

# A Simple Test of Learning Theory\*

Jim Engle-Warnick<sup>†</sup>

Department of Economics

McGill University

Montreal, Qué H3A 2T7, Canada

Ed Hopkins<sup>‡</sup>

Edinburgh School of Economics

University of Edinburgh

Edinburgh EH8 9JY, UK

August, 2004

Provisional

## Abstract

We report experiments designed to test the theoretical possibility, first discovered by Shapley (1964), that in some games learning fails to converge to any equilibrium, either in terms of marginal frequencies or of average play. Subjects played repeatedly in fixed pairings one of two  $3 \times 3$  games, each having a unique Nash equilibrium in mixed strategies. The equilibrium of one game is predicted to be stable under learning, the other unstable. We find that in the experiments, average play is close to equilibrium in both cases. However, stronger cycles seem to be present in the data from the unstable treatment.

---

\*We thank Steffen Huck and Martin Sefton for helpful discussions. We thank Sara Constantino and Julie Héroux for research assistance. We gratefully acknowledge financial support from the Economic and Social Research Council, UK, the Nuffield Foundation, and Nuffield College, University of Oxford, UK. We thank the Economics Department, Royal Holloway, University of London and the Centre for Interuniversity Research and Analysis on Organizations (CIRANO) and the Bell University Laboratory in Electronic Commerce and Experimental Economy for use of the experimental laboratory. Errors remain our own.

<sup>†</sup>[jim.enge-warnick@staff.mcgill.ca](mailto:jim.enge-warnick@staff.mcgill.ca); <http://www.mcgill.ca/economics/faculty/enge-warnick/>

<sup>‡</sup>[E.Hopkins@ed.ac.uk](mailto:E.Hopkins@ed.ac.uk), <http://www.homepages.ed.ac.uk/ehk>

# 1 Introduction

Recent work on learning is marked by the integration of theory and experimental investigation. There has been significant recent success in explaining observed laboratory decision-making through the use of learning theories (for example, Camerer and Ho, 1999; Erev and Roth, 1998). This has spurred theorists to do new work that advances in our understanding of how learning models work. This field is surveyed in Fudenberg and Levine (1998). However, there have been fewer experiments designed explicitly to test learning theory's predictions.<sup>1</sup>

In this paper, we examine one particularly striking implication of learning: there are games in which play should never converge. We investigate experimentally two games that possess a unique Nash equilibrium in mixed strategies. In one game, learning theory predicts convergence to the mixed strategy equilibrium. In the other game, theory predicts that the equilibrium is unstable. That is, if the learning model we consider, stochastic fictitious play, accurately predict the behavior of experimental subjects, their play should never settle down but rather should continue to cycle. Our experiments reveal limited support for these hypotheses. In the both treatments, average play seems to converge to close to Nash equilibrium. However, cycles are detectable in the data and are larger in the unstable treatment than in the stable.

Hofbauer and Hopkins (2004b) recently have analyzed a variant of stochastic fictitious play where agents place greater weight on more recent experience. This new research suggests the difference between divergence from and convergence to equilibrium may not be as clear cut as with classical assumptions on players' beliefs. Specifically, average behaviour always converges, even when the only equilibrium is unstable. It is only players' actual mixed strategies, which are not directly observable in experiments, that are predicted to diverge in the unstable case.<sup>2</sup> Thus, we should not be too surprised if average behaviour is close to equilibrium in both stable and unstable treatments. We will have to look more closely to distinguish differences between them.

These results are not of purely theoretical interest. They also address important and timely questions about the phenomenon of price dispersion. Price dispersion is the charging of different prices by competing sellers for the same good. This phenomenon has and continues to be extensively documented even on the internet where search costs are particularly low (see, for example, Brynjolfsson and Smith, 2000; Baye, Morgan, and Scholten, 2001). There are a number of competing theoretical models that seek to explain price dispersion as an equilibrium phenomenon. Perhaps the most popular model is Varian (1980), in which there is a unique equilibrium in mixed strategies.<sup>3</sup> Price

---

<sup>1</sup>Exceptions include Van Huyck et al. (1994), Huck, Normann and Oechssler (1999), Duffy and Hopkins (2004).

<sup>2</sup>This seems broadly consistent with the analysis of Cason, Friedman and Wagener (2004) of data from experiments on a game with an equilibrium only in mixed strategies.

<sup>3</sup>This model has also been the basis for more recent work for example, Baye and Morgan (2001).

dispersion is therefore a consequence of sellers randomizing over prices. One conclusion of both theoretical and experimental work on learning is that convergence to mixed strategy equilibria is difficult. There is therefore reasonable doubt whether actual price behavior will be similar to the equilibrium behavior predicted by such models.

Testing these models with data has been rare. Villas-Boas (1995), for example, examines actual field price data, and Cason and Friedman (2003) and Morgan, Orzen, and Sefton (2001) provide important tests of comparative static predictions in the experimental laboratory. The advantage to these studies is that they are realistic; sellers have many prices from which to choose. The disadvantage is that the multiplicity of choices makes individual decisions difficult to analyse. Our project differs from the existing literature by, first, simplifying the framework by reducing the number of choices sellers have to set prices so that we can, second, formally test qualitative predictions of learning theory as an explanation for strategic behaviour that emerges in environments that model price dispersion.

We study two-player games with three actions available to each player; for those familiar with economics experiments and/or children's games they are related to the game "rock-scissors-paper". These games are a simplified theoretical model of price dispersion because they mimic the following reasoning behind the construction of equilibrium in the base model (Hopkins and Seymour (2002)). A seller would never set the highest price in the market because it would pay to steal customers by undercutting everyone else's price slightly, but then it would pay for someone to undercut a little more; this reasoning continues until undercutting results in an unprofitably low price, leading to a completion of the cycle where it pays to set the high price, at which point the undercutting reasoning begins anew.

In the experiments, subjects were randomly and anonymously matched to play one game 100 times. The length of repetition is required to give every chance for learning to converge. Each player always plays against the same opponent. This contrasts with many other experiments in which a random matching protocol is implemented. We chose constant pairing so as to be able to investigate individual learning. In a situation where a group of players are repeatedly randomly matched, the behaviour of every subject is linked to the behaviour of every other. In contrast, in constant pairing it is possible to treat the behaviour of an individual pair as being independent of the actions of other subjects.

The reason often advanced against common pairing is that, in such circumstances, subjects will play repeated game strategies, not adjust myopically as learning theory suggests. In some games, such as the prisoner's dilemma, this has been documented to occur, but how widespread is this phenomenon is unclear. In order to reduce collusion, each subject was only informed of her own payoff matrix, but not that of her opponent. Note that knowledge of others' payoffs is not required by either of the two most popular learning models, reinforcement learning and stochastic fictitious play. As we will see, though there were potential gains to collusion in all the games, learning theory describes

play well. That is, constant pairing in itself does not invalidate the use of adaptive learning theory.

In the first experimental treatment the game is a rescaled constant sum game. That is, it is a linear transformation of a constant sum game. In this case the learning models predict convergence to Nash equilibrium (i.e., the equilibrium is “stable”). This means that subjects choose randomly from their available actions, i.e., their actions are not predictable. In the second treatment only the game payoffs are altered so that the outcome of the learning models is cyclic behaviour (i.e., the equilibrium is “unstable”). Thus the next action should become predictable from one round to the next. The design includes one stable and three unstable experimental manipulations. The three unstable games are essentially the same game, but with different levels of monetary incentives. This is because, as we will see, theory predicts that instability of Nash equilibrium under learning depends on the level of incentives in a game.

## 2 RSP Games and Edgeworth Cycles

The children’s game of Rock-Scissors-Paper (RSP) is well known around the world. But it is also an interesting metaphor for various price setting games (Hopkins and Seymour (2002); Cason, Friedman and Wagener (2004)). What they have in common is that there is a cycle of best responses. In RSP, Rock beats Scissors which beats Paper which beats Rock. In a number of different oligopoly games, the best response to a high price is a medium price to which the best response is a low price to which the best response is a high price, restarting the cycle. This phenomenon was first noted by Edgeworth (1925), in the context of price-setting oligopolists under capacity constraints. But the same best response cycle is present, as Hopkins and Seymour (2002) point out, in a number of models of price dispersion, including that of Varian (1980). The fact that there is a cycle of best responses indicates that there can be no equilibrium in pure strategies, the only Nash equilibria that these games possess are mixed.

In this paper, we study a number of simple games that have these characteristics. We chose three two player  $3 \times 3$  games and one three player game. The first game which we call A\*B\* is a game of opposed interest. Player 1 would like to play along the diagonal, whereas Player 2 wants to avoid the diagonal.

	LEFT	CENTRE	RIGHT	
“A*B*”	UP	323, 0	17, 432	153, 108
	MIDDLE	289, 108	323, 0	0, 405
	DOWN	85, 324	187, 108	187, 0

(1)

This game’s unique Nash equilibrium consists of the first player placing the following weights  $(17, 20, 24)/61 \approx (0.279, 0.328, 0.393)$  on her three strategies and  $(63, 65, 116)/244 \approx (0.258, 0.266, 0.475)$  being the corresponding probabilities for Player 2. Equilibrium pay-offs are approximately 161 for Player 1 and 163 for Player 2. The second game B\*B\* is constructed by giving the payoffs of Player 2 to both players.

	LEFT	CENTRE	RIGHT	
“B*B*”	LEFT	0, 0	108, 432	324, 108
	CENTRE	432, 108	0, 0	108, 405
	RIGHT	108, 324	405, 108	0, 0

(2)

This has a unique mixed strategy equilibrium with relative probabilities of approximately  $(0.28, 0.32, 0.39)$ , which gives an equilibrium payoff of approximately 163.

Shapley (1964) was the first to offer the example of games where learning failed to converge. Those games were also  $3 \times 3$  games with a cycle of best responses. However, what more recent research indicates (Ellison and Fudenberg (2000); Hofbauer and Hopkins (2004)) is that in some games of this class, the mixed equilibrium is unstable under learning and in others, it is stable. Indeed, while the games given above might seem quite similar, they have quite different theoretical properties. The equilibrium of the

first game  $A^*B^*$  is predicted to be stable under learning, the equilibrium of the second  $B^*B^*$  is potentially unstable.

Whether it is actually unstable depends on whether incentives are sufficiently sharp (we discuss this issue in detail in the next section). With that question in mind, we introduce two new games based on  $B^*B^*$ , but with higher incentives. In fact, they are simple multiples of this game. The next game we call  $B^{**}B^{**}$  and is simply  $B^*B^*$  multiplied by  $4/3$ . Therefore, the Nash equilibrium of  $B^{**}B^{**}$  is identical to that of  $B^*B^*$ , but the equilibrium payoffs are approximately 217.

	LEFT	CENTRE	RIGHT	
“ $B^{**}B^{**}$ ”	LEFT	0, 0	144, 576	432, 144
	CENTRE	576, 144	0, 0	144, 540
	RIGHT	144, 432	540, 144	0, 0

(3)

The game  $B^{***}B^{***}$  is  $B^*B^*$  multiplied by  $16/9$ . Again the unique Nash equilibrium is mixed with probabilities  $(17, 20, 24)/61 \approx (0.279, 0.328, 0.393)$ , with equilibrium payoff now equal to approximately 290.

	LEFT	CENTRE	RIGHT	
“ $B^{***}B^{***}$ ”	LEFT	0, 0	192, 768	576, 192
	CENTRE	768, 192	0, 0	192, 720
	RIGHT	192, 576	720, 192	0, 0

(4)

All these games have cycles of best responses and no equilibrium in pure strategies. The simple analysis of Edgeworth would therefore suggest play would continue to cycle and never settle down. More modern learning theory, however, predicts convergence to the mixed strategy equilibrium in one of these games, but not in the others, as we will now see.

### 3 Theoretical Predictions

In this section we introduce our learning model. Stochastic fictitious play was introduced by Fudenberg and Kreps (1993) and is further analysed in Benaïm and Hirsch (1999) and Ellison and Fudenberg (2000). It has been applied to experimental data by Cheung and Friedman (1997) and Battalio et al. (2001) among others. We will see that under the classical case of fictitious play beliefs, where every observation is given an equal weight, that stochastic fictitious play gives clear predictions. Specifically, some mixed equilibria are stable, others unstable and the behaviour of learning in the two different cases is quite different. However, we will also see that if instead we assume that players place greater weight on more recent events, this difference is significantly weakened, with very little difference in terms of average play.

Stochastic fictitious play embodies the idea that players play, with high probability, a best response to their beliefs about opponents' actions. Imagine that one of the  $3 \times 3$  games introduced in the previous section was played repeatedly by the same pair of players at discrete time intervals  $n = 1, 2, 3, \dots$ . Let the payoff matrix for player  $i$  be  $A^i$ . We suppose that both players have beliefs about the probability of different strategies being chosen by their opponent. We write the beliefs about player  $i$  as  $(b_{1n}^i, b_{2n}^i, b_{3n}^i)$ , where in this context  $b_{1n}^i$  is  $j$ 's subjective probability in period  $n$  that her opponent  $i$  will play his first strategy in that period. That is,  $b_n^i \in S^3$  where  $S^N$  is the simplex  $\{x = (x_1, \dots, x_N) \in \mathbb{R}^N : \sum x_i = 1, x_i \geq 0, \text{ for } i = 1, \dots, N\}$ . This implies that the vector of expected payoffs of the different strategies of player  $i$ , given her beliefs about  $j$ , will be  $u_n^i = A^i b_n^j$  where  $A^i$  is player  $i$ 's payoff matrix. We will write  $b_n = (b_n^1, b_n^2)$  as a summary of the two players' beliefs.

We assume that each period, each player chooses one of her actions, randomly and independently. Each period, each player receives an independent random shock to her payoffs, and chooses the action with the highest perturbed payoff. We have

$$\tilde{u}_n^i = u_n^i + \epsilon_n^i / \beta = A^i b_n^j + \epsilon_n^i / \beta \quad (5)$$

where  $u_{kn}^i$  is Player  $i$ 's expected payoff to strategy  $k$  given the beliefs over her opponent's actions in period  $n$ . The parameter  $\beta$  is a precision parameter, which scales the noise. Importantly, we assume that  $\epsilon_n^i$  is a vector of random variables, identically and independently distributed. Let the probability that agent  $i$  plays strategy  $k$  in period  $n$  be  $p_{kn}^i$ . If each  $\epsilon_{kn}^i$  is distributed according to the double exponential or extreme value distribution, it is well known that the probability of taking each action will be given by the exponential or logit rule,

$$p_{kn}^i = \frac{\exp \beta u_{kn}^i}{\sum_{m=1}^3 \exp \beta u_{mn}^i} = \phi(b_n^j). \quad (6)$$

Note that for the logit rule, if  $\beta$  is large, the strategy with highest expected payoff is chosen with probability close to one. If  $\beta$  is (close to) zero, then each strategy is chosen with probability (close to) one third, irrespective of the relative expected payoffs.

As is now well known, the limit points of stochastic fictitious play are not Nash equilibria. Rather, they are perturbed equilibria known as quantal response equilibria (QRE) or logit equilibria.<sup>4</sup> Specifically, a perturbed equilibrium  $\hat{p}$  satisfies

$$\hat{p}^i = \phi(\hat{p}^j), \quad \hat{p}^j = \phi(\hat{p}^i) \quad (7)$$

When the parameter  $\beta$  in (6) is large, the set of QRE, will be close to that of Nash equilibria (see McKelvey and Palfrey (1995)). In the class of games we look at which have a unique mixed strategy equilibrium  $p^*$ , we would expect a QRE that was a close approximation to  $p^*$  when  $\beta$  is large.

<sup>4</sup>The literature on these stochastic equilibria is now extensive. See, for example, McKelvey and Palfrey (1995).

One problem in experimental work is that we cannot directly observe either players' beliefs  $b_n$  or their intended mixed strategies  $p_n$ . One thing that can be observed is which choices are actually made. We will therefore find it useful to consider the actual play, defined as

$$X_n^i = (I_{1n}^i, I_{2n}^i, I_{3n}^i) \quad (8)$$

where  $I_{mn}^i = 1$  if player  $i$  plays strategy  $m$  at time  $n$ , else  $I_{mn}^i = 0$ . That is, the vector  $X_n^i$  simply records the choice of player  $i$  in period  $n$ . We will also consider the historical or time average of past play which evolves according to

$$h_{mn+1}^i = h_{mn}^i + \frac{I_{mn}^i - h_{mn}^i}{n}, \text{ for } m = 1, \dots, 3 \quad (9)$$

where again  $I_{mn}^i = 1$  if player  $i$  plays strategy  $m$  at time  $n$ , else  $I_{mn}^i = 0$ . Of course, in the first period we have no history, so  $h_1^i = (0, 0, 0)$  for  $i = 1, 2$ .

Thus, to summarise, we have  $b_n^i$ , player  $j$ 's beliefs about what  $i$  will do in period  $n$ . We have  $p_n^i$ , which is  $i$ 's actual mixed strategy in period  $n$ . The vector  $X_n^i$  gives  $i$ 's choice in period  $n$ , and  $h_n^i$  the history of play prior to the period.

### 3.1 Fictitious Play Beliefs

One possible method of forming beliefs is the fictitious play assumption that beliefs are based on the average of past play. That is, suppose a player observes his opponent  $i$  plays action  $k$  in period  $n$ . Then,

$$b_{mn+1}^i = b_{mn}^i + \frac{I_{mn}^i - b_{mn}^i}{n + t}, \text{ for } m = 1, \dots, 3 \quad (10)$$

where  $I_{mn}^i = 1$  if  $m = k$ , else  $I_{mn}^i = 0$ . The expression for beliefs allows for an initial belief  $b_1^i$  which is non-zero and has weight  $t \geq 0$ . Note, consequently, there is a difference between  $b_n^i$  and  $h_n^i$ : beliefs are based on past play but they are not identical. Subjects may come to an experiment with priors over their opponent's play. That is, specifically

$$b_{mn+1}^i = \frac{I_{mn}^i + I_{mn-1}^i + \dots + I_{m1}^i + tb_{m1}^i}{n + t}.$$

Of course, however, these beliefs are what has been called in the literature "asymptotically empirical", or in other words, asymptotically any priors are washed out by experience and so, if  $\lim_{n \rightarrow \infty} b_n^i$  exists then  $\lim_{n \rightarrow \infty} h_n^i$  exists and is equal.

We now give two simple results due to Ellison and Fudenberg (2000) and Hofbauer and Hopkins (2004a), whose predictions we test experimentally. We are interested in the class of games that possess a fully-mixed equilibrium  $p^* = (p^{1*}, p^{2*})$ . For the first result, we need to introduce the concept of a rescaled zero sum game. A two player

zero sum game is one where  $A^1 = -(A^2)^T$ . A rescaled zero sum game (Hofbauer and Sigmund (1998, 11.2)) is a game that can be made into a zero sum game by a linear transformation, that is, multiplying the payoff matrices by positive constants and/or adding constants to each column. The game AB is an example of such a game. Here, the learning process converges to the perturbed equilibria with probability one (omitted proofs are in the Appendix).

**Proposition 1** *Suppose the game is a rescaled zero sum game. Then, there is a unique perturbed equilibrium point  $\hat{p}$  satisfying the equilibrium conditions (7). And, for stochastic fictitious play (choice rule (6) and updating rule (10)), we have*

$$\Pr(\lim_{n \rightarrow \infty} b_n = \hat{p}) = \Pr(\lim_{n \rightarrow \infty} p_n = \hat{p}) = \Pr(\lim_{n \rightarrow \infty} h_n = \hat{p}) = 1.$$

Since the influential critique of Brown and Rosenthal (1990), it is important to be clear about the nature of this convergence. As they pointed out, for play to be close on average to equilibrium is necessary but not sufficient for true convergence to equilibrium. Here, however, not only does average play,  $h_n$  in current notation, converge to equilibrium, but the actual mixed strategies of both players  $p_n$  do so as well. Furthermore, the beliefs of each player, which under fictitious play assumptions are just past average play, must also converge to that point. Finally, Brown and Rosenthal argued that if experimental subjects were truly playing a mixed equilibrium there should be no serial correlation in play. Here, since the choice probabilities converge to a constant and as the random shocks assumed in (5) are iid, so must be each player's choices. That is, we have the following corollary which implies in rescaled zero sum games, under stochastic fictitious play, there is convergence to a mixed strategy equilibrium that satisfies even the strict criteria of Brown and Rosenthal (1990).

**Corollary 1** *Asymptotically, in a rescaled zero sum game, play  $I_n^i$  for each player will be an iid sequence.*

We have a corresponding result on instability.

**Proposition 2** *Suppose the game is not a rescaled zero sum game and is symmetric so that  $A = B$ . Let  $\hat{x}$  be a perturbed equilibrium point corresponding to a completely mixed Nash equilibrium. Then, for stochastic fictitious play (choice rule (6) and updating rule (10)), there is a  $\hat{\beta} > 0$  such that for all  $\beta > \hat{\beta}$ ,*

$$\Pr(\lim_{n \rightarrow \infty} b_n = \hat{p}) = \Pr(\lim_{n \rightarrow \infty} p_n = \hat{p}) = \Pr(\lim_{n \rightarrow \infty} h_n = \hat{p}) = 0.$$

That is, in symmetric games mixed strategy equilibria are unstable. This implies in the games we consider that only have an equilibrium in mixed strategies, learning should

not converge at all. Notice that this applies both to mixed strategies and to beliefs or equivalently, average play. That is, the time average of play should also diverge from the perturbed equilibrium. In this class of games, we have behaviour which is the diametric opposite to the previous case.

There is one caveat, however. This latter result requires that the parameter  $\beta$  be sufficiently large. Note that if  $\beta$  were zero, the exponential choice rule in effect requires players to choose entirely at random. Hence, if  $\beta$  is very small, then the perturbed equilibrium will be at approximately  $(1/3, 1/3, 1/3)$  irrespective of the game payoff matrix and this equilibrium will be asymptotically stable, as noise swamps the payoffs. One important consideration, however, is that given the way the perturbed best response choice rule is constructed, a proportional increase in all payoffs is equivalent to an increase in  $\beta$ . In particular, given the exponential choice rule (6), if all expected payoffs  $u_n^i$  were doubled, this would have exactly the same effect on the choice probabilities and the stability properties of an equilibrium as doubling  $\beta$ .<sup>5</sup> Thus, a way of translating Proposition 2 into an empirical hypothesis is that the mixed equilibrium of a symmetric game will be unstable, provided incentives are high enough.

There has been one other stochastic learning model that has been popular in the recent literature. Reinforcement learning is found in many different forms in several different disciplines. In economics, references include Erev and Roth (1998), Beggs (2002), Hopkins and Posch (2004). It can be shown that mixed equilibria of symmetric games are also repulsive for reinforcement learning (Hofbauer and Hopkins (2004a)), and that mixed equilibria of rescaled zero sum games are locally attractive for a perturbed form of reinforcement learning (Hopkins (2002)).

### 3.2 Cournot Beliefs and Edgeworth Cycles

Up to now, it has been assumed that all observations are given equal weight. This has the effect that as the number of periods progresses, the marginal impact of new experiences upon behaviour decreases, asymptotically approaching zero. As we have seen this has a certain mathematical convenience. However, both Erev and Roth (1998) and Camerer and Ho (1999) find that experimental data seems to support the hypothesis that agents discount previous experience, which implies that learning will not come to a complete halt even asymptotically (as Cheung and Friedman (1997) point out, disaggregated data indicates that the level of discounting varies enormously between individuals). It has been hypothesised that this form of learning would be useful in non-stationary environments but this claim has received little analysis.

Again, suppose a player observes that his opponent  $i$  plays action  $k$  in period  $n$ . But now

$$b_{mn+1}^i = b_{mn}^i + (1 - \delta)(I_{mn}^i - b_{mn}^i) \text{ for } m = 1, \dots, 3 \quad (11)$$

---

<sup>5</sup>Note that adding a constant to all payoffs, in contrast, has no effect at all.

where again  $I_{mn}^i = 1$  if  $m = k$  and zero otherwise. The parameter  $\delta \in [0, 1)$  is a recency parameter. More recent experiences are given greater weight. In the extreme case of  $\delta = 0$ , only the last experience matters (“Cournot beliefs”). In contrast, as  $\delta$  approaches 1, beliefs approach the previous classical case, where all experience is given equal weight.

Edgeworth (1925) was the first to notice that the Cournot adjustment process could lead to cycles in behaviour. For the games we consider here also, if each player simply plays the best response to the previous choice of her opponent then we would have perpetual cycles. For the game A\*B\* there are six strategy profiles where only one player has an incentive to deviate and they are joined by the following cycle (where the symbol  $\wr$  indicates a return to the first state of the cycle):

$$UL \rightarrow UC \rightarrow MC \rightarrow MR \rightarrow DR \rightarrow DL \wr \quad (12)$$

In the three remaining profiles, both players have an incentive to deviate and the profiles are connected by this cycle:

$$UR \rightarrow DC \rightarrow ML \wr \quad (13)$$

For the game B\*B\*, the equivalent cycles are

$$UM \rightarrow DM \rightarrow DU \rightarrow MU \rightarrow MD \rightarrow UD \wr \quad (14)$$

and

$$UU \rightarrow MM \rightarrow DD \wr \quad (15)$$

As we will see, there is some evidence for these cycles being apparent in experimental play.

Our first result concerns the case of Cournot beliefs. Suppose that players only remember or care about the opponent’s action in the previous period. Suppose further for simplicity that choices are characterised by the simple best response rule. That is, each player chooses the action that has the highest expected payoff (or randomises over the different actions that have the joint highest expected payoff). Then we have the following

**Proposition 3** *Cournot Behaviour: If players use the updating rule (11) with  $\delta = 0$  and play a best response given their beliefs, the time average of play  $h_n^i$  converges to  $(1/3, 1/3, 1/3)$  for  $i = 1, 2$  for both game A\*B\* and game B\*B\*.*

**Proof:** Given strict best responses and Cournot beliefs, play must follow one of the Edgeworth cycles given above.<sup>6</sup> It is easy to verify that the time average of each of the cycles is  $(1/3, 1/3, 1/3)$ .

---

<sup>6</sup>Which will depend on initial beliefs. If initial beliefs are such that one or more players are indifferent, any such tie is broken by randomisation, and from then on, play must follow one of they cycles.

### 3.3 The Intermediate Case: $0 < \delta < 1$

We are able to show the following. Take any 2 player game. Then, stochastic fictitious play, with memory parameter  $\delta$  strictly less than 1, is ergodic. That is, irrespective of the game being played, the time average of play must converge. This is in contrast to the classical results with fictitious play beliefs, where in some games, the time average of play does not converge.

**Proposition 4** *The Markov process defined by stochastic fictitious play with forgetting is ergodic, with an invariant distribution  $\nu_\delta(b)$  on  $S^N \times S^M$ . This implies that*

$$\Pr(\lim_{n \rightarrow \infty} h_n = \tilde{p}) = 1$$

where  $\tilde{p} \in S^N \times S^M = \int p(b) d\nu_\delta(b)$ .

The weakness of this result is that it does not say anything about  $\tilde{p}$ , the point to which the time average converges. However, in some cases it is possible to characterise some aspects of the invariant distribution  $\nu_\delta(b)$ .

Under stochastic fictitious play, the updating rule (11) implies that the step size is exactly  $1 - \delta$ . The first conclusion we can draw from the theory of stochastic approximation is that in the stable case the limit distribution will be in the following sense clustered around the perturbed equilibrium  $\hat{p}$ . The second is that when the equilibrium is unstable, the distribution places no weight near the equilibrium.

**Proposition 5** *Let  $\nu_0 = \lim_{\delta \rightarrow 1} \nu_\delta$  and  $\hat{p}$  be a perturbed equilibrium satisfying (7). Then for any rescaled zero sum game  $\hat{p}$  is unique and*

$$\nu_0(\hat{p}) = 1;$$

(b) *for any symmetric two player game where  $A = B$ , there exists a  $\hat{\beta} > 0$  such that for any  $\beta > \hat{\beta}$ , there is a neighbourhood  $V$  of  $\hat{p}$  such that*

$$\nu_0(V) = 0;$$

What this implies is that for the game AB, for values of  $\delta$  close to one, play  $p_n$  and the long term time average  $h_n$  will be close to  $\hat{p}$ . For the game BB, however, players' mixed strategies  $p_n$  will not be close to the mixed strategy equilibrium. However, even in the stable case, play will not satisfy the strict conditions of Brown and Rosenthal (1990), in that it will not be iid. Specifically, the probability each period that a player plays a particular action will be an independent draw, but from a changing distribution.

**Proposition 6** *At any time under stochastic fictitious play with  $\delta \in [0, 1)$ , play  $X_n$  is a vector autoregressive process.*

**Proof:** Given the updating dynamic (11), beliefs  $b_n$  follow a AR(1) process. Using standard time series results, one can calculate the variance of  $b_{mn}^i$  as  $p_{mn}^i(1 - p_{mn}^i)(1 - \delta)/(1 + \delta)$  which is non-vanishing. The probabilities  $p_n$  that generate the play are a time invariant function (6) of the beliefs, and, therefore, also follow an AR(1) process.

### 3.4 Numerical Analysis of the Games in the Experiments

In this section we try to apply the above theoretical results more closely to the actual payoff matrices used experimentally. We have seen that while the stability of the equilibrium of the game  $A^*B^*$  is independent of the level of incentives, the equilibrium of the games  $B^*B^*$ ,  $B^{**}B^{**}$  and  $B^{***}B^{***}$  will be unstable only if the parameter  $\beta$  is higher than some critical level  $\hat{\beta}$ . This threshold level  $\hat{\beta}$  depends critically on the level of incentives. As  $B^{***}B^{***}$  offers higher incentives than  $B^{**}B^{**}$  than  $B^*B^*$  than  $B^*B^*$ , the critical level  $\hat{\beta}$  will be lower in  $B^{**}B^{**}$  than in  $B^*B^*$  than in  $B^*B^*$ .

Note that it is possible to calculate the critical  $\hat{\beta}$  numerically, and in the the case of the game  $B^*B^*$ , we find that  $\hat{\beta} \approx 0.0125$ .<sup>7</sup> As payoffs in  $B^{**}B^{**}$  are 4/3 those in  $B^*B^*$ ,  $\hat{\beta}$  for  $B^{**}B^{**}$  will be approximately 0.009 ( $0.0125 \times 4/3 \approx 0.009$ ) and for  $B^{***}B^{***}$  approximately 0.007. Battalio et al. (2001) estimate the current model of stochastic fictitious play from experimental data and find values of  $\beta$  ( $\lambda$  in their notation) from 0.14 to 0.3, when payoffs are in cents per game. To be directly comparable, we have to divide their estimates by 10, as in our experiments only one out of 10 games were paid. Furthermore, the subjects in our experiments were paid in Canadian dollars. There are arguments that the real exchange rate is much closer to 1:1, but using the nominal exchange rate (about 1:1.4 at the time the experiments were carried out), converting their estimates to our scale, they range from 0.01 to 0.021.

This highlights an important issue. The level of incentives commonly used in experiments may not in fact be adequate to generate unstable behaviour. Specifically, the game  $B^*B^*$ , given the parameter estimates of Battalio et al. (2001), may not generate unstable behaviour, even if subjects play according to stochastic fictitious play. However, our approximate calculations indicate the games  $B^{**}B^{**}$  and  $B^{***}B^{***}$  should provide adequate incentives.

When beliefs exhibit recency, that is,  $\delta < 1$ , we know that the learning process is

---

<sup>7</sup>We first find an analytic expression for the linearisation of the dynamics (17) given in the Appendix. We fix a certain value for  $\beta$ , we solve numerically the equations (7) to calculate the perturbed equilibrium  $\hat{p}$  corresponding to that level of  $\beta$ , substitute these values into the linearisation, and then calculate the eigenvalues numerically. If all eigenvalues are negative, which they will be for low values of  $\beta$ , we raise the value of  $\beta$  and repeat.

ergodic, and that its time average exists. Hence, we know also that the time average of any simulation will converge to that time average. We conclude this section with some numerical simulations of stochastic fictitious play. The following table gives time averages of the first player's choices  $(h_{1n}^1, h_{2n}^1)$  where  $n$  is sufficiently large for convergence to 3 decimal places (between 500 and 100,000 periods for different parameter values) for the game A\*B\* (the frequency of the third strategy is omitted for reasons of space but can be easily calculated by subtracting the sum of the figures given from 1). The final column gives the perturbed equilibrium for the corresponding value of the precision parameter  $\beta$ . This is calculated independently by numerical solution to the equilibrium equations (7).

$\beta$	$\delta = 0$	$\delta = 0.5$	$\delta = 0.85$	$\delta = 0.999$	$\hat{p}$
0.001	(0.334, 0.341)	(0.330, 0.346)	(0.331, 0.342)	(0.330, 0.341)	(0.330, 0.343)
0.0125	(0.314, 0.400)	(0.295, 0.398)	(0.293, 0.389)	(0.294, 0.383)	(0.294, 0.384)
0.1	(0.330, 0.341)	(0.328, 0.342)	(0.273, 0.381)	(0.278, 0.340)	(0.278, 0.339)
0.5	(0.333, 0.333)	(0.333, 0.333)	(0.266, 0.387)	(0.278, 0.330)	(0.278, 0.330)

Similarly, for Player 2 in game A\*B\* we have:

$\beta$	$\delta = 0$	$\delta = 0.5$	$\delta = 0.85$	$\delta = 0.999$	$\hat{p}$
0.001	(0.328, 0.336)	(0.330, 0.335)	(0.325, 0.339)	(0.324, 0.338)	(0.326, 0.338)
0.0125	(0.276, 0.326)	(0.250, 0.329)	(0.255, 0.317)	(0.256, 0.314)	(0.256, 0.313)
0.1	(0.329, 0.331)	(0.327, 0.327)	(0.255, 0.280)	(0.248, 0.274)	(0.250, 0.273)
0.5	(0.333, 0.333)	(0.333, 0.333)	(0.261, 0.266)	(0.255, 0.269)	(0.256, 0.268)

The numerical simulations conform with the theoretical predictions. See that, for very short memory and for  $\beta$  reasonably large, the time average is effectively at the centre of the simplex, that is  $(1/3, 1/3)$ , which mirrors the prediction from Section 3.2 on Edgeworth cycles. However, for  $\delta$  close to 1, the time average of play is extremely close to the perturbed equilibrium, given in the final column. And, consequently, for the precision parameter  $\beta$  high, the time average is close to the Nash equilibrium.

We then repeat the exercise for the game B\*B\*.

$\beta$	$\delta = 0$	$\delta = 0.5$	$\delta = 0.85$	$\delta = 0.999$	$\hat{p}$
0.001	(0.325, 0.340)	(0.327, 0.337)	(0.326, 0.337)	(0.328, 0.338)	(0.327, 0.338)
0.0125	(0.322, 0.338)	(0.300, 0.344)	(0.295, 0.342)	(0.295, 0.344)	(0.295, 0.343)
0.1	(0.333, 0.333)	(0.308, 0.328)	(0.299, 0.343)	(0.296, 0.344)	(0.281, 0.332)
0.5	(0.333, 0.333)	(0.333, 0.333)	(0.297, 0.338)	(0.298, 0.344)	(0.277, 0.329)

Notice the difference here. For  $\beta \leq \hat{\beta} \approx 0.0125$  and when  $\delta$  is close to one, the time average is close to the perturbed equilibrium. However, this is no longer true when  $\beta > \hat{\beta}$ . Here, the theory says that, for  $\delta$  closer to 1, play should not be close to the perturbed equilibrium, however, the time average should converge. It turns out that time average converges to a point that is close but not identical to the perturbed equilibrium (compare for  $\beta \leq \hat{\beta}$  and  $\delta$  close to 1 where the time average hits the perturbed equilibrium exactly).

### 3.5 Summary of Predictions from Theoretical Findings

Under fictitious play beliefs, where agents place an equal weight on all previous observations, the main predictions from stochastic fictitious play for the games conducted experimentally are the following.

1. In game  $A^*B^*$ , mixed strategies, beliefs and historical frequencies should all converge to the perturbed equilibrium. Any serial dependence should disappear with time.
2. Theory predicts that for high enough incentives, in games of type  $B^*B^*$ ,  $B^{**}B^{**}$ ,  $B^{***}B^{***}$ , the equilibrium is unstable. It is uncertain what monetary incentives in practice would be sufficient. However, since incentives are higher in  $B^{***}B^{***}$  than in  $B^{**}B^{**}$  than in  $B^*B^*$ , the dispersion of mixed strategies, beliefs and historical frequencies away from the mixed equilibrium should be no smaller in  $B^{***}B^{***}$  than in  $B^{**}B^{**}$  than in  $B^*B^*$ .

Under beliefs with recency where agents place a greater weight on more recent observations, the main predictions from stochastic fictitious play for the games conducted experimentally are the following.

1. In game  $A^*B^*$ , historical frequencies should converge to a point close to the perturbed equilibrium. Mixed strategies and beliefs do not converge to a point but remain close to the perturbed equilibrium. Serial dependence is persistent.
2. Theory predicts that for high enough incentives, in games of type  $B^*B^*$ ,  $B^{**}B^{**}$ ,  $B^{***}B^{***}$ , the equilibrium is unstable. It is uncertain what monetary incentives in practice would be sufficient. However, since incentives are higher in  $B^{***}B^{***}$  than in  $B^{**}B^{**}$  than in  $B^*B^*$ , the dispersion of mixed strategies and beliefs away from the mixed equilibrium should be no smaller in  $B^{***}B^{***}$  than in  $B^{**}B^{**}$  than in  $B^*B^*$ . However, in all three games the historical frequencies converge to a point, close to but not identical to the perturbed equilibrium.

## 4 Experimental Design and Procedures

We ran four different experimental treatments labelled  $B^*B^*$ ,  $B^{**}B^{**}$ ,  $B^{***}B^{***}$ , and  $A^*B^*$ . Thirty-four subjects (i.e., seventeen subject pairs) participated in  $B^*B^*$  and  $B^{***}B^{***}$ , and thirty subjects (i.e., fifteen subject pairs) participated in  $B^{**}B^{**}$  and  $A^*B^*$ . No subject participated in more than one session.

All the rules of the game were common knowledge and subjects were given complete information regarding their opponent's behavior. In the upper-right corner of their computer screen the subjects were presented with a 3x3 payoff matrix called the earnings table. At the left side of the screen they could scroll through the entire history of their and their opponent's decisions. At the bottom of the screen they were shown their and their opponent's past frequency and proportion of play of each of the three stage-game strategies. We included these summaries of play to minimize forgetting, or discounting, past play. The history information was updated after each period.

All subjects viewed the game as if they were the row player, and the screen revealed only the subject's, and not the opponent's, payoff in each cell. We hid the opponent's payoffs to suppress behavior due to other-regarding preferences, for which there is a mountain of experimental evidence. Note that the learning theory is silent with respect to the opponent's payoff. More specifically, the learning theory assumes that own payoffs are utilities.

We presented the "earnings table" in the instructions in symbolic form, using the letters A through I to represent payoffs. Thus the only difference between experimental treatments was the numbers presented to the subjects in the earnings table on the computer screen. Identical instruction sheets were used for all treatments. In fact, in principle it would have been possible to run multiple treatments in a session, but we did not do so.

Subjects were told that they would make one-hundred decisions, that they would be randomly chosen to be the row or column player (each subject was always the row player, but the opponent was always presented as the column player), and that the participant with whom they were randomly paired would stay the same throughout the entire session. They were informed that they would be paid for a randomly-chosen ten periods of play to be determined by the computer (to control for wealth effects). The sessions never lasted more than an hour and a half. The subjects had to correctly answer questions on a quiz indicating that they understood how to read the earnings table, and that they understood that they did not know the content of their opponent's earnings table.

A total of 128 subjects, who were English-speaking university students in Montreal, participated in the four experimental treatments. The experiment was programmed and conducted with the software z-Tree (Fischbacher 1999). The experiments were run in

May and June, 2004 at the Bell Experimental Laboratory for Commerce and Economics at the Centre for Research and Analysis on Organizations (CIRANO). Subjects earned CAD \$10.00 for showing up on time and participating fully (our show-up fee must take into account the fact that the laboratory is not on campus), and an average of \$23.90 for the results of their decisions and the decisions of their opponent. Alternative opportunities for pay in Montreal for our subjects is considered to be approximately \$8.00 per hour.

## 5 Experimental Results

Table 1 presents the proportion of times each stage-game strategy was played, and the average earnings per round, averaged over all one-hundred rounds for the four experimental treatments. Results for the treatment  $A^*B^*$  are divided between the row player and the column player since the roles in this game are asymmetric. The table also presents the stage-game equilibrium point predictions for comparisons.

Table 1 reveals that for the most part, marginal choice frequencies were very close to stage-game equilibrium predictions. For example, in  $B^*$  strategies 1, 2, and 3 were chosen with frequencies 0.29, 0.36, and 0.35 while equilibrium predicts the frequencies 0.28, 0.32, and 0.39. This close correspondence with equilibrium holds for the treatments  $B^*$ ,  $B^{**}$ , and  $B^{***}$ . In treatment  $A^*B^*$ , however, the row players again matched the marginal choice frequencies of equilibrium, but the column players deviated substantially: they overplayed strategy 2 (0.36 vs. 0.27) and underplayed strategy 3 (0.34 vs. 0.48). And this deviation appears to have been profitable: they out-earned their row-player opponents 187 to 155 per round, plus they earned more than equilibrium predicts.

While marginal choice frequencies were close to equilibrium, it does not appear that equilibrium was being played: in every case except for the row players in treatment  $A^*B^*$  players earned too much per round. For example, in treatment  $B^*$  the average per-round earnings were 71 cents above equilibrium earnings, and one can see the same pattern in the other treatments. This suggests that the players may have picked up opportunities for collusive equilibria in the repeated game.

Our main results are displayed in Figures 1 through 4, which show how dispersed the marginal choice frequencies were throughout the sessions. To construct the figures we grouped the choice data from each player into ten blocks of ten consecutive periods: the first block consisted of periods 1-10, the second block periods 11-20, etc., up to the tenth block consisting of periods 91-100. For each block we computed the proportion of times the subject played each of her three strategies.

We then aggregated these results across subjects. Since we divided the data into blocks of ten periods, the feasible proportions of play of any strategy are limited to the set  $0, 0.1, 0.2, \dots, 1.0$ ; the sum of proportions of play for the three strategies within a block for a subject is always 1. For each block of ten periods we counted across subjects the number of times a particular set of proportions of play occurred. For example, in a particular block we counted three subjects who selected strategies 1, 2, and 3 with proportions 0.3, 0.2, and 0.5.

Figures 1 through 4 present the data: they reveal the frequency with which particular proportions of play occurred within blocks of ten periods. We divided the sessions into the first and second half (i.e., rounds 1-50 and rounds 51 - 100) to observe possible differences in behavior before and after the subjects gained experience playing the game.

Figures 1 and 2 present the B and AB treatments in rounds 1-50, and Figures 3 and 4 present the same information for rounds 51-100.

In the figures, the horizontal axis represents the proportion of times strategy 1 was played during a block of ten consecutive games, and the vertical axis represents the same for strategy 2 (the proportion computed for strategy 3 is not necessary to plot). The width of the circles that locate each set of proportions is proportional to the number of times that particular set of proportions was realized in the data.

For example, in the top-most panel of Figure 1, one can see that in the B\*B\* treatment, strategies 1, 2, and 3 were played with proportion 0.4, 0.3, 0.3 by the largest number of subjects in the five blocks contained within rounds 1 - 50; this is so because the largest circle is located at 0.4 on the horizontal axis and 0.3 on the vertical axis. Strategy choice proportions 0.2, 0.3, 0.5 and 0.2, 0.4, 0.4 were played second-most frequently. And strategy choice proportions 0.1, 0.1, 0.8, for example, were played once.

Figure 1 reveals a great deal of dispersion in all three B matrix treatments, but a hint of a possible decrease in dispersion in the treatment B\*\*\*. One can see this by noting the increasing number of larger circles, indicating higher concentration of frequency of play, and by noticing that observed frequency of play clusters more tightly around the stage-game equilibrium. This tends to work against the theory that higher incentives to deviate increase instability of play. Figure 2 shows the dispersion of frequency of play for the row players in A\*B\*. We only plot the A player for this game because theoretical results depend first on A players randomizing which causes B players to follow. It is obvious that play is very dispersed here as well.

Figures 3 and 4 allow a comparison of play over time as well as between treatments. Figure 3 shows play in B\* and B\*\* moving in towards equilibrium and possibly becoming less dispersed over time. Again B\*\*\* appears less dispersed than the other two treatments. The presence of more large circles clustering inside the boundary in Figure 4 suggests that play in A\*B\* may be becoming less dispersed over time.

How should these figures look? Figures 5 and 6 present simulation of stochastic fictitious play, which were run under the same conditions (i.e., same number of subjects and same number of rounds) as the experimental conditions. We present simulation data from the last fifty periods of play. Model parameters were similar to those estimated from experimental data by Battalio et al. (2001). Specifically, the recency parameter  $\delta$  was 0.85 and the precision parameter  $\beta$  was 0.01. This latter value was at the low end of the estimates of Battalio et al. as discussed in Section 3.4. It is such that the mixed equilibrium in B\* is stable, but in B\*\*\* it is unstable. Figure 5 shows increasing instability from B\* to B\*\*\*: play moves sharply to the boundaries. And Figure 6 shows A\*B\* clustered more tightly within the triangle of possible play, near equilibrium.

To test for significance in the experimental data we observe the simple fact that as play becomes unstable, it becomes more predictable. This is because unstable play

moves through best-response cycles in a deterministic manner. Thus the B treatments should be more predictable than the AB treatment, and we test this predictions in two ways.

First, we constructed a table of action pairs taken in the last fifty rounds of each treatment. Table 2 presents these aggregated data, with the rows labelled 1, 2, and 3 representing actions taken by the row player and columns representing the same for the column player. The striking result here is the low frequency of play along the diagonal in the B treatments, with a greater frequency occurring in the A\*B\* treatment. By eye it appears that A\*B\* is qualitatively different from the B treatments. However, we performed a chi square test, testing formally for statistical independence between the rows and the columns, and rejected the null hypothesis of independence in every case (p-value  $\leq 0.002$  in each case, although the t-statistic was quite a bit lower in the A\*B\* case).<sup>8</sup>

## 6 Conclusions

We tested the theoretical possibility, first discovered by Shapley (1964), that in some games learning fails to converge to any equilibrium, either in terms of marginal frequencies or of average play. Subjects played repeatedly in fixed pairings one of two  $3 \times 3$  games, each having a unique Nash equilibrium in mixed strategies. The equilibrium of one game is predicted to be stable under learning, the other unstable.

We found a high degree of dispersion in play in both games, and we found play from round-to-round not to be independent in either treatment as well. However, we found that play is less predictable in the stable game under learning than in two of the unstable games, and this result is consistent with learning theory's prediction. One unanswered question is whether increasing incentives to deviate in our A\*B\* treatment would change behavior with respect to instability. Another is to characterize play in the B\*\*\* treatments, which did not seem to increase in level of instability.

Our experimental design provided a tough test of the theory by studying play in fixed pairings, and by providing the subjects with the entire history of play as well as history of marginal choice frequencies, which would presumably make randomizing easier or more natural. However, this control for the forgetting parameter did not cause the unstable games to be as unpredictable as the stable game.

## Appendix

Most of the theoretical results follow from application of the theory of stochastic approximation. Note that that, given the updating rules (10) or (11), the expected change in beliefs can be written

$$E[b_{mn+1}^i | b_n] - b_{mn}^i = \gamma_n(p_{mn}^i - b_{mn}^i), \text{ for } m = 1, \dots, 3 \quad (16)$$

where  $\gamma_n$  is the step size (equal to  $1/(n+t)$  under fictitious play beliefs, equal to  $1-\delta$  under recency). In the terminology of stochastic approximation, the associated system of ordinary differential equations (ODE) can be written in vector form as

$$\dot{b}^i = \phi(b^j) - b^i, \quad \dot{b}^j = \phi(b^i) - b^j. \quad (17)$$

It is possible to predict the behaviour of the stochastic learning process by the analysis of the behaviour of these ODE's.

**Proof of Proposition 1:** In this game, the perturbed equilibrium is unique and globally stable under the associated ODE (17) by Theorem 3.2 of Hofbauer and Hopkins (2004a). This in turn implies convergence with probability 1 of stochastic fictitious play by standard results in the theory of stochastic approximation (see, for example, Benveniste et al. (1990), Chapter 2, Corollary 6).

**Proof of Proposition 2:** Any perturbed equilibrium of this game is unstable under the ODE (17) for sufficiently large  $\beta$  by Theorem 4.7 of Hofbauer and Hopkins (2004a). For  $\beta$  larger than the critical value  $\hat{\beta}$ , stochastic fictitious play converges to the equilibrium with probability zero, by the result of Pemantle (1992).

**Proof of Proposition 4:** Some results follow based on techniques developed by Norman (1968). The focus is now on the Markov process defined by the appropriate choice rule and the updating rule (11). The state of the process at any time can be given by  $b_n \in S^N \times S^M = S$ , that is, a vector of beliefs for each player. This obviously evolves according to the actions chosen by the two players. If the first player chooses action  $i$ , and the second  $j$ , then denote that event as  $ij$  and event operator  $f_{ij}$ . Norman (1968) defines a Markov process on a metric space with metric  $d$  to be strictly “distance diminishing” if  $\rho(f_{ij}) < 1$  for all  $ij$  where

$$\rho(f) = \sup_{b \neq b'} \frac{d(f(b), f(b'))}{d(b, b')}.$$

**Lemma 1** *The Markov process defined by stochastic fictitious play with forgetting is distance diminishing with respect to the standard Euclidean metric.*

**Proof:** Given arbitrary states  $b, b'$ ,  $f_{ij}(b) = ((1-\delta)I^1 + \delta b^1, (1-\delta)I^1 + \delta b^2)$  and  $f_{ij}(b') = (((1-\delta)I^1 + \delta b'^1, (1-\delta)I^1 + \delta b'^2)$ . It is easy to show therefore that  $d(f_{ij}(b), f_{ij}(b')) = \delta d(b, b')$  and  $\rho(f_{ij}) = \delta$  for all possible events.

Let  $T_n(b)$  be the set of states reached with positive probability in  $n$  steps if we start at  $b$ . Let  $d(S_1, S_2)$  be distance between two subsets  $S_1$  and  $S_2$  of the state space. That is,

$$d(S_1, S_2) = \inf_{b \in S_1, b' \in S_2} d(b, b')$$

Then Norman (1968, Theorem 2.2, p66) is able to show that if the following condition holds

$$\lim_{n \rightarrow \infty} d(T_n(b), T_n(b')) = 0 \text{ for all } b, b' \in S$$

then a distance diminishing Markov process is ergodic. That is, its limit distribution is independent of initial conditions. From an arbitrary initial state  $b_0$  there is a positive probability that both players continue to choose their first action for an indefinite number of periods. As this run of play continues,  $b_n$  will approach the state  $(I_1^1, I_1^2)$ . This state is therefore accessible from any initial state and from the theorem of Norman the Markov process is ergodic.

**Proof of Proposition 5:** (a) The perturbed equilibrium is unique and globally asymptotically stable under the ODE (17) and has all negative eigenvalues (Theorems 3.2 and 4.5 of Hofbauer and Hopkins (2004a)). The result then follows from Benveniste et al. (1990, Ch2, Theorem 3). (b) By Theorem 4.7 of Hopkins and Hofbauer (2004a) there exists a  $\hat{\beta}$  such that for  $\beta > \hat{\beta}$ , the perturbed equilibrium is a saddlepoint under the ODE (17). This result then follows from Theorem 3.7 of Benaïm (1998).

## References

- Battalio, R., Samuelson, L. Van Huyck, J. (2001) "Optimization Incentives And Coordination Failure In Laboratory Stag Hunt Games," *Econometrica*, **69**, 749-764.
- Baye, M.R., Morgan, J. (2001): "Information gatekeepers on the internet and the competitiveness of homogeneous product markets", *American Economic Review*, 91: 454-474.
- Benaïm, M. (1998) "Recursive algorithms, urn processes and chaining number of chain recurrent sets," *Ergodic Theory and Dynamical Systems*, **18**, 53-87.
- Benaïm, M. (1999): "Dynamics of Stochastic Algorithms," in *Séminaire de Probabilités XXXIII*, J. Azéma et al. Eds, Berlin: Springer-Verlag.
- Benaïm, M., Hirsch, M.W. (1999) "Mixed equilibria and dynamical systems arising from fictitious play in perturbed games," *Games Econ. Behav.*, **29**, 36-72.
- Benveniste, A., M. Métivier, and P. Priouret (1990): *Adaptive Algorithms and Stochastic Approximations*. Berlin: Springer-Verlag.
- Brown, J., and Rosenthal, R. (1990). "Testing the minimax hypothesis: a reexamination of O'Neill's game experiment," *Econometrica*, **58**, 1065-1081.
- Camerer, C. (2003). *Behavioral Game Theory*, Princeton University Press, Princeton.
- Camerer, C., Ho, T-H. (1999). "Experience-weighted attraction learning in normal form games," *Econometrica*, **67**, 827-874.
- Cason, T., Friedman, D. (2003). "Buyer search and price dispersion: a laboratory study," *Journal of Economic Theory*, **112**, 232-260.
- Cason, T., Friedman, D., and Wagener F., (2004) "The Dynamics of Price Dispersion or Edgeworth Variations", forthcoming *Journal of Economic Dynamics and Control*.
- Cheung, Y-W., Friedman, D. (1997) "Individual learning in normal form games: some laboratory results," *Games Econ. Behav.*, **19**, 46-76.
- Duffy J., Hopkins, E. (2004) "Learning, information and sorting in market entry games: theory and evidence," forthcoming *Games Econ. Behav.*
- Edgeworth, F. Y. (1925), *The Pure Theory of Monopoly*, in *Papers Relating to Political Economy*, vol. 1, by F. Y. Edgeworth (New York: Burt Franklin).
- Ellison, G., Fudenberg, D. (2000) "Learning purified mixed equilibria," *J. Econ. Th.*, **90**, 84-115.

- Erev, I., Roth, A.E., “Predicting how people play games: reinforcement learning in experimental games with unique, mixed strategy equilibria,” *Amer. Econ. Rev.*, **88** (1998), 848-881.
- Fort, J-C., Pagès, G. (1999) “Asymptotic behavior of a Markovian stochastic algorithm with constant step,” *Siam Journal of Control and Optimization*, **37**, 1456-1482.
- Fudenberg, D., Levine D., *The Theory of Learning in Games* (Cambridge, MA: MIT Press, 1998).
- Gaunersdorfer, A., and J. Hofbauer (1995): “Fictitious play, Shapley Polygons, and the Replicator Equation,” *Games and Economic Behavior*, **11**, 279-303.
- Hofbauer, J., Sigmund K., *Evolutionary Games and Population Dynamics* (Cambridge: Cambridge University Press, 1998).
- Hofbauer J., Hopkins, E. (2004a). “Learning in Perturbed Asymmetric Games,” forthcoming *Games Econ. Behav.*
- Hofbauer J., Hopkins, E. (2004a). “When Learning “Converges” to Unstable Equilibria: Fictitious Play with Recency,” work in progress.
- Hopkins, E. (2002): “Two Competing Models of How People Learn in Games,” *Econometrica*, **70**, 2141-2166.
- Hopkins, E., Posch, M. (2002): “Attainability of Boundary Points under Reinforcement Learning”, forthcoming *Games Econ. Behav.*
- Hopkins, E., Seymour, R. (2002) “The Stability of Price Dispersion under Seller and Consumer Learning”, *International Economic Review*, **43**, 1157-1190.
- McKelvey, R.D., Palfrey, T.R. (1995). “Quantal response equilibria for normal form games,” *Games Econ. Behav.*, **10**, 6-38.
- Morgan, J., H. Orzen, and M. Sefton (2001), “An Experimental Study of Price Dispersion”, working paper, University of Nottingham.
- Norman, M.F. (1968). “Some convergence theorems for stochastic learning models with distance diminishing operators,” *J. Math. Psych.*, **5**, 61-101.
- Pemantle, R. (1990). “Nonconvergence to Unstable Points in Urn Models and Stochastic Approximations,” *Ann. Probability*, **18**, 698-712.
- Shapley, L., “Some topics in two person games,” in M. Dresher et al. eds., *Advances in Game Theory*, (Princeton: Princeton University Press, 1964).
- Tang, Fang-Fang, (2001), “Anticipatory learning in two-person games: some experimental results” *Journal of Economic Behavior and Organization*, **44**: 221-232.

Van Huyck, J.B., J.P. Cook and R.C. Battalio (1994), "Selection Dynamics, Asymptotic Stability, and Adaptive Behavior" *Journal of Political Economy*, 102: 975-1005.

Varian, H.R.,(1980), "A model of sales," *Amer. Econ. Rev.*, **70**, 651-659.

**Table 1: Aggregate Results and Equilibrium Predictions**

	<b>B*B*</b>		<b>B**B**</b>		<b>B***B***</b>		<b>A*B*</b>			
	<b>Results</b>	<b>Equilibrium</b>	<b>Results</b>	<b>Equilibrium</b>	<b>Results</b>	<b>Equilibrium</b>	<b>Row Player</b>		<b>Column Player</b>	
							<b>Results</b>	<b>Equilibrium</b>	<b>Results</b>	<b>Equilibrium</b>
<b>Strategy 1</b>	0.29	0.28	0.3	0.28	0.31	0.28	0.31	0.28	0.3	0.26
<b>Strategy 2</b>	0.36	0.32	0.36	0.32	0.35	0.32	0.29	0.32	0.36	0.27
<b>Strategy 3</b>	0.35	0.39	0.34	0.39	0.34	0.39	0.4	0.39	0.34	0.48
<b>Profit</b>	191	120	264	217	326	290	155	161	187	163

Notes: Experimental results are computed as proportion of actual play of strategies 1, 2, and 3, and average profit per round.

**Table 2: Strategy Pairs Played in Rounds 51-100**

		Column			
		B*	1	2	3
Row	1	45	112	106	
	2	88	78	139	
	3	89	118	75	

		Column			
		B**	1	2	3
Row	1	38	128	118	
	2	86	47	122	
	3	75	85	51	

		Column			
		B***	1	2	3
Row	1	48	124	119	
	2	104	68	124	
	3	81	119	63	

		Column			
		A*B*	1	2	3
Row	1	66	98	73	
	2	54	67	88	
	3	109	114	81	

**Table 3: Multinomial Regression Results**

Decision	Variable	B*		B**		B***		A*B*	
		row	column	row	column	row	column	row	column
1	s12	1.51 (3.26)	0.89 (1.81)	1.98 (4.09)	1.03 (1.86)	0.46 (1.08)	-0.74 (-1.49)	-0.6 (-1.51)	0.82 (1.66)
	s13	1.5 (3.26)	0.95 (1.94)	2.4 (4.75)	2.74 (4.61)	1.42 (3.26)	1.2 (2.71)	-0.08 (-0.18)	1.44 (3.87)
	s21	-0.12 (-0.26)	-0.61 (-1.33)	0.14 (-0.29)	-0.63 (-1.35)	0.36 (0.88)	0.29 (0.73)	-0.66 (-1.49)	-0.66 (-1.54)
	s22	1.11 (2.26)	-0.57 (-1.14)	0.05 (0.10)	1.37 (2.21)	0.62 (1.52)	0.33 (0.69)	-0.81 (-1.93)	-0.08 (-0.16)
	s23	0.8 (1.81)	-0.42 (-0.93)	1.22 (2.71)	1.66 (3.49)	0.43 (1.05)	0.27 (0.67)	-0.26 (0.63)	0.6 (1.67)
	s31	-0.31 (-0.63)	-0.86 (-1.81)	-1.3 (-2.27)	-1.57 (-2.77)	-0.89 (-1.85)	-0.87 (-2.00)	-0.53 (-1.35)	-0.38 (-0.84)
	s32	1.01 (2.12)	-0.19 (-0.43)	-0.19 (-0.35)	-0.51 (-0.09)	0.2 (0.47)	0.3 (0.70)	0 (0.00)	0.52 (-1.19)
	s33	0.98 (2.03)	0.69 (1.36)	1.44 (2.68)	1.76 (3.04)	0.54 (1.11)	0.56 (1.18)	0.08 (0.18)	0.8 (2.02)
	Constant	-0.94 (-2.39)	-0.22 (-0.58)	-0.74 (-1.85)	-0.75 (-1.85)	-0.43 (-1.21)	-0.45 (-1.32)	0.41 (1.33)	-0.66 (-2.22)
	3	s12	0.98 (2.27)	0.65 (1.43)	1.55 (3.30)	2.14 (4.33)	0.38 (0.95)	1.08 (2.68)	0 (0.01)
s13		0.67 (1.55)	0.6 (1.31)	1.51 (3.02)	2.01 (3.41)	0.76 (1.80)	0.69 (1.49)	0.49 (1.11)	0.1 (0.21)
s21		-0.44 (-0.99)	-0.6 (-1.45)	-0.51 (1.07)	-1.7 (-3.38)	-1.04 (-2.34)	-0.91 (-2.05)	-0.58 (-1.21)	0.55 (1.42)
s22		0.89 (1.99)	-0.38 (-0.88)	-0.15 (-0.28)	1.93 (3.41)	0.36 (0.79)	0.96 (2.12)	-0.56 (-1.27)	1.42 (3.39)
s23		0.3 (0.73)	-0.79 (-1.89)	-0.53 (-1.08)	-0.35 (-0.63)	-0.51 (-1.24)	0.38 (0.93)	0.53 (1.21)	-0.16 (-0.39)
s31		-0.06 (-0.15)	-0.54 (-1.32)	-0.73 (-1.54)	-0.11 (-0.24)	-0.39 (-0.94)	0.02 (-0.05)	0.23 (0.57)	0.66 (1.62)
s32		1.41 (3.36)	-0.02 (-0.04)	1.06 (2.38)	2.07 (4.58)	0.39 (0.99)	0.96 (2.37)	1.3 (3.09)	1.66 (4.05)
s33		0.49 (1.1)	0.75 (1.61)	0.63 (1.14)	1.41 (2.49)	0.41 (0.89)	0.7 (1.50)	0.81 (1.83)	0.52 (1.20)
Constant		-0.57 (-1.64)	0.18 (-0.52)	-0.55 (-1.44)	-0.55 (-1.44)	-0.16 (-0.498)	-0.45 (-1.32)	0.05 (0.16)	-0.86 (2.69)

Notes: Coefficient estimate and t-stat in parenthesis. The coefficients are not marginal effects, but can be interpreted relative to each other and to the base categories (decision = 2; state = 11).

Figure 1: Proportion of Play Within Blocks of Ten Consecutive Games Rounds 1 - 50, B Matrix Treatments

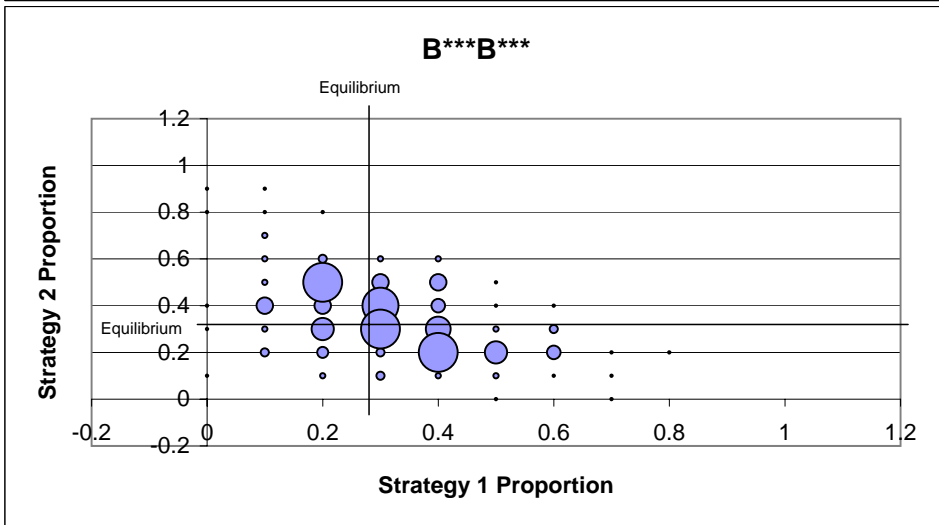
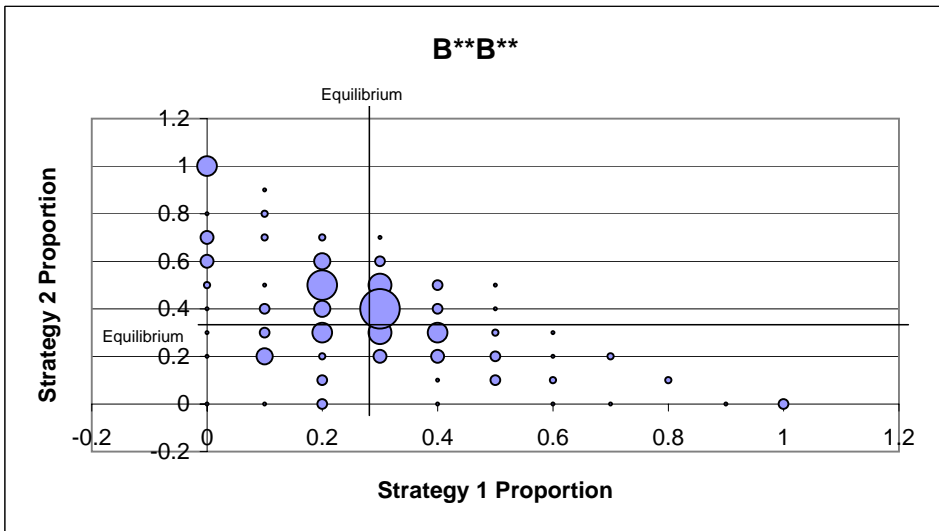
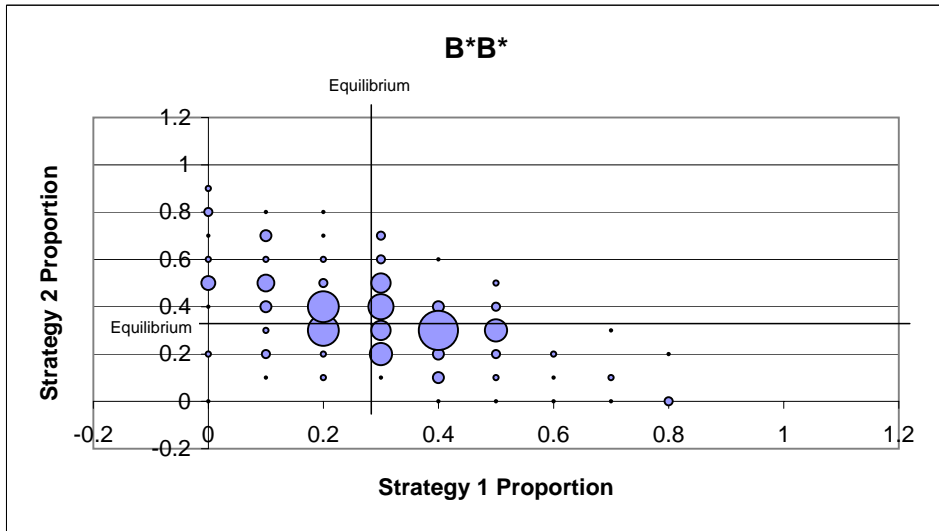


Figure 2: Proportion of Play Within Blocks of Ten Consecutive Games  
Rounds 1 - 50, AB Matrix Treatment

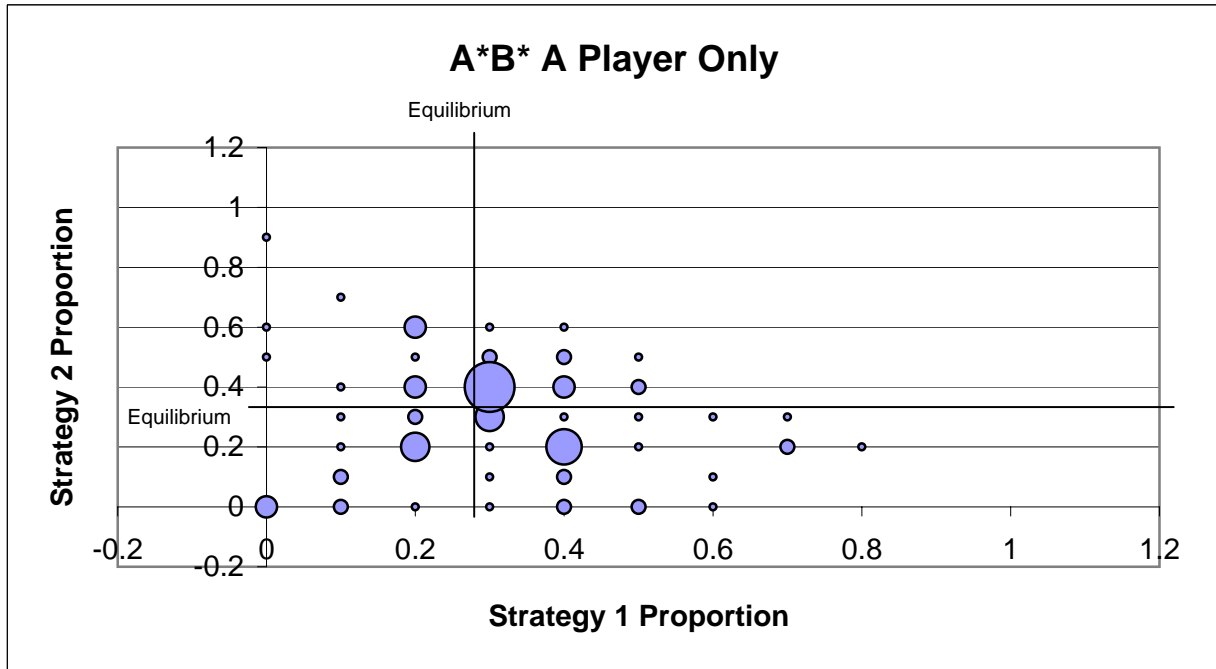


Figure 3: Proportion of Play Within Blocks of Ten Consecutive Games  
Rounds 51 - 100, B Matrix Treatments

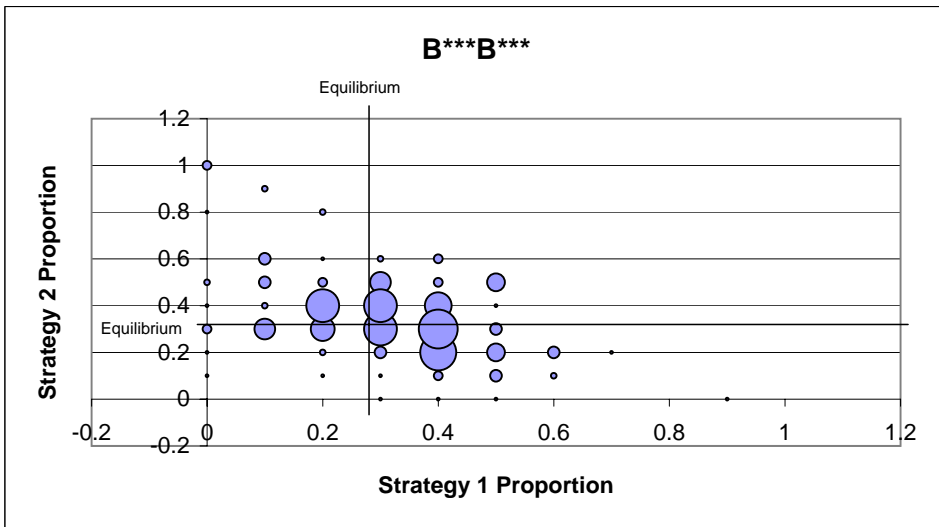
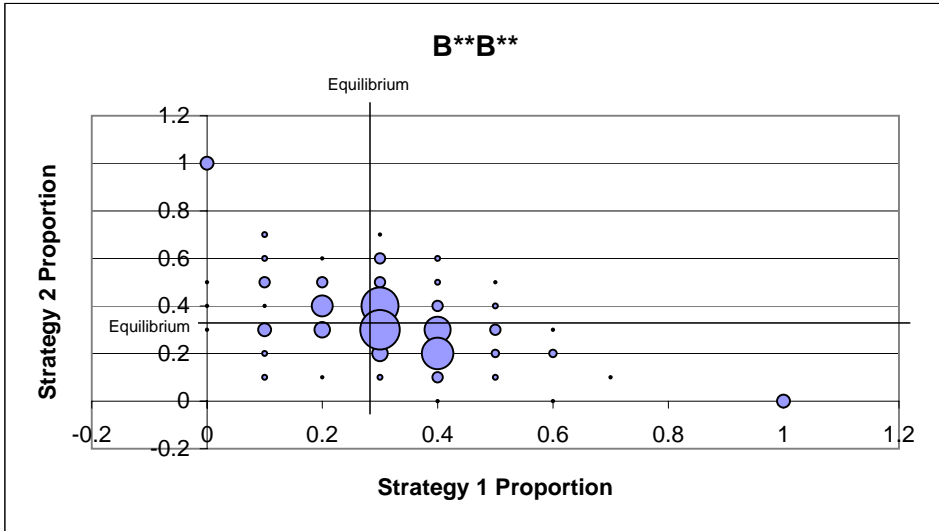
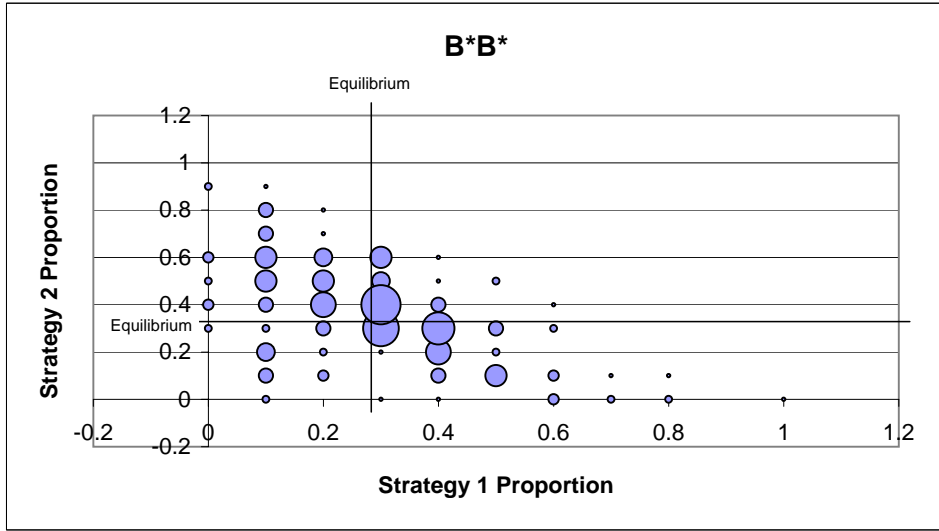


Figure 4: Proportion of Play Within Blocks of Ten Consecutive Games  
Rounds 51 - 100, AB Matrix Treatment

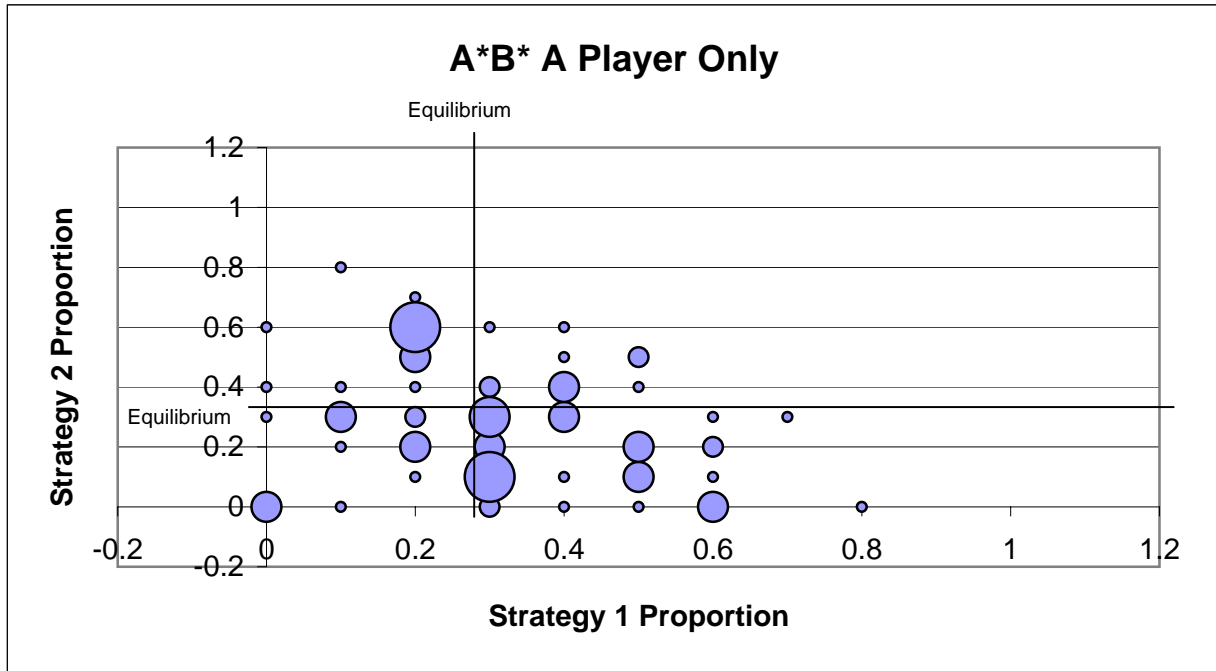
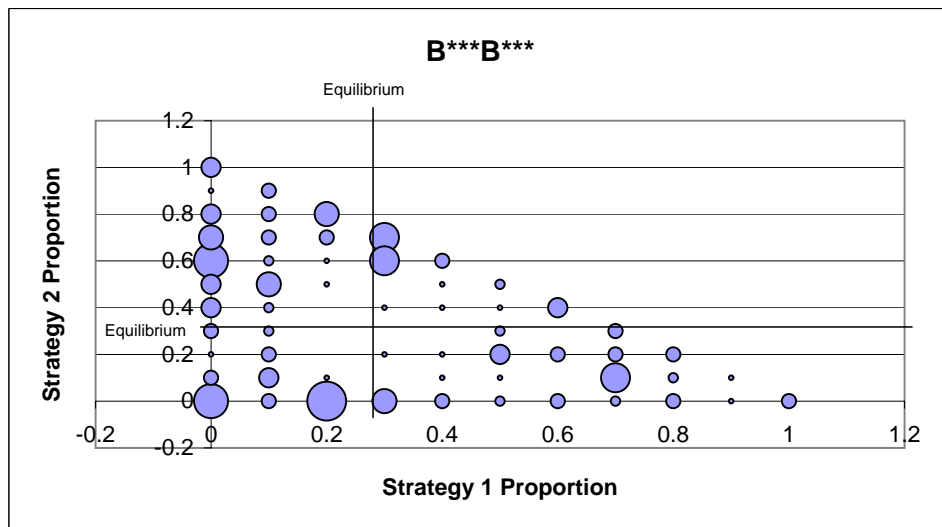
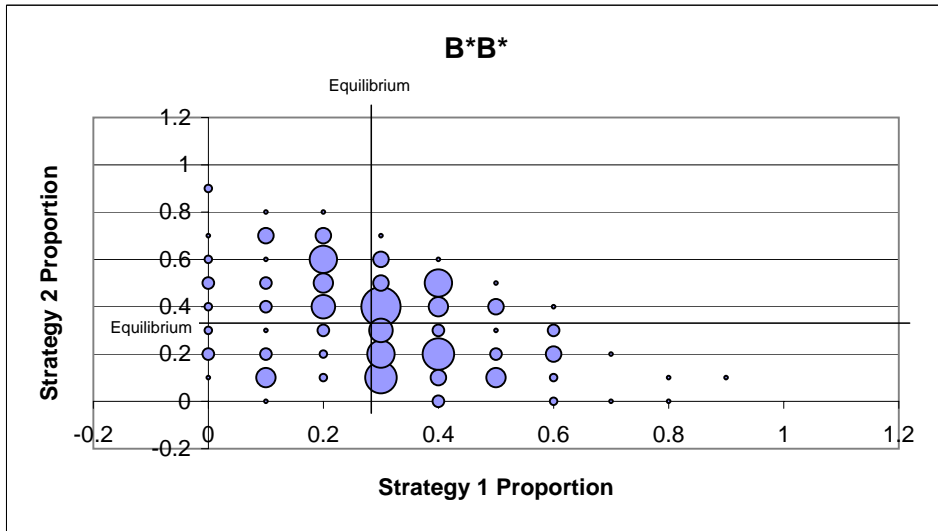


Figure 5: Proportion of Play Within Blocks of Ten Consecutive Games Rounds 51 - 100, B Matrix Treatments, Simulations



**Figure 6: Proportion of Play Within Blocks of Ten Consecutive Games  
Rounds 51 - 100, A and B Matrix Treatment, Simulations**

