B),(B,A),...$ — again a deterministic cycle. The time average converges to $(\frac{1}{2}A + \frac{1}{2}B, \frac{1}{2}A + \frac{1}{2}B)$, which is a mixed strategy equilibrium of the game. But players never successfully coordinate!!

A few more results for convergence of fictitious play.

**Proposition 2** *Let $\{s^t\}$ be a fictitious play path, and suppose that for some $t$, $s_t = s^*$, where $s^*$ is a strict Nash equilibrium of $G$. Then $s_\tau = s^*$ for all $\tau > t$.*

**Proof.** We'll just show that $s_{t+1} = s^*$, the rest follows from induction. Note that:

$$\mu_i^{t+1} = (1-\alpha)\mu_i^t + \alpha s_{-it}$$

where $\alpha = 1/\left(\sum_{s_{-i}} \eta_t^t(s_{-i}) + 1\right)$, so:

$$g_i(a_i; \mu_i^{t+1}) = (1-\alpha) g_i(a_i, \mu_i^t) + \alpha g_i(a_i, s_{-i}^*)$$

but $s_i^*$ maximizes both terms, to $s_i^*$ will be played at $t+1$. *Q.E.D.*

**Proposition 3** *(Robinson, 1951; Miyasawa, 1951) If $G$ is a zero-sum game, of if $G$ is a $2 \times 2$ game, then fictitious play always converges in the time-average sense.*

An example of this is matching pennies, as studied above.

**Proposition 4** *(Shapley, 1964) In a modified version of Rock-Scissors-Paper, fictitious play does not converge.*

**Example** Shapley's Rock-Scissors-Papers game has payoffs:

| | $R$ | $S$ | $P$ |
|---|---|---|---|
| $R$ | $0,0$ | $1,0$ | $0,1$ |
| $S$ | $0,1$ | $0,0$ | $1,0$ |
| $P$ | $1,0$ | $0,1$ | $0,0$ |

Suppose that $\eta_1^0 = (1,0,0)$ and that $\eta_2^0 = (0,1,0)$. Then in Period 0: play is $(P,R)$. In Period 1, player 1 expects $R$, and 2 expects $S$, so play is $(P,R)$. Play then continues to follow $(P,R)$ until player 2 switches to $S$. Suppose this takes $k$ periods. Then play follows $(P,S)$, until player 1 switches to $R$. This will take $\beta k$ periods, with $\beta > 1$. Play then follows $(R,S)$, until player 2 switches to $P$. This takes $\beta^2 k$ periods. And so on, with the key being that each switch takes longer than the last.

To prove Shapley's result, we'll need one Lemma. Define the time-average of payoffs through time $t$ as:

$$U_i^t = \frac{1}{t+1}\sum_{\tau=0}^{t} g_i(s_i^\tau, s_{-i}^\tau).$$

Define the expected payoffs at time $t$ as:

$$\tilde{U}_i^t = g_i(s_i^t, \mu_i^t) = \max_{s_i} g_i(s_i, \mu_i^t)$$

**Lemma 1** *For any $\varepsilon > 0$, there exists $T$ s.t. for any $t \geq T$, $\tilde{U}_i^t \geq U_i^t - \varepsilon$.*

**Proof.** Note that:

$$\begin{aligned}
\tilde{U}_i^t = g_i(s_i^t, \mu_i^t) &\geq g_i(s_i^{t-1}, \mu_i^t) \\
&= \frac{1}{t+1} g_i(s_i^{t-1}, s_i^{t-1}) + \frac{t}{t+1} g_i(s_i^{t-1}, \mu_i^{t-1}) \\
&= \frac{1}{t+1} g_i(s_i^{t-1}, s_i^{t-1}) + \frac{t}{t+1} \tilde{U}_i^{t-1}
\end{aligned}$$

Expanding $\tilde{U}_i^{t-1}$ in the same way:

$$\tilde{U}_i^t \geq \frac{1}{t+1} g_i(s_i^{t-1}, s_i^{t-1}) + \frac{1}{t+1} g_i(s_i^{t-2}, s_i^{t-2}) + \frac{t-1}{t+1} \tilde{U}_i^{t-2}$$

and iterating:

$$\tilde{U}_i^t \geq \frac{1}{t+1}\sum_{\tau=0}^{t-1} g_i(s_i^\tau, s_{-i}^\tau) + \frac{1}{t+1} g_i(s_i^0, \mu_i^0)$$

Define $T$ such that

$$\varepsilon > \frac{1}{T+1} \max_s g_i(s)$$

and we're done. *Q.E.D.*

The Lemma says that if the game goes on long enough, expected payoffs must ultimately be almost as big as time-average payoffs. Of course, they can in principle be a lot bigger. In the mis-coordination example above, expect payoffs converge to 1/2. But actual payoffs are always zero.

We can use the Lemma to prove Shapley's result.

**Proof of Shapley Non-convergence.** In the Rock-Scissors-Paper game, there is a unique Nash equilibrium with expected payoffs of 1/3 for both

players. Therefore, if fictitious play converged, then ultimately $\tilde{U}_i^t \to 1/3$, meaning that $\tilde{U}_1^t + \tilde{U}_2^t \to 1/3 + 1/3 = 2/3$. But under fictitious play, the empirical payoffs *always* sum to 1, so $U_1^t + U_2^t = 1$ for all $t$. This contradicts the Lemma, meaning fictitious play can't converge. *Q.E.D.*

Shapley's example highlights a few key points about fictitious play. One is that because it jumps around, it is not particularly well-suited for learning mixed equilibria. Another is that because players only are thinking about their opponent's actions, they're not paying attention to whether they've actually been doing well.

More recent work on belief-based learning has tried to get around these difficulties in various ways. A good reference for this material is the book by Fudenberg and Levine (1999). Interesting papers include Fudenberg and Kreps (1993, 1995) on smoothed fictitious play, Kalai and Lehrer (1993), and also Nachbar (2003) on Bayesian learning, and Foster and Vohra (1998) on calibrated learning rules. I'll have something to say about this literature in class.

## 2 Reinforcement Learning

A different style of learning model derives from psychology and builds on the idea that people will tend to use strategies that have worked well in the past. These adaptive learning models do not incorporate beliefs about opponent's strategies or require players to have a 'model' of the game. Instead player respond to positive or negative stimula. References include Bush and Mosteller (1955) and Borgers and Sarin (1996).

The following model is based on Erev and Roth (1998). Let $q_{ik}(t)$ denote player $i$'s propensity to play his $k$th pure strategy at time $t$. Initially, player $i$ has the uniform propensities across strategies:

$$q_{i1}(1) = q_{i2}(1) = ... = q_{iK}(1)$$

After each period, propensities are updated using a *reinforcement* function. Suppose that at time $t$, player $i$ played strategy $k_t$ and obtained a payoff $x$. Then,

$$q_{ik}(t+1) = \begin{cases} q_{ik}(t) + R(x) & \text{if } k = k_t \\ q_{ik}(t) & \text{otherwise} \end{cases},$$

for some increasing function $R(\cdot)$. The idea is that if $k_t$ was successful, the player is more likely to use it again. If it was unsuccessful, he will be less likely to use it.

Propensities are mapped into choices using a choice rule. For instance, letting $p_{ik}(t)$ denote the probability that $i$ will choose $k$ at time $t$, a simple rule would be:

$$p_{ik}(t) = \frac{q_{ik}(t)}{\sum_j q_{ij}(t)}.$$

While this sort of model is very simple, it can sometimes do very well explaining experimental results. Not surprisingly, these models tend to fit the data better with more free parameters.

## References


[1] Borgers, Tilman and Rajiv Sarin (1997) "Learning Through Reinforcement and Replicator Dynamics," *J. Econ. Theory*, 77, 1–14.

[2] Brown, George (1951) "Iterative Solutions of Games by Fictitious Play," in *Activity Analysis of Production and Allocation*, New York: John Wiley & Sons.

[3] Bush, Robert and Frederick Mosteller (1955) *Stochastic Models for Learning.*

[4] Crawford, Vincent (1997) "Theory and Experiment in the Analysis of Strategic Interaction," in *Advances in Economics and Econometrics: Theory and Applications*, Seventh World Congress of the Econometric Society, ed. D. Kreps and K. Wallis.

[5] Erev, Ido and Alvin Roth (1998) "Predicting How Play Games: Reinforcement Learning in Experimental Games with Unique Mixed-Strategy Equilibria," *Amer. Econ. Rev.,* 88, 848–881.

[6] Foster, Dean and R. Vohra (1998) "Asymptotic Calibration," *Biometrika.*

[7] Fudenberg, Drew and David Kreps (1993) "Learning Mixed Equilibria," *Games and Econ. Behav.*, 5, 320–367.

[8] Fudenberg, Drew and David Kreps (1995) "Learning in Extensive Form Games I: Self-Confirming Equilibria," *Games and Econ. Behav.*, 8, 20–55.

[9] Fudenberg, Drew and David Levine (1993) "Self-Confirming Equilibrium," *Econometrica*, 61, 523–546.

[10] Fudenberg, Drew and David Levine (1999) *The Theory of Learning in Games,* Cambridge: MIT Press.

[11] Kalai, Ehud and Ehud Lehrer (1993) "Rational Learning Leads to Nash Equilibria," *Econometrica.*

[12] Kalai, Ehud and Ehud Lehrer (1993) "Subjective Equilibrium in Repeated Games," *Econometrica.*

[13] Nachbar, John (2003) "Beliefs in Repeated Games," Working Paper, Washingon University.

[14] Robinson, Julia (1951) "An Iterative Method of Solving a Game," *Annals of Mathematical Statistics,* 54, 296–301.