Learning Nash Equilibrium

Dean P. Foster
H. Peyton Young

SFI WORKING PAPER: 2003-02-007

SANTA FE INSTITUTE

Learning Nash Equilibrium*

DEAN P. FOSTER
Department of Statistics, Wharton School, University of Pennsylvania

H. PEYTON YOUNG
Department of Economics, Johns Hopkins University
Center on Social and Economic Dynamics, The Brookings Institution
The Santa Fe Institute

January 24, 2003

*Acknowledgements. We are indebted to seminar participants at the Santa Fe Institute, University of Chicago, Northwestern University, Hebrew University, the University of Siena, Stanford, the Stockholm School of Economics, University College London and the NSF/NBER Decentralization Conference at Georgetown University. The second author's research was supported by NSF grant 9818975.
versionVersion Id: easyNash.tex,v 1.41 2003/01/17 12:23:39 foster Exp
32k/47k allowed.

Abstract

While the concept of Nash equilibrium is central to economic theory, to date no one has provided a credible model of how economic agents come to play such equilibria. In particular, there is no model of how agents acting in their own self-interest — without any form of coordination — can learn equilibrium behavior starting from out-of-equilibrium conditions. This paper proposes a simple learning process with this property. Agents learn about their opponents by periodically testing and rejecting alternative hypotheses about their behavior. If they use sufficiently powerful tests and their responses are sufficiently close to being optimal, this process eventually leads them to play close to equilibrium a large proportion of the time.

1 Nash equilibrium

Over fifty years ago, John Nash established the existence of an equilibrium concept that has since become one of the cornerstones of economic theory (1). The idea is simple: a system of interacting agents is in equilibrium if each individual's strategy maximizes her expected utility given the strategies of the others. The existence of equilibrium strategies follows from standard assumptions about payoff functions and Kakutani's fixed point theorem (1,2). To date, however, no one has given a satisfactory explanation of what *dynamical process* actually yields Nash equilibria when the system begins in out-of-equilibrium conditions. While there have been a number of attempts to solve this problem, every learning method in the literature that we are aware of either fails to converge in some situations, converges to something other than a Nash equilibrium, or relies on *ex ante* information about the opponents' payoffs that the players are unlikely to know in practice.

In this paper we propose a simple learning process that leads to Nash equilibria and *only* Nash equilibria in any finite game. The idea is based on statistical hypothesis testing: players learn about their opponents by testing and rejecting alternative hypotheses about their behavior. If they use sufficiently powerful tests and are sufficiently rational in their responses (but not perfectly rational), then eventually they will play close to Nash equilibrium — in fact they will eventually play close to subgame perfect equilibrium, which is considerable more demanding (3). This conclusion holds no matter how far from equilibrium the process is to begin with. Furthermore the method is *robust* in the sense that it presupposes no prior information about the opponents' payoffs or their intended strategies.

2 Fictitious play

To fix ideas we shall first consider one of the standard learning algorithms in the game theory literature — fictitious play (4, 5) — and show why it does not solve the learning problem except under very special circumstances. (The reader familiar with the literature on fictitious play may wish to skip

ahead to section 3.)

Consider a two person game that is played infinitely often. At each point in time, each player computes the empirical frequency with which her opponent has played each action in the past. She assumes that the opponent will use these same frequencies in every future period, and chooses an action with highest expected payoff given this assumption. After each play the empirical frequencies are recomputed, and the process is repeated. (The choice of actions in the first period is arbitrary.)

To illustrate, consider the following simple coordination game. Two players are driving towards each other, and each has the choice of staying to the left (L) or to the right (R). If they coordinate each gets a payoff of 1, if they miscoordinate they crash and each gets a payoff of 0:

		Player 2	
		L	R
Player 1	L	1, 1	0, 0
	R	0, 0	1, 1

In a single period of play this game has three Nash equilibria: i) both choose R; ii) both choose L; iii) both choose randomly with equal odds on R and L. The first two are *coordination equilibria*, while the third is a *mixed equilibrium*. Notice that the coordination equilibria yield a payoff of 1 to each player, whereas the mixed equilibrium yields an expected payoff of .5 and is therefore inefficient.

In the absence of communication it is not clear which equilibrium will be played, or in fact whether any equilibrium will be played. Suppose they use fictitious play as a learning rule. Suppose also that, in the first period, they happen to miscoordinate. Then in the second period they will miscoordinate again, because each thinks the other is going to continue to do what she did in the first period. This leads to a situation such as the following:

Period	1	2	3	...
Player 1	L	R	?	
Player 2	R	L	?	

In the third period each believes, based on the empirical distribution in periods 1 and 2, that the opponent will play L or R with equal probability. Given these beliefs, both L and R have an expected payoff of zero, hence any probability mixture over L and R would be a rational response in the third period. A particularly natural choice is to play L or R with equal probability. When both players do this, they coordinate with probability one-half. Say they coordinate on (R, R). Then in the fourth period each thinks that the other is more likely to play R than to play L, so they both play R again. This continues in every subsequent period, that is, their behaviors have converged to playing equilibrium (R, R). Similarly if they happen to coordinate on L in the third period, their behaviors converge on equilibrium (L, L).

But what if they miscoordinate in the third period? Then they will miscoordinate again in the fourth period, but in the fifth period there will be a fifty-fifty chance that they coordinate, in which case they continue to coordinate from then on. Extending this reasoning we find that the probability is one that eventually they coordinate on R or on L, assuming they randomize fifty-fifty between R and L whenever they are indifferent. In this case fictitious play solves the learning problem in the sense the behaviors converge with probability one to a Nash equilibrium.

But this does not hold in other situations. Consider the following example:

		Player 2	
		L	R
Player 1	L	$\sqrt{\frac{1}{2}}, 1-\sqrt{\frac{1}{2}}$	0, 1
	R	0, 1	1, 0

Here player 1 wants to imitate player 2, but player 2 wants to do the opposite of what player 1 is doing. In a single period of play, this game has a unique equilibrium in which each player chooses L with probability $.58578\ldots = 1/(1 + 1/\sqrt{2})$. No matter how the players start out, there is never a time at which they are indifferent between playing L and R, due to the irrational payoff. Thus, each player chooses L for sure or R for sure in

every period. But this means that neither comes close to playing the unique equilibrium no matter how long the process continues.

It is important to note here that, while the *period-by-period behavior* of the players does not converge to equilibrium, the *time-average behavior* converges to the mixed equilibrium. That is, the empirical frequency with which each player chooses L converges to $1/(1 + 1/\sqrt{2})$. Thus, while the players themselves are not learning Nash equilibrium behavior, it is true that the average behavior of the system converges to equilibrium. (It can be shown that fictitious play has this property for almost all games in which each player has just two strategies (6,7)).

3 Notions of convergence

The preceding example illustrates that a learning process may converge to equilibrium in two different senses. We now formalize this observation. Let G be a two person game in which player 1 (the row player) has m strategies, $1 \leq i \leq m$, and player 2 (the column player) has n strategies, $1 \leq j \leq n$. When player 1 chooses action i and player 2 chooses action j, the payoffs are a_{ij} and b_{ij} to players 1 and 2 respectively. Suppose that player 1 chooses action i with probability p_i, where $\sum_i p_i = 1$, and player 2 chooses action j with probability q_j, where $\sum_j q_j = 1$. Then the expected payoffs are $\alpha(p,q) = \sum\sum a_{ij}p_iq_j$ for player 1 and $\beta(p,q) = \sum\sum b_{ij}p_iq_j$ for player 2. The pair of distributions (p,q) is a *Nash equilibrium* if p maximizes $\alpha(p,q)$ given q, and q maximizes $\beta(p,q)$ given p.

Suppose now that the players change their behaviors over time according to some learning rule (not necessarily fictitious play). Let p^t denote the probability distribution that describes how player 1 chooses her action in period t (given the history to date), and let q^t be the analogous distribution for player 2. Let T be a positive integer. The expected proportion of times that action i will be played in the first T periods is $\overline{p}_i^T = \sum_{t \leq T} p_i^t/T$. Similarly, the expected proportion of times with which action j is played is $\overline{q}_j^T = \sum_{t \leq T} q_j^t/T$. (As T becomes large, the expected proportions come arbitrarily close to the actual proportions with probability one.)

Let $\overline{p}^T \equiv (\overline{p}_1^T, \ldots, \overline{p}_m^T)$ and $\overline{q}^T \equiv (\overline{p}_1^T, \ldots, \overline{p}_n^T)$. We say that the learning process *converges in time average to equilibrium* if the sequence $(\overline{p}^T, \overline{q}^T)$ comes close to some Nash equilibrium with high probability when T is sufficiently large. More formally, for every small $\epsilon > 0$, there is a time T_ϵ such that, for all $T \geq T_\epsilon$, the Euclidean distance between $(\overline{p}^T, \overline{q}^T)$ and some Nash equilibrium is less than ϵ with probability at least $1 - \epsilon$.

This is a relatively weak form of convergence to equilibrium. A stronger and more satisfactory definition is the following: the learning process *converges in behaviors to equilibrium* if the sequence of strategies (p^t, q^t) comes close to equilibrium with high probability. More formally, for every small $\epsilon > 0$, there is a time t_ϵ such that, for all $t \geq t_\epsilon$, the Euclidean distance between (p^t, q^t) and some Nash equilibrium is less than ϵ with probability at least $1 - \epsilon$. (For games with a unique Nash equilibrium it is straightforward to show that convergence in behaviors implies convergence in time average. More generally it implies convergence in time average to the convex hull of Nash equilibria.)

We have just seen that fictitious play converges in behaviors for the left-right driving game and hence it also converges in time average. However, the example with irrational payoffs shows that fictitious play does not necessarily converge in behaviors, while still converging in time averages. It turns out that this second behavior is more typical for games with two actions. Indeed, every two-person game in which each player has just two actions converges in time averages provided that ties are resolved in a fixed way (6, 7).

If there are more than two strategies, however, fictitious play may no longer converge even in time averages. The following standard example is due to Shapley (8):

		Player 2		
		Red	Yellow	Blue
Player 1	Red	1, 0	0, 0	0, 1
	Yellow	0, 1	1, 0	0, 0
	Blue	0, 0	0, 1	1, 0

We can interpret this as a fashion game in which player 2 is a fashion

leader and player 1 is a fashion follower: 1 wants to wear whatever 2 is wearing, but 2 wants to *avoid* what 1 is wearing. It can be shown that, starting from any pair of choices in period 1, fictitious play leads to cyclic behavior in which fashion rotates from Red to Blue to Yellow to Red to Blue and so forth. A particular sequence generated by this process is shown below:

Period	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Player 1	R	R	B	B	B	B	B	Y	Y	Y	Y	Y	Y	Y
Player 2	R	B	B	B	Y	Y	Y	Y	Y	Y	Y	Y	Y	R

It can be shown that each cycle is exponentially longer than the previous one, hence the time average of Red, Blue, and Yellow also cycles. But if the players are myopic (not forward-looking) the only Nash equilibrium of the game is for each player to choose each color with probability one-third in each period. Hence neither the time average nor the period-by-period behaviors come close to equilibrium (see figure 1).

4 Lack of information

Before describing our new learning procedure, we need to be clear about what kinds of information the players have at their disposal. When we write down a game (e.g., the three examples given so far) we specify the payoffs of each player. This does not mean, however, that in an actual game a player could be likely to know the payoff numbers of her opponent. Consider the following game with goods as prizes:

The soda game

	L	R
L	Coke, Coke	Sprite, Seven-up
R	Seven-up, Sprite	Pepsi, Pepsi

Notice that without more information about the players' preferences, we do not even know the number of Nash equilibria in this game. If both players like dark drinks better than light drinks (or the other way around), then it

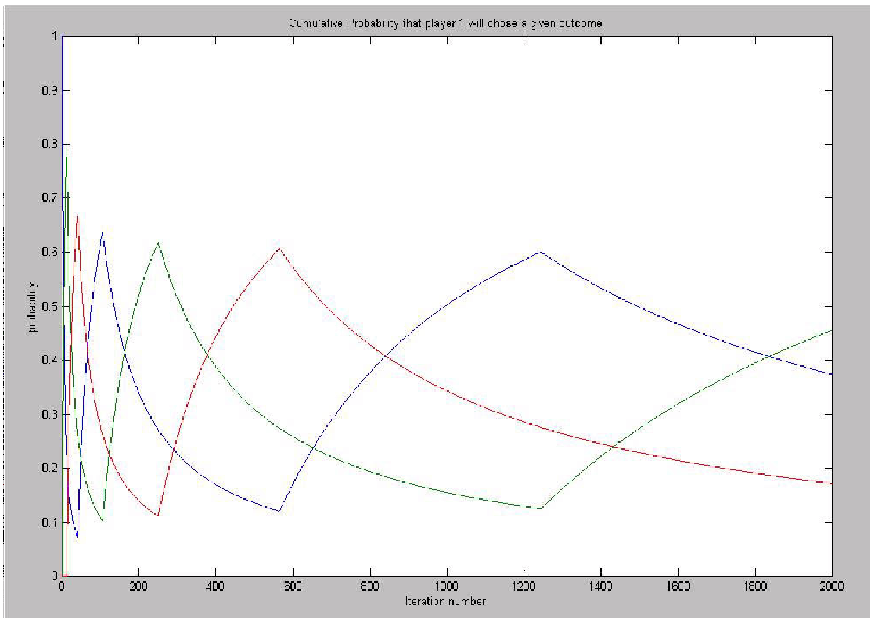

Figure 1: Frequency of each action in a simulation of fictitious play for the Shapley fashion game. The length of each successive cycle grows exponentially longer.

is a coordination game with three equilibria (two pure and one mixed). But if one player prefers dark to light and the other prefers light to dark, then it has a single Nash equilibrium that involves randomizing between L and R.

In the game theory literature, it is common to treat the problem of incomplete information as follows. The payoffs to a player are determined by the realization of a random variable which is called the *type* of the player. It is assumed that the distribution of types is known to both players but the realization is known only for each player's own payoff. These are known as *games of incomplete information* (9). One can now define a *Bayesian Nash equilibrium* to be a strategy that, at each stage of the game, maximizes expected future payoff conditional on the ex ante knowledge of the opponent's payoff distribution together with any additional information revealed by play up to that stage. Unfortunately this does not solve the problem of learning equilibrium, because there are often many Bayesian Nash equilibria; it merely pushes the learning problem onto another level.

An alternative approach is to assume that the players have prior information about their opponents' possible *strategies*, rather than their payoffs. In fact, if each has enough information that she attaches positive probability to the strategy the opponent actually intends to use (the *grain of truth* assumption), then ordinary Bayesian updating leads them to learn these strategies, and their behaviors come close to equilibrium with probability one (10). This is a very demanding assumption, however, because the set of possible strategies of an opponent is uncountably large. Indeed it can be shown that, under quite reasonable conditions, it would be very unlikely that players have the necessary information at their disposal for the grain of truth assumption to hold (11, 12, 13, 14, 15, 16).

In this paper we demonstrate a simple learning rule that leads to equilibrium behavior without assuming any coordination by the players or any *ex ante* knowledge of the opponent's payoffs or their distribution. That is, unlike the Bayesian approaches discussed earlier, out learning procedure is *robust* in the sense that it uses as input only the player's *own* utility function, together with the actions taken by the opponent up to that time. This learning process will have the property that, for all realizations of the pay-

offs, the behaviors will be within ϵ of a Nash equilibrium at least $1-\epsilon$ of the time.

5 A new learning rule

For simplicity of exposition we shall state the approach in terms of two-person games; it extends to n-person games with no difficulty. The learning process has the following three-part structure, which is based on classical statistical hypothesis testing:

1. At each time a player has a model or hypothesis about the future behavior of her opponent, that is, a conjectured probability distribution over the opponent's future actions.

2. From time to time the player compares her model with recent past data. If the data is unlikely to have occurred given that the model is true (according to some hypothesis test), the model is jettisoned and a new model is selected; otherwise the model is retained.

3. At all times the player chooses a smoothed best response to her current model, that is, she almost optimizes given her current forecast.

We now show how this process works for a particularly simple class of hypothesis tests; the more general case is treated in (16). We shall describe the algorithm from the perspective of player one: an analogous description holds for player two. Suppose we are playing a 2×2 game like the Left-Right driving game. Both players have two actions called 1 and 2. Player one will play action 1 with probability p and player two will play 1 with probability q. Player 1 *thinks* that player two will play action 1 with probability $\hat{q}$ ($\hat{q}$ is player one's *model* of player two). Given $\hat{q}$, player one computes p in order to expect to obtain a high payoff. Assume that player one is myopic, that is, he cares only about expected payoffs in the current period. (The framework extends to forward-looking players as we show in (16).)

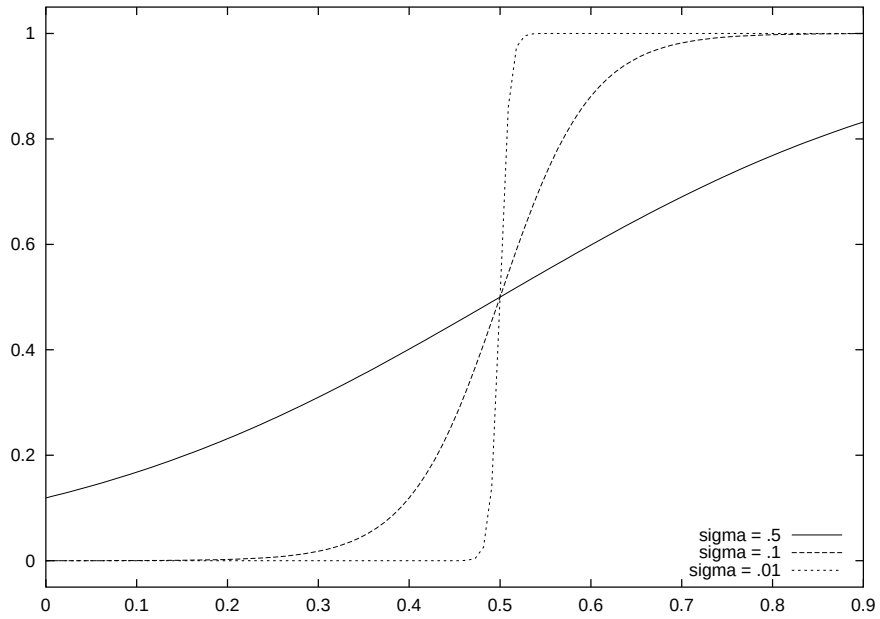

Figure 2: The function $p = R^{\sigma}(\hat{q})$ relates player 1's model ($\hat{q}$) to his probability (p) of playing L, where the underlying utilities are as specified in the left-right driving game. As σ approaches zero, the reply functions look more and more like the pure best reply function.

Assume that p is determined from $\hat{q}$ by the following *quantal response function* (17):

$$p = R^{\sigma}(\hat{q}) = \frac{e^{(\hat{q}a_{11}+(1-\hat{q})a_{12})/\sigma}}{e^{(\hat{q}a_{11}+(1-\hat{q})a_{12})/\sigma} + e^{(\hat{q}a_{21}+(1-\hat{q})a_{22})/\sigma}}$$

where a_{ij} is the payoff to player one if he plays i and player two plays j. See figure 2 for a plot of this function.

Let $U_{\hat{q}}(1) \equiv \hat{q}a_{11}+(1-\hat{q})a_{12}$ and $U_{\hat{q}}(2) \equiv \hat{q}a_{21}+(1-\hat{q})a_{22}$ be the utilities from player actions 1 and 2 respectively given one's belief that player two is using the distribution $\hat{q}$. Then the response function can be written in a more suggestive fashion as follows:

$$p = R^{\sigma}(\hat{q}) = \frac{e^{U_{\hat{q}}(1)/\sigma}}{e^{U_{\hat{q}}(1)/\sigma} + e^{U_{\hat{q}}(2)/\sigma}}.$$

The sharpness of the response depends on the size of σ. When σ is small, player one plays an optimal response with high probability for most values of $\hat{q}$. When σ is large, he chooses the two actions with nearly equal probability. The idea is that players' responses are modified by small trembles or unobservable utility shocks, an assumption that comports with a fair amount of experimental evidence (18, 19, 20, 21).

Figure 2 shows plots of R^{σ} for three values of σ. When σ is small (e.g., $\sigma = .01$) the player plays left with high probability if his model of the opponent is that she is going to play left with probability greater than one-half (i.e. $\hat{q} > .5 + O(\sigma)$). Similarly he plays Right with high probability if his model says the opponent is going to play Right with probability greater than one-half.

In general, when player one has m actions and player two has n actions, p and $\hat{q}$ are vectors. That is $p = (p_1, \ldots, p_m)$ where p_i is the probability with which player one takes action i. Similarly, $\hat{q} = (\hat{q}_1, \ldots, \hat{q}_n)$ where $\hat{q}_j$ is the probability with which player one *thinks* player two is going to choose action j. The quantal response function in this case is

$$p_i = R_i^{\sigma}(\hat{q}) = \frac{e^{U_{\hat{q}}(i)/\sigma}}{\sum_x e^{U_{\hat{q}}(x)/\sigma}} \tag{1}$$

where $U_{\hat{q}}(i) = \sum_j a_{ij}\hat{q}_j$ is the expected utility to player one when he plays i and he thinks player two is using $\hat{q}$. (To avoid confusion with Bayesian terminology, we have used the word "model" instead of "belief" to describe $\hat{q}$.)

To make the absolute size of the smoothing parameter σ meaningful, we will assume that the payoffs are all between zero and one. When σ is large, we know the actions chosen are a very smooth function of the models, when σ is small the response function is close to knife-edge behavior. Plots of R^σ for various values of σ are shown in figure 2.

The hypothesis testing rule compares the observed frequency of play over s rounds with the current model $\hat{q}$. If this difference is larger than a prespecified tolerance τ, the hypothesis that the other player is playing $\hat{q}$ is rejected. If the hypothesis is rejected a new hypothesis is chosen; otherwise the current hypothesis is maintained.

Definition 1 (The learning parameters) *The parameters σ, τ and s are called the* learning parameters. *The size of σ determines the closeness to optimality of the response, that is, the utility is close to the utility generated by a best reply. The size of τ determines the fineness of the test, that is, how accurate the hypothesis must be in relation to the observed data in order to accept the hypothesis. Finally, s is the amount of data on which tests are based. A large value of s combined with a small value of τ yields a powerful and accurate test.*

We can now describe the algorithm in terms of the learning parameters (which may be different for different players):

1. Initialization: Pick the initial hypothesis $\hat{q}$ uniformly at random.
2. Play: For each round of play use the distribution $p = R^\sigma(\hat{q})$ to randomly pick an action in the current round of play.
3. Test:
   (a) If not currently collecting data, start to collect it with probability $\frac{1}{s}$ and continue collecting it for the next s rounds of play.

(b) If s rounds of data have just been collected, stop collecting data and compute the empirical distribution of the opponents' actions, $\overline{q}$. If $|\hat{q} - \overline{q}| > \tau$, reject the current hypothesis and pick a new hypothesis $\hat{q}$ uniformly at random (i.e. go to step 1). If the test is not rejected leave $\hat{q}$ as it was (i.e. go to step 2).

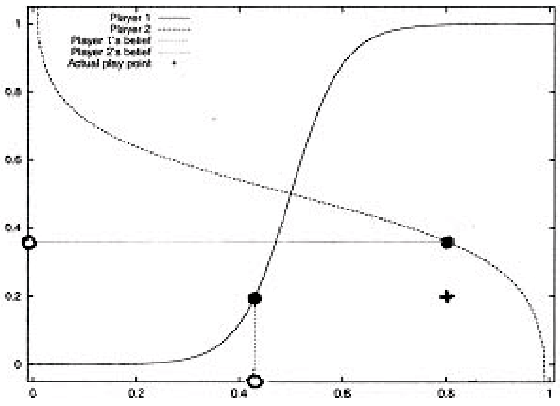

Figure 3: Response functions for the soda game, with models and responses. The true model (indicated by a cross) is not close to either hypothesis (indicated by solid dots).

We illustrate the situation in Figure 3. At this particular point in time, player one has model $\hat{q} = .43$ and player two has model $\hat{p} = .38$ (these are indicated in the figure by hollow dots). The actual probabilities of choosing action 1 for each player are given by $R(\hat{q}) = .2$ and $R(\hat{p}) = .8$. For each player, the combination of the model and the play probability is a point on that player's response curve (indicated by a solid dot). These

are the probability distributions that each player *thinks* is governing the process. The real probability distribution (the true model) is determined by the combination of play probabilities, that is, by the point labeled with a cross in the figure. Since the dots do not lie near the cross, the players are likely to reject their hypotheses, and typically the true model will shift also. Only when the dots both lie close to the cross are the hypotheses likely to be accepted. This can only happen if the dots and the cross all lie close to the crossing point of the two response curves. When the response curves are close to being optimal (i.e., when σ is small), this means that play is close to being a Nash equilibrium.

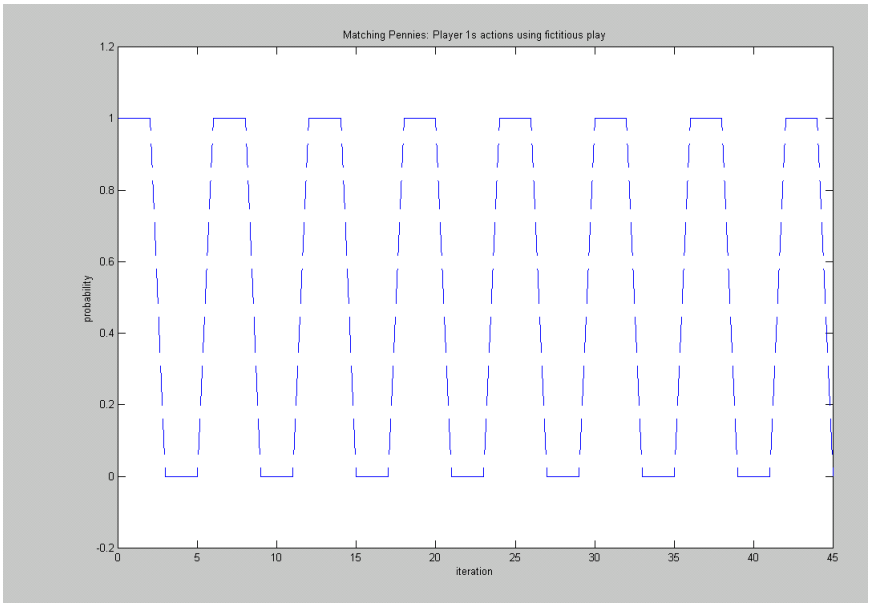

Figure 4: The learning process illustrated for the driving game. The learning parameters for both players are $\sigma = .5$, $\tau = .025$ and $s = 1000$. Sometimes play is close to the drive-right equilibrium, while at other times it is close to the drive-left equilibrium. Note the sharp transitions between the two regimes. The interior equilibrium occurs infrequently enough that it does not appear in this run.

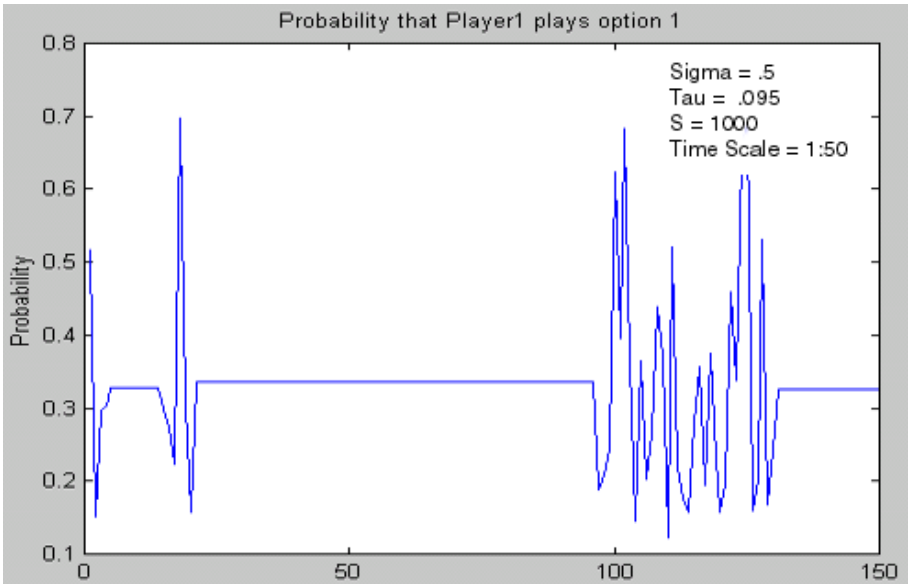

Figure 5: Shapley game simulation for 150,000 rounds of play with $\sigma = .5$, $\tau = .095$, and $s = 1000$. Note the long periods of stability and short bursts of searching. (The graph shows player 1's probability of choosing Red in each period; the behavior of the other players exhibits the same phases of stability and searching.) In the stable periods play is close to the unique Nash equilibrium in which each player chooses each action with probability 1/3.

We illustrate the qualitative behavior of this algorithm for the driving game (figure 4) and the fashion game (figure 5). Note that the process goes though two phases of behavior: search and stable play. During the search phase, both players frequently change their hypotheses and responses. This continues until they hit on a lucky combination in which both of their hypotheses are close to the actual play distribution of the other player. In this situation the behaviors are close to being a Nash equilibrium. During the ensuing stable phase, both players keep testing their hypotheses and do not reject them. A stable phase typically lasts much longer than a search phase.

Eventually, one of the players rejects her model even though it is reasonably accurate (one might say a type I error has occurred). This destabilizes the "near equilibrium" phase and another search phase is entered. After this new search settles into stable play, the process may be close to the same Nash equilibrium as before, or it may be close to a new Nash equilibrium. From the figures, we see that the fraction of time that the process is in a stable phase (and hence close to equilibrium behavior) much higher than the fraction of time that it is in a search phase. Establishing this fact proves the following theorem (see 16 for details):

Theorem 1 *For any finite game G, and any $\epsilon > 0$, there is a range of values of the learning parameters such that hypothesis testing yields behaviors that are within ϵ of a Nash equilibrium of G at least $1 - \epsilon$ of the time.*

6 Extensions of the result

An important feature of this process is that the choice of new hypothesis can be carried out in many different ways and behaviors will still converge to equilibrium (though the rate of convergence may be affected). In particular, when hypotheses are rejected, the players could use quite sophisticated techniques to choose new hypotheses that take into account information revealed by play so far. This added sophistication does not necessarily lead to better convergence properties; in fact it may promote cycling, as in fictitious play. It can be shown, however, that if the space of possible hypotheses is

searched completely with positive probability after each rejection (in a sense made precise in (16)), then behaviors eventually come close to equilibrium and stay close a large proportion of the time.

Another extension involves players who take account of the repeated nature of the game. For example, some players might be interested in payoffs from future play in addition to play in the current round, and thus they may be motivated to play very sophisticated, forward-looking strategies. Other players might play complex strategies that condition on previous behavior, such as tit-for-tat. Theorem 1 can be extended to take both of these possibilities into account (16). In fact it can be stated in a form that guarantees that behavior comes close to subgame perfect equilibrium, which is considerably more demanding than Nash equilibrium.

7 Discussion

It may seem somewhat odd that such a straightforward approach to learning has been overlooked to date in the game theory literature. After all, hypothesis testing is routinely applied to statistical estimation problems: why not use it to learn the strategy of an opponent? In this case a hypothesis would amount to a "guess" about the probabilities governing the opponent's choices. A hypothesis testing procedure involves subjecting successive batches of data to tests, rejecting those that do not fit the data reasonably closely. This approach would clearly work if the opponent's strategy were the same in each period, but in fact it keeps changing as the opponent learns. In other words, the learning process induces a feedback loop in which the thing being learned reacts to the players' attempts to learn it. The substance of our result is that, even in this highly non-stationary environment where multiple agents are learning about each other, the hypothesis testing approach can be successfully applied without any special assumptions about the structure of the game or the information available to the players.

References

1. Nash, John (1950). "Equilibrium points in n-person games," *Proceedings of the National Academy of Sciences of the USA*, **36**: 48-49.

2. Nash, John (1951): "Non-cooperative Games." *Annals of Mathematics* **54**, 289-295.

3. Selten, Reinhard (1975) "A Re-examination of the perfectness concept for equilibrium points in extensive games." *International Journal of Game Theory* **4**, 25-55.

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

5. Robinson, J. (1951): "An iterative method of solving a game," *Annals of Mathematics*, **54**, 296 - 301.

6. Miyasawa, K. (1961): "On the Convergence of the Learning Process in a 2 x 2 Non-Zero-Sum Two-Person Game," Economic Research Program, Research Memorandum no. 33, Princeton University, Princeton NJ.

7. Monderer, Dov, and Loyd Shapley (1996): "Fictitious Play Property for Games with Identical Interests," *Journal of Economic Theory* 68, 258-265.

8. Shapley, L. S. (1964): "Some topics in two-person games," in *Advances in Game Theory* M. Dresher, L. S. Shapley, and A. W. Tucker, eds., pp. 1 - 28, Princeton, NJ: Princeton Univ. Press.

9. Harsanyi, John (1968) "Games with incomplete information played by Bayesian players." *Management Science* **14**, 159-182; 320-384; 486-502.

10. Kalai, Ehud, and Ehud Lehrer (1993): "Rational Learning Leads to Nash Equilibrium," *Econometrica*, **61**, 1019-1045.

11. Jordan, James S. (1993): "Three Problems in Learning Mixed-Strategy Equilibria," *Games and Economic Behavior*, **5**, 368-386.

12. ———— (1995): "Bayesian Learning in Repeated Games," *Games and Economic Behavior*, **9**, 8-20.

13. Nachbar, John. H. (1997): "Prediction, Optimization, and Learning in Games," *Econometrica*, **65**, 275-309.

14. ———————— (2001): "Bayesian Learning in Repeated Games of Incomplete Information," *Social Choice and Welfare*, **18**, 303-326.

15. Foster, Dean P. and H. Peyton Young (2001) "On the impossibility of predicting the behavior of rational agents." *Proceedings of the National Academy of Sciences of the USA*, **98**, 12848-12853.

16. Foster, Dean P. and H. Peyton Young (2001): "Learning, Hypothesis Testing, and Nash Equilibrium," Santa Fe Institute Working Paper.

17. McKelvey, Richard and Thomas Palfrey (1995) "Quantal response equilibria for normal form games," *Games and Economic Behavior*, **7**, 6 - 38.

18. ———— (1992) "An experimental Study of the Centipede Game," *Econometrica*, **60**, 803 - 836.

19. Banks, J. Camerer, C. and Porter, D. (1994) "Experimental tests of nash refinements in signaling games," *Games and Economic Behavior*, **6**, 1 - 31.

20. Brandts, J. and Holt, C. A. (1993) "Adjustment patterns and equilibrium selection in experimental signaling games," *International Journal of Game Theory*, **22**, 279 - 302.

21. Schotter, A., Weigelt, K., and Wilson, C. (1994) "A laboratory investigation of multiperson rationality and presentation effects," *Games and Economic Behavior*, **6**, 445 - 468.