C.E.S.S. Working Paper Series

CESS
center for experimental social science

Working Paper #2007-08

Convergence: An Experimental Study of Teaching and Learning in Repeated Games

Wolf Ze'ev Ehrblatt, Kyle Hyndman, Erkut Y. Ozbay, and Andrew Schotter

Key Words: Game Theory; Belief Formation; Learning; Convergence
JEL Classification: C70, C91, D83, D84

19 West Fourth Street, New York, New York 10003, 212-992-8660

The Center for Experimental Social Science

The Center for Experimental Social Science (C.E.S.S.) is an inter-disciplinary research Center at New York University engaged in laboratory experimental work in the social sciences that combines high quality economic theory, psychology and political science. The over arching research tool used in the Center is laboratory experiments in which human subjects engage in decision-making tasks whose purpose is to test both economic theory and/or the properties of proposed or existing economic, political, or social institutions. The working papers in this series reflect these interests.

Convergence: An Experimental Study of Teaching and Learning in Repeated Games*

Wolf Ze'ev Ehrblatt
New York University

Kyle Hyndman
Southern Methodist University

Erkut Y. Özbay
New York University

Andrew Schotter
New York University

25th November 2006

Abstract

Nash equilibrium can be interpreted as a steady state of a game where players hold correct beliefs about the other players' behavior and act rationally. In this paper, we experimentally examine the process that leads to this steady state. We find: (1) A teacher is needed for successful convergence. (2) Non-convergence appears to be the result of faulty beliefs and excessively sluggish updating, rather than inability to best respond. (3) When teaching becomes difficult, *e.g.,* when players cannot observe other players' payoffs, convergence becomes rare. (4) A successful model of belief formation should include a component taking a player's opponent's payoffs into account.
Keywords: Game Theory, Belief Formation, Learning, Convergence.
JEL Classification: C70, C91, D83, D84

1 Introduction

Nash equilibrium can be interpreted as a steady state of a game where players hold correct beliefs about the other players' behavior and act rationally. But how do players achieve

*This paper has benefited from comments from participants at the C.E.S.S. Experimental Design Workshop, the NYU-CREED Graduate Student Workshop in Amsterdam, the 2005 ESA Meetings and seminar participants from Johns Hopkins University and University College London. We are also grateful to Colin Camerer, Guillaume Fréchette, Ed Hopkins, David Levine and three anonymous referees for valuable suggestions. This research was supported by C.E.S.S. at NYU and by the NSF under grant SES-0425118 (Schotter). A previous version of this paper circulated under the title of "Convergence: An Experimental Study." *Corresponding Author:* Andrew Schotter, Department of Economics, New York University, 269 Mercer St. 7th Floor, New York, NY 10003. E-mail: `andrew.schotter@nyu.edu`.

this steady state, i.e., how do they learn to play a Nash equilibrium? Is it a belief-led process where people's beliefs converge and then their actions follow (because of best response behavior) or do actions converge first and then pull beliefs into line? In this paper, we experimentally examine this question.

If we relied on economic theory for guidance, the answer would be that typically either beliefs converge first or both beliefs and actions converge simultaneously. This is true because most belief-learning theories constructed by economists depict behavior in repeated games as a backward-looking, myopic and mechanistic (BLM&M) process in which beliefs are formed using historical data on the actions of one's opponents, and behavior is described as some type of deterministic or stochastic best response (*e.g.,* fictitious play). In such a world, it would be hard for an agent to choose the equilibrium action unless his or her beliefs had already arrived in that portion of the beliefs simplex where the Nash action was a best response (what we will call the Nash Best-Response Belief Set).

Another common feature of learning models is that they tend to ignore the payoffs of one's opponents. For example, reinforcement learning models take into account only a player's own payoffs while belief learning models take into account only an opponent's previous actions. EWA models take into account both an opponent's previous actions and one's own payoffs (either real or counterfactual), but not the payoffs of one's opponent. Models that fail to take into account an opponent's payoff have been called uncoupled by Hart and Mas-Colell (2006) who demonstrate that in infinitely repeated games there is no uncoupled learning model that guarantees convergence to a pure Nash equilibrium of the stage game. Furthermore, Young (2006) has argued that it is the lack of knowledge about one's opponent's payoffs that is the crucial factor in this impossibility result and not an inability to perfectly best respond to one's beliefs. In this paper we introduce a belief formation model that is capable of taking an opponents previous actions and payoffs into account and show that when subjects have access to the payoffs of their opponent, they use this information in their belief formation process and so behave in a more sophisticated (less myopic and mechanical) manner.

Our experiments lead us conclude that for the 3×3 games we examine, convergence is an action-led process in which actions converge before beliefs. As we will see, this means that in order to get to an equilibrium some player must choose the Nash action repeatedly even when his beliefs are outside of the Nash best response belief set. Such a player we will call a "teacher" and one of the main punch lines of this paper is that, in the class of games we investigate, the existence of a teacher greatly facilitates convergence.

Given the apparent importance of teaching in facilitating convergence, we wanted to examine the robustness of this result. To do this we ran a total of three treatments. In our baseline experiment, subjects were in fixed pairs for 20 periods and had full knowledge of the payoffs (own and opponent) in the game being played. This is the environment most kind to teaching. We ran two other treatments in which teaching was substantially more difficult, if not impossible. In one treatment, subjects were randomly re-matched with a new opponent each period, while in another subjects were in fixed pairs but only had knowledge of their own payoff and not the payoff of their opponent. For learning models which are BLM&M, such as Boylan and El-Gamal (1993), Cheung and Friedman (1997), Erev and Roth (1998), Camerer and Ho (1999), along with evolutionary models such as those discussed in Weibull (1995) and Samuelson (1997), these changes should make no difference in the play of the game since myopic behavior is only enhanced when one is not playing against a fixed opponent and the payoff of one's opponents is not an input into any of these models. On the contrary, if teaching is necessary for convergence we would expect that games played under these conditions would generally fail to converge. This is indeed what we find.

The results above focus on convergence. However our experiments also comment on what happens when players fail to converge. More precisely, we demonstrate that subjects fail to converge not because they are unable to best respond (convergent and non-convergent pairs of players do so equally often) but rather they fail to converge because of the sluggish way such subjects update their beliefs using historical data. More precisely, since convergence requires a teacher, if the follower in the game learns too slowly, the teacher tends to give up on such costly teaching and revert to myopically best responding to the follower's actions.

While we are not the first to focus on teaching and sophisticated play, the literature is rather limited. To demonstrate this it is interesting to note that in the classic treatise on learning by Fudenberg and Levine (1998) only the last chapter is devoted to what they call "sophisticated learning," and only the last three pages are devoted to teaching — the focus of our paper. Camerer, Ho and Chong (2002) extend their EWA model by assuming that a certain portion of population is sophisticated and that the sophisticated ones use EWA to forecast the behavior of the adaptive ones. Additionally, they introduce another parameter which represents the weight of future payoffs. They defined a teacher as the player who takes into account the effects of current actions. Their paper differs from ours in two important respects. First, we elicit beliefs from subjects and so can directly observe teaching behavior by contrasting actions with elicited beliefs. Second, their paper is largely concerned with how a player learns in a "population" environment, while we are concerned with the strategic interaction of two players. Hence, the difference between a non-myopic player and a teacher in their formulation is not clear. Additionally, they cannot identify the reason for non-convergent behavior.

In an interesting and related paper, Terracol and Vaksmann (2006) study teaching in games with multiple Pareto efficent equilibria. They argue that agents try to teach their opponents to play that equilibrium which is best for them. In contrast, our focus is on the role of teaching in facilitating convergence a unique pure strategy equilibrium.

The rest of the paper proceeds as follows. In Section 2 we provide details of the experimental design. In Section 3 we will start by discussing our results in the AP treatment first by describing behavior in our experiments and then examining our main contention that teaching greatly facilitates convergence. In our discussion we move from descriptive analysis to more formal econometric analysis as we progress. We then move on to discussing behavior in our OP and RM treatments and demonstrate that since teaching is more difficult here, convergence is less frequent. We next present a model of belief formation that is both descriptive and predictive of how people form beliefs. Finally, in Section 4 we offer some concluding remarks.

2 Experimental Design and Procedures

In order to answer the questions posed at the beginning of this paper we conducted a number of different experiments, the details of which are given in Table 1. All experiments were run on inexperienced subjects recruited from the general undergraduate population at New York University. The experiments were run in the laboratory of the Center for Experimental Social Science and typically lasted 1 to 1.5 hours. Subjects' mean payoffs were \$19.14 across all treatments, not including a \$7.00 participation fee.

In the first treatment, called the AP (All-Payoff) Treatment, subjects played one of the games depicted in Figure 1 for 20 periods with a fixed partner and with the payoffs of both players visible. They were then randomly rematched and played the other game in Figure 1 for 20 periods. Figure 1.a depicts a dominance solvable game (DSG), with a unique Nash equilibrium which is in pure strategies. In contrast, Figure 1.b presents a game which is not dominance solvable (nDSG) but which has a unique pure strategy Nash equilibrium.[1]

Figure 1: Games Used in the Experiments

	$A1$	$A2$	$A3$
$A1$	$51, 30$	$35, 43$	$93, 21$
$A2$	$35, 21$	$25, 16$	$32, 94$
$A3$	$\mathbf{68, 72}$	$45, 69$	$13, 62$

(1.a) DSG

	$A1$	$A2$	$A3$
$A1$	$12, 83$	$39, 56$	$42, 45$
$A2$	$24, 12$	$12, 42$	$\mathbf{58, 76}$
$A3$	$89, 47$	$33, 94$	$44, 59$

(1.b) nDSG

[1] The game in Figure 1.b actually has two mixed strategy equilibria, one of which is fully mixed, while the other requires subjects to place zero weight on their pure Nash equilibrium action. We focus our attention on the pure strategy equilibrium and find no evidence for convergence to either of the two mixed strategy equilibria. We will discuss this later in Section 3.

Table 1: Summary of Experimental Treatments

Treatment	Task	Game(s)	# Subj's	Matching	Payoffs	# Per's
AP	Beliefs/Actions	DSG/nDSG	64	fixed	both vis.	20/20†
RM	Beliefs/Actions	DSG	20	random	both vis.	20+40‡
RM	Beliefs/Actions	nDSG	20	random	both vis.	20+40‡
OP	Beliefs/Actions	DSG/nDSG	72	fixed	own vis.	20/20+40♯

† Subjects played one game for 20 periods and then (after being rematched) the other game for 20 periods.
‡ Subjects played for an initial 20 periods and were then asked to play 40 more periods.
♯ Subjects played one game for 20 periods, and then (after being rematched) the other game for 20+40 periods as in ‡.

The games chosen had the following features:

- A unique pure strategy Nash equilibrium in the stage game.
- Nash payoffs are on the Pareto frontier.
- Payoffs in the Nash equilibrium were not symmetric.

In each period, the subjects were asked to make two decisions. The first was to choose the action for that period. The second was to state their beliefs regarding their partner's action in that period. The action decision was rewarded according to the relevant game matrix, while the belief reports were rewarded using the Quadratic Scoring Rule (QSR).[2] All payoffs from the action choices and the belief predictions were then summed up to give subjects their final payoff. This treatment has two features: subjects could see both their and their opponent's payoffs and were matched in fixed pairs for the full 20 period length of the experiment. The AP treatment serves as our baseline, and will be our main focus.

[2]Under the assumption the subjects are risk neutral, the use of the QSR should make the subjects state their true beliefs regarding their opponent's action. Sonnemans and Offerman (2001) find that the QSR is incentive compatible and that subjects tend to report their true beliefs when the QSR is used. Nyarko and Schotter (2002a) also use a quadratic scoring rule and offer substantial evidence that subjects best respond to the beliefs they state. Moreover, Wilcox and Feltovich (2000) report that belief elicitation does not always affect subjects' behaviour. However, Rutström and Wilcox (2004) argue that the act of solicitation may focus the attention of the subject on his or her beliefs in a way that may be unnatural.

In addition to the AP treatment we also ran two others to help substantiate our conclusions. In the RM (Random Matching) treatment, subjects were randomly matched each period over the 20 period horizon of the experiment — however, they kept their same role, as row or column player, throughout. They were informed only of the outcome of their interaction at the end of each period. (This is one of the three ways in which random matching feedback could be given (see Fudenberg and Levine (1998 pp. 4–7) and Hopkins (2002)). In contrast to the AP treatment, subjects only played either the DSG or nDSG game. After the initial 20 periods were completed, we surprised the subjects and told them that the experiment would last for 40 more periods. This was done to check if behavior would change if the horizon was increased. In our third treatment, OP (Own-Payoff), we replicated the conditions of the original AP treatment except that subjects are only able to see their own payoffs and *not* the payoffs of their opponent. As with the RM treatment, we surprised subjects after their final 20 period interaction and asked them to play the game for 40 more periods.

As we mentioned in the introduction, these treatments were run in order to better isolate the role of teaching. If teaching is important for convergence and we make it more difficult to teach, then we should see less convergence. In contrast, if the BLM&M models are correct, having a random matching protocol should highlight myopic play and *increase* the rate of convergence relative to the AP treatment.[3] Similarly for the OP treatment: for most of the popular learning theories (*e.g.,* reinforcement learning, EWA, fictitious play, noisy fictitious play, *etc.*), the elimination of one's opponent's payoffs should have no impact on behavior or convergence rates.[4]

[3] As Fudenberg and Levine (1998, p. 4) point out, with fixed pairs, subjects may think "that they can 'teach' their opponent to play a best response to a particular action by playing that action over and over."

[4] For more on how behaviour is different when players do not have access to their opponent's payoffs, see Partow and Schotter (1993), Mookherjee and Sopher (1994) and Costa-Gomes, Crawford, and Broseta (2001).

3 Results

3.1 Definitions

In this section we will describe the difference in the behavior of pairs of subjects whose play converged to the Nash equilibrium over the 20 periods in which they interacted in the AP treatment with those who did not. Our aim will be to demonstrate that teaching is essential for convergence.

Before we begin let us first define what we mean by convergence in actions (and beliefs) to the Nash equilibrium and also define what we mean by a "teacher". We consider a player i playing a two-person finite-strategy game repeatedly to have converged in actions to the Nash equilibrium, $a^* = (a_1^*, a_2^*)$ at time $t_i^{a^*}$ if $t_i^{a^*} \leq 18$ is the earliest time that player i played the Nash equilibrium and continued to do so for the rest of the game. In other words, for his actions the Nash equilibrium is an absorbing state and $t_i^{a^*}$ is the first entry point into that state. In a two player game if player i converges in actions to the Nash equilibrium in period $t_i^{a^*}$ and player j does so in period $t_j^{a^*}$, then we say the game converges in period $\tilde{t} = \max\{t_i^{a^*}, t_j^{a^*}\}$.

To describe convergence in beliefs let $b_{i,t}$ denote player i's belief about player j's period t action choice. We say that player i's beliefs have converged in period t_i^b if for all time periods greater than or equal to t_i^b, player i's beliefs remain inside the Nash Best Response Belief Set. In other words, a player's beliefs have converged once they enter the set of beliefs for which the Nash action is a best response and never leave until the game ends.

These definitions are the strictest definitions we can use and make convergence the hardest to achieve. We feel they are less ad hoc than other possible definitions, however, since for any other definition we could use we would have to call a play path of actions convergent even if at some points along the path subjects would not be playing their Nash actions. In our definition, once a game converges it converges and no deviations are allowed.[5]

Finally, we call a member of a convergent pair a teacher (early converger) if his actions converged first to the Nash equilibrium and, before convergence of the game, he chose the Nash action repeatedly even though that action was not a best response to his beliefs. This differentiates a teacher from a person who is simply best responding to his beliefs and happens to choose the Nash action before his opponent. We call a subject a follower (late converger) if his actions converged second.

3.2 The AP Treatment: Teaching, Convergence and Non-Convergence

3.2.1 Teaching and Convergence

We begin with a thorough analysis of the results from our AP treatment. We first descriptively differentiate between convergent and non-convergent pairs of subjects and then continue with a more formal econometric analysis. We start with the AP treatment because, of all the environments we considered, it is the most conducive to teaching since both players knew the entire payoff matrix and were paired for the entire 20 period interaction. Hence, if teaching is essential for convergence, we should see it here.

Using the definition of convergence above, 17 of 32 pairs converged in the dominance solvable game, while 16 of 32 pairs converged in the non-dominance solvable game. Figure 2 presents all of those pairs whose actions converged to the Nash equilibrium and shows the time period each member of the pair converged in actions. Obviously, in each pair one subject converged early and one later.[6] Notice that there are a number of pairs (pair numbers 9, 14

[5]There were three instances of subjects having a high frequency of Nash equilibrium play, with a final period deviation that we labeled as convergent. There was also one pair that we labeled as non-convergent because one of the pair members only chose his/her Nash action in periods 19 and 20, despite his/her opponent having played the Nash action from period 1.

[6]Or they converged simultaneously; however, their behavior is qualitatively identical to early convergers and so we group them together in our analysis.

and 15 in DSG and 4, 5, 6, 12 and 14 in nDSG) where there were long periods of apparent teaching before the follower converged. The reader can also see substantial variation in the time at which convergence occurs.

Looking only at the actions data of Figure 2, we cannot conclude that teaching was actually occurring. To say that teaching was happening, we must first demonstrate that for early convergers, actions converged before beliefs, while to conclude that learning was occurring, it should be that for late convergers, beliefs and actions converged at approximately the same time. In Table 2 we divide all those whose actions converged to Nash equilibrium into two distinct subgroups — those who converged first (in a given pair) on their Nash action and those who converged second on their Nash action — and calculate the mean difference between the period in which beliefs and actions converged.

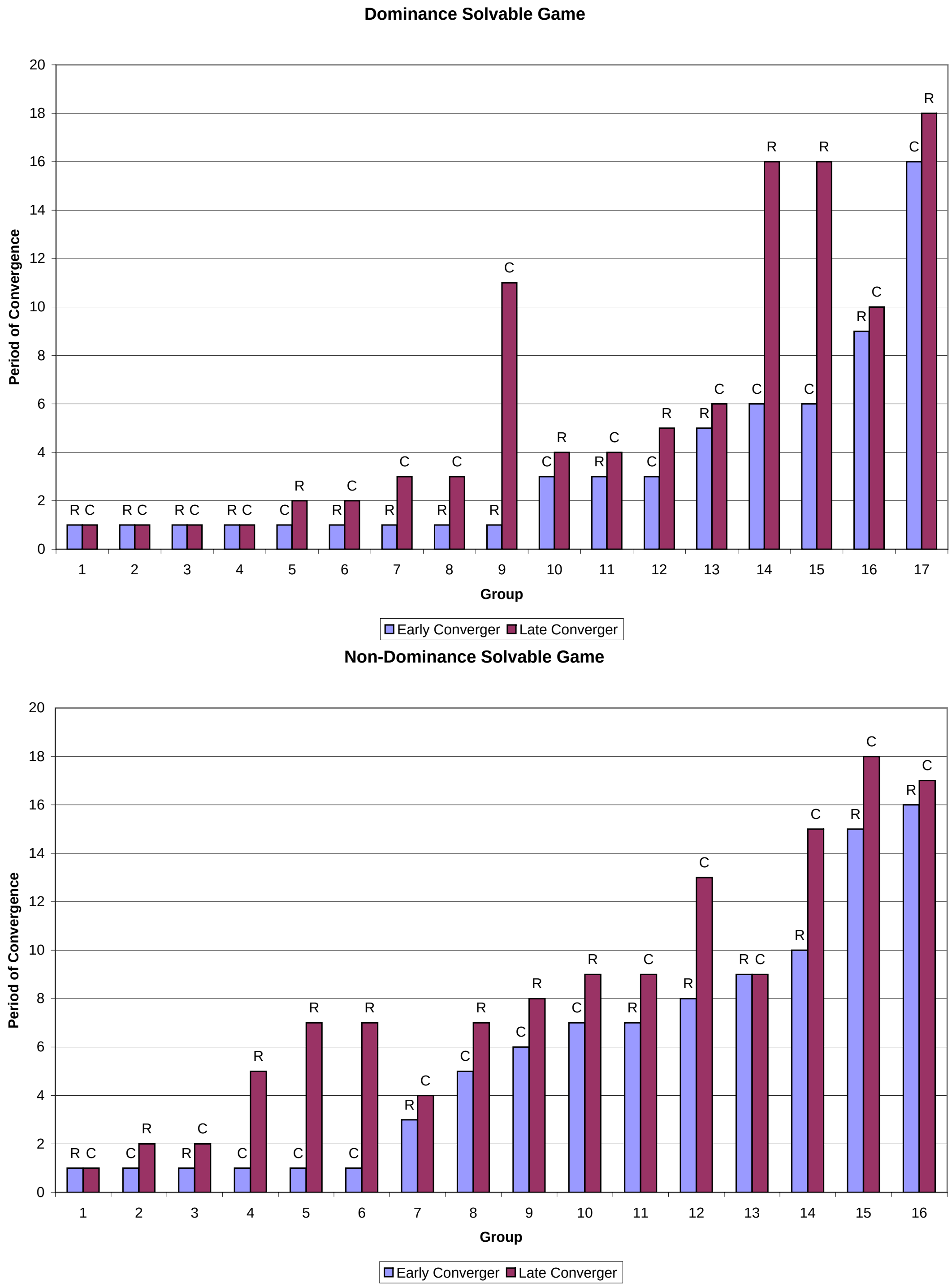
Table 2: Mean Difference in Convergence Times of Beliefs and Actions: Early and Late Convergers in DSG and nDSG Games

	DSG		nDSG	
	Early Converger	Late Converger	Early Converger	Late Converger
$\mu_{t^b - t^a}$	3.95	0.69	3.17	-0.14
N	21	13	18	14

† $\mu_{t^b - t^a}$ is the mean difference between the period of convergence in beliefs and actions.

Indeed, those who converged to equilibrium first are, in fact, teachers: their actions converged to equilibrium *much* before beliefs converged to the Nash Best Response Belief Set. For example, in the dominance solvable game, the difference between belief and action convergence for early convergers is approximately 3.95 periods, while for the non-dominance solvable game, it is 3.17 periods. On the other hand, those who converged to their part of the Nash equilibrium second appear to be followers: the difference between the period of convergence in actions and beliefs is not distinguishable from zero. Note that for both games, the Mann-Whitney Rank Sum Test allows us to reject the null hypothesis that early convergers and late convergers come from the same population in favor of the alternative that early convergers have beliefs which converge after actions, while late convergers do not

Figure 2: The Period of Convergence in Actions: AP Treatment



($Z_{DSG} = 2.605$ and $Z_{nDSG} = 3.936$).

To reinforce the idea that early convergers are teachers according to our definition, consider the fact that for 36 of the of the 39 subjects who converged either first or simultaneously, their actions converged *strictly* before their beliefs: In two cases, beliefs and actions converged simultaneously, while in another beliefs converged one period before actions.[7] In other words, 36 of the 39 early convergers strictly satisfied our definition of what a teacher is while 2 satisfied it weakly.

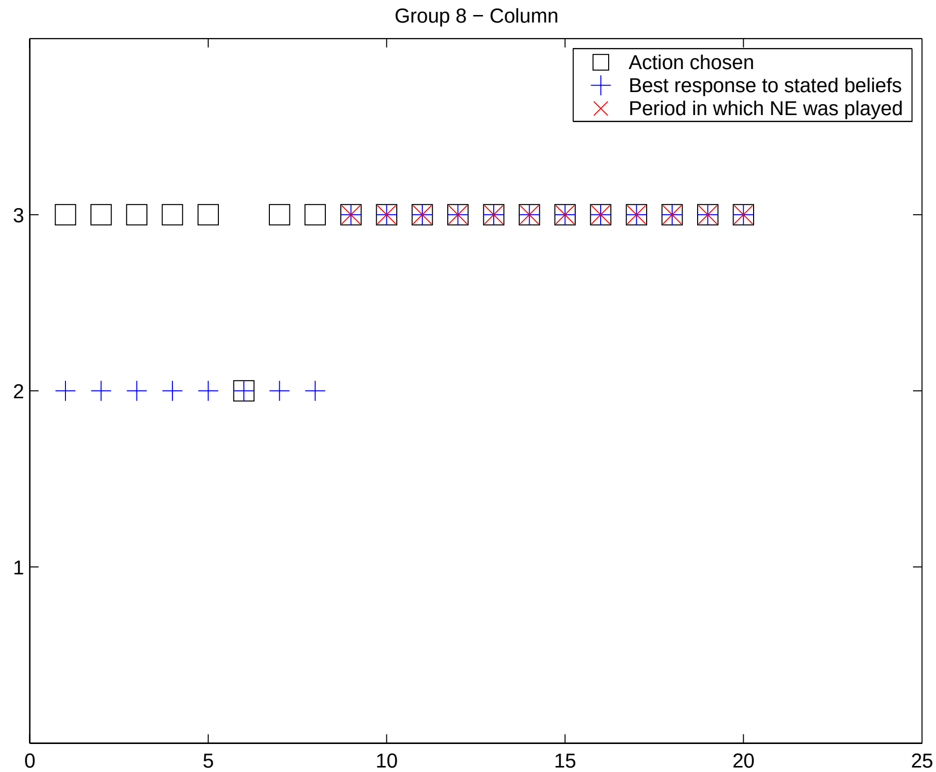
Another interesting fact is that teachers appear to be born and not raised. What we mean by this is that whether a subject becomes a teacher in our experiment is highly correlated with that person's first period action. In the dominance solvable game, 15 out of 21 subjects that we classify as teachers or simultaneous convergers chose their Nash action in period 1 as opposed to only 3 of the 13 followers, while in the non-dominance solvable game these numbers were 10 out of 18 as opposed to 3 out of 14. For both the DSG and nDSG games, a two-sample proportions test reject the null hypothesis that teachers and followers are as likely to play the Nash action in the first period: $Z_{DSG} = 2.745$ ($p < 0.01$) and $Z_{nDSG} = 1.95$ ($p = 0.051$), respectively.

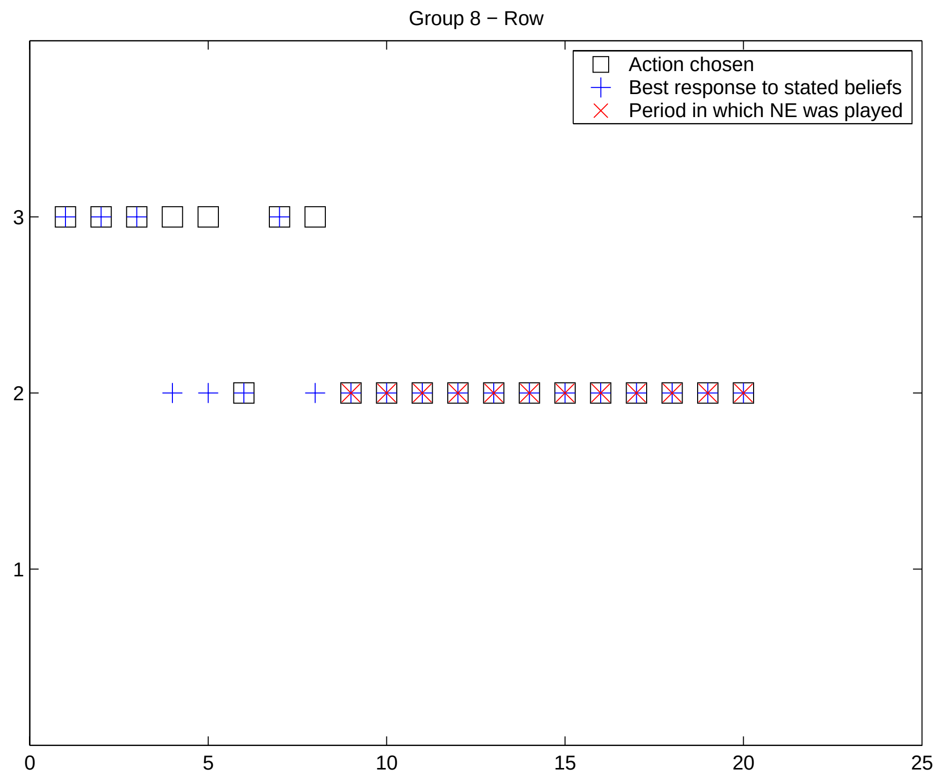
3.2.2 A Teaching Episode

We would like to present one illustration of the process through which teaching leads to convergence in order to give you a flavor of what goes on in a teaching episode. To do this consider Figure 3, which shows the history of one convergent pair in the nDSG game. While the time series for other convergent pairs would obviously look different, they all tell a similar story: teachers recognize the Nash equilibrium relatively early and choose their part of it repeatedly in an effort to teach. The only question on their part is whether their opponent understands what they are trying to teach.

[7] All of these three cases were for players who converged simultaneously.

Figure 3: Actions, Beliefs and Best Response: A Typical Teacher and Follower





In this figure we place the period number along the horizontal axis and the action taken by the row or column player along the vertical axis. Boxes indicate the action chosen by the subject in any given period, an $\times$ indicates whether the Nash equilibrium was played in that period and $+$ indicates the action that was a best response to the subject's stated beliefs in that period. Hence in any period where we see an empty box we know that the action taken was not a best response to the subject's beliefs, while if we see a box with a $+$ but not an $\times$ inside we know that the action taken was a best response but at least one subject did not play their Nash action. When both beliefs and actions have converged we should see boxes with both $\times$'s and $+$'s in them.

Since this pair of subjects played the non-dominance solvable game, the equilibrium actions are action 3 for column and action 2 for row. There are many interesting features of this interaction that are illustrative of our point. First, notice that according to our definition of convergence, this game converges in period 9. In this pair, the column player is the teacher and starts to play his Nash action in period 1, despite the fact that he does not think his opponent will reciprocate, and continues to do so until period six even though his teaching has failed to get the row player to budge from choosing action 3. In period 6 he gives up and chooses action 2, which is a best response to his beliefs in that period. This might have ruined this pair's chances of convergence except for the fact that in period 6 the row player seemed to get the point and chose his Nash action. Seeing this, the column player resumes teaching by reverting back to action 3 again, despite the fact that he still is not convinced that row has understood the Nash equilibrium. Finally, in period 9 the game converges.

According to our definitions we would claim that the teacher's actions converged in period 7 and his beliefs in period 9 while the followers actions converged in period 9 and his belief in period 8. This classification does us a disservice since it is clear that if we do not count the period in which the teacher gave up (period 6), his actions actually converges in period 1. Also, note that up until convergence the teacher best responded to his beliefs only once, while the follower best responded in five out of nine periods. This demonstrates that teaching behavior is clearly different than follower behavior which tends to be much more like best-response behavior.

3.2.3 Non-Convergence

Until now we have spent our time on convergence and the role that teachers play is fostering it. However, half of our pairs in the AP treatment failed to converge so it would be interesting to discover why. It is our claim that a failure to converge is a failure, first and foremost, of belief formation and not of an ability to best respond. As can be seen in Table 3, which presents the fraction of times subjects either best responded or chose a second or third best response to their elicited beliefs, those who converged best responded in approximately the same proportions as those who did not, and approximately in the same proportions across games. Instead, what non-convergers do wrong is to update too sluggishly, *i.e.,* put to much weight on distant history. This sluggishness creates two problems. First, if a subject's beliefs are far from the Nash Best Response Belief Set, slow adjustment means it will take a long time for those beliefs to converge. Second, if a teacher is trying to lead the way to the Nash equilibrium, sluggish responses can cause the teacher to give up. So we conjecture that those subjects that converged would be those who give more weight to the recent past and best respond to those beliefs, while those who do not converge put too little weight on what recently happened and also best respond to their (out-of-equilibrium) beliefs.

Table 3: Percentage of Best Responses- Non-Convergent vs. Convergent up to Convergence Period

	DSG Game		nDSG Game	
	Convergent	Non-Convergent	Convergent	Non-Convergent
Action was a B.R.	53.8	52.2	51.9	60.0
Action was a 2^{nd} B.R.	26.9	29.4	35.5	25.8
Action was a 3^{rd} B.R.	19.2	18.8	12.6	14.2

We will make this all more formal in the next subsection but before we do let us pause and take a look at the beliefs of subjects who did and did not converge. Consider Figure 4 where we present the simplex of beliefs and the time paths of beliefs of two subjects whose play failed to converge in the nDSG game. The point (0,0) represents the case in which a player holds degenerate beliefs that her opponent will play his Nash strategy. The beliefs on the two non-Nash actions are then given by a point in the (x, y) plane and the area enclosed by the dashed line represents the Nash Best-Response Belief Set. That is, if beliefs lie inside this set, it is a best response for the player to choose *her* Nash action. Finally, the points labeled S and F depict where the subject's beliefs started and finished in the simplex; when it is not clear, an arrow points to the direction in which the subject updated his/her beliefs.

What is obvious in these figures is that for these players, who are very representative of all players who failed to converge, throughout the entire game their beliefs never entered into the Nash Best Response Belief Set. In other words, if subjects are capable of best responding and best respond to the beliefs we elicited, then this is clear proof that failure to converge is a result of a failure in beliefs not in actions. To contrast this to representative convergent pairs, consider Figure 5 where we show two typical sequences of beliefs for players whose actions converged. Here, after some initial periods outside the Nash best-response set, the beliefs of the subjects enter the set and very shortly become degenerate.

The subjects depicted in these figures are the rule and not the exception: there is a dramatic difference in the frequency with which beliefs enter the Nash Best Response Belief Set when comparing pairs that converged to those that did not. For convergent pairs their beliefs were in the Nash Best Response Belief Set 79.7% and 68.0% of the time in the dominance solvable and non-dominance solvable games, respectively, while for non-convergent pairs these same percentages were 18.7% and 12.5%, respectively. This is strong evidence that the beliefs of those players who do not converge to Nash equilibrium spent very little time inside the Nash Best Response Belief Set. Also interesting is the fact that beliefs, once inside the Nash Best Response Belief Set, are often degenerate on their opponent's Nash action for convergent pairs (DSG: 78.0%, nDSG: 80.1%), while beliefs are much less often

Figure 4: Belief Data, A Non-Converging Pair — nDSG

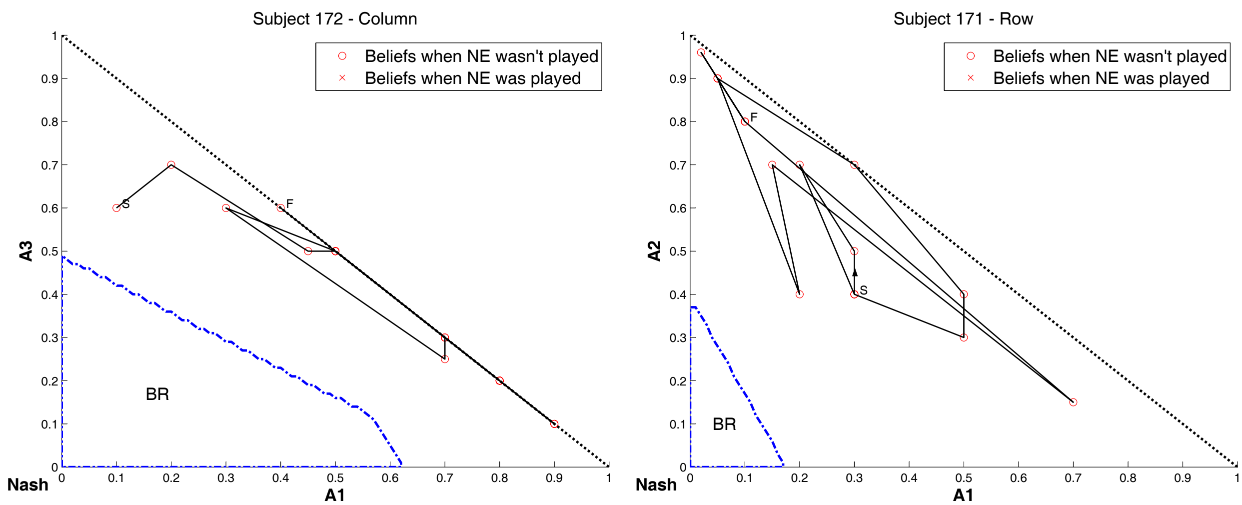
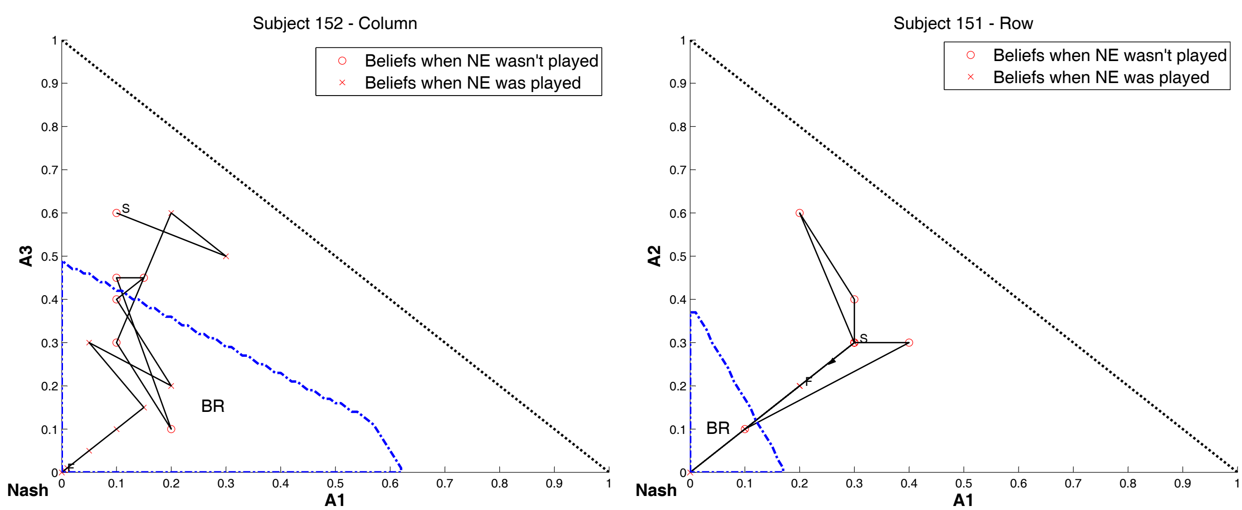


Figure 5: Belief Data, A Converging Pair — nDSG



degenerate for non-convergent pairs (DSG: 39.0%, nDSG: 35.2%).

3.2.4 A Formal Analysis of Convergence and Non-Convergence

The results above indicate that the belief formation process is a key element in whether player pairs converge. What appears to be needed for convergence is a teacher along with a follower who updates quickly. We will formalize this conjecture into a hypothesis that we test in this section but in order to state this hypothesis correctly, let us pause to define what we mean by fast and slow convergers, and to look more carefully at teaching.

Suppose that people form beliefs based on the history of play, say according to Cheung and Friedman's (1997) $\gamma-$weighted beliefs. That is, the belief of player i that j will choose strategy k in period $t+1$ is given by:

$$\Gamma_i^k(t+1) = \frac{1_t(a_k^j) + \sum_{u=1}^{t-1} \gamma^u 1_{t-u}(a_j^k)}{1 + \sum_{u=1}^{t-1} \gamma^u}. \tag{1}$$

Here $1_t(a_k^j)$ is an indicator function which takes on the value of 1 when j plays action k in period t and 0 otherwise. This is a "historical" belief function which places geometrically declining weights on past actions of a player's opponent. We use the Cheung-Friedman model as a diagnostic tool to evaluate the manner in which our subjects use history in their updating and not necessarily as an accurate description of how people form beliefs. Clearly, the smaller a subject's γ the faster he or she updates. Those subjects with γ's close to 1 (ficticious-play beliefs) weigh all observations equally and hence update slowly by not paying any extra attention to the recent past. Those with γ's near zero (Cournot beliefs) give heavy weight to recent events and hence adjust their beliefs quickly. Indeed, the closer γ is to 1, the less beliefs respond to recently chosen actions and the more sluggish they are.

To explain the relationship between the γ and pair convergence consider a teacher playing the Nash equilibrium for a couple of periods. If her opponent has a relatively low γ, her opponent's beliefs will update rather quickly and rapidly enter the best response set. If her opponent also best responds to his beliefs, then the game will converge. On the other hand, suppose that the teacher's opponent has a relatively high γ. In this case, his beliefs will update only very slowly in response to teaching and may not enter the Nash Best Response Belief Set before the teacher finally gives up teaching. In such a case the game will not converge. Since many games, including those that do not converge, have teachers, one might conjecture that what separates successful from unsuccessful teaching is the γ of the follower in the pair.[8] If the follower has a low γ and the teacher is persistent, then we expect convergence while if the follower has a sluggish reaction to what he sees, a high γ, then we expect convergence to be less likely. These facts allow us to state our first hypothesis:

Convergence Hypothesis 1a — The γ distribution: *The distribution of γ's for non-convergers stochastically dominates the distribution of γ's for convergers.*

To estimate the γ used by each subject we take the sequence of elicited beliefs $\{b_{i,t}\}_{t=1}^{20}$ for each player i over the 20 period horizon of the experiment and compare it to what that sequence would have been if the subject formed their beliefs using the Cheung-Friedman γ-historical belief model which would produce, for a given γ, a sequence of beliefs denoted as $\{b_{i,t}(\gamma)\}_{t=1}^{20}$. We estimate γ by searching for that γ that minimizes the sum of squared prediction errors. That is, given a sequence of choices, $\{b_{i,t}\}_{t=1}^{20}$ we can find the γ that minimizes:

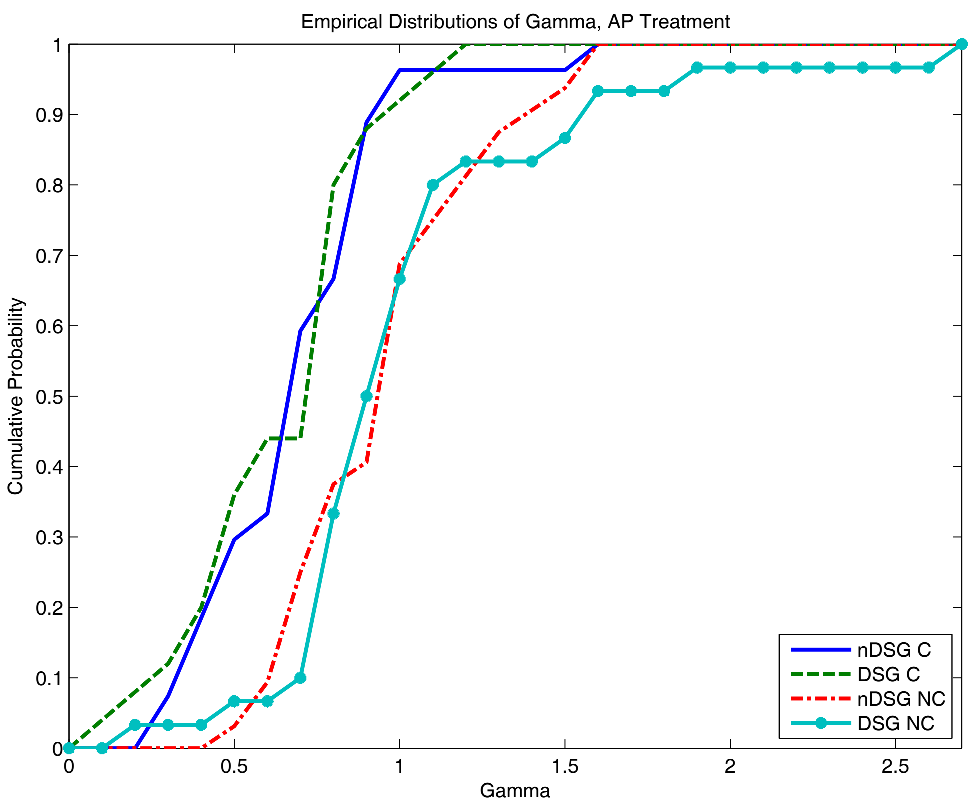
$$SSE(\gamma) = \sum_{t=1}^{20} \sum_{k=1}^{3} (b_{i,t}^k - b_{i,t}^k(\gamma))^2 \tag{2}$$

where we sum over all periods $t = 1, \ldots, 20$ and all three possible actions $k = 1, 2, 3$.

The results of our calculations are seen in Figure 6 where we present the cumulative distributions of the γ's we estimate subject by subject. There are several things to notice here. First, our hypothesis is immediately verified for the dominance solvable game and also for the non-dominance solvable game except on the extreme lower end of the support.

[8] Indeed, in our games, more than 30% of the players not part of a pair that converged to Nash equilibrium engaged in at least one unsuccessful teaching episode, defined as three or more consecutive periods of playing the Nash equilibrium, despite it not originally being a best response.

Figure 6: Empirical Distributions of γ: AP Treatment



That is, the distribution of γ's for non-convergers is skewed (relative to the distribution for convergers) towards higher values of γ. Second, Kolmolgorov-Smirnov tests allow us to reject the null hypothesis that the distributions for convergers and non-convergers are the same in both games that we considered ($D_{DSG} = 0.5467$, $p < 0.01$; $D_{nDSG} = 0.4826$, $p < 0.01$). Third, there is substantial variation in γ across players, regardless of game type and regardless of whether the game converges. Finally, the distribution of γ's across game types are not distinguishable for each other. This supports the idea that γ is based more on individual characteristics of the player than on the specific game being played.

While we prefer to focus on individual level data, we can also conduct a similar exercise pooling across players and estimate γ jointly with the subjects' best response precision. Since we also argue that a failure to converge is *not* a failure of best responding, we estimate a model of γ-weighted beliefs with stochastic best response (henceforth the γ/λ-model). More precisely, using the beliefs defined by (1), if we define the expected utility of choosing action k in period t as $\mathbb{E}_t[\pi(a_k, a_{-i})] + \epsilon_k$, for $k = 1, 2, 3$, where ϵ_k has a Type I extreme value distribution, with ϵ_i and ϵ_j for $i \neq j$ independently distributed, then we can define the probability that action k will be chosen in period t as:

$$\Pr[A_t = k] = \frac{\exp(\lambda \mathbb{E}_t \pi[(a_k, a_{-i})])}{\sum_j \exp(\lambda \mathbb{E}_t[\pi(a_j, a_{-i})])} \tag{3}$$

where λ measures the precision with which this player best responds.

Our convergence hypothesis can then be restated and expanded to include both γ and λ. Before stating our hypothesis, however, there is an important point to make: Strictly speaking, it does not make sense to estimate this model for teachers since, while they are teaching, teachers are explicitly not best responding to their stated beliefs, making the model potentially mis-specified for them. However, the model is well specified for followers and non-convergers since they are responding to what they see their opponent doing. Therefore, we will state our hypothesis in terms of followers and nonconvergers as follows:

Convergence Hypothesis 1b — The joint γ/λ hypothesis: *The estimate, γ^{FOL}, for followers should be less than the estimate, γ^{NC}, for non-convergent players. Moreover, the estimated λ's should not be different.*

We test this hypothesis on the pooled data of all subjects in each treatment. However, there is some question about what the appropriate sample to use in the estimation is. We argue that when comparing λ it only makes sense to use the data up to convergence. After convergence has occurred, all subjects are best responding, almost by default, and this would lead to abnormally high estimates of λ. On the other hand, the dependence on history of beliefs, and hence on γ, still seems important, even after convergence has occurred. As a compromise, we present results on both the full sample *and* on a restricted sample which considers the data up to two periods after convergence. For the estimates of γ, the results are not qualitatively changed. Table 4 presents the results of our estimation.

Table 4: Estimates of γ & λ: γ-weighted beliefs with Stochastic BR[†]

	nDSG (FOL)		nDSG (NC)	DSG (FOL)		DSG (NC)
	conv. + 2	all data		conv. + 2	all data	
λ	0.051	0.096	0.048	0.043	0.098	0.033
	(0.0082)	(0.0075)	(0.0043)	(0.0104)	(0.0096)	(0.0044)
γ	0.497	0.608	0.727	0.423	0.503	0.913
	(0.1565)	(0.0715)	(0.0513)	(0.2829)	(0.1178)	(0.0980)
n	151	280	640	126	260	600
LL	-137.91	-187.44	-620.28	-126.41	-216.15	-623.9

[†] Standard errors (in parentheses) were calculated using BHHH with a step size of 0.02 for λ and 0.2 for γ.

[‡] NC: non-convergent pairs; FOL: followers

The reader can see that Table 4 lends support to Convergence Hypothesis 1b. In particular, and consistent with Table 3, the estimated λ's for followers and non-convergers are not statistically distinguishable from each other. We base this result on the partial data set, which we argued was the relevant one for our the estimation of λ.[9] We also see that

[9]When we speak of statistical significance here, we do not use the standard errors calculated using BHHH, but instead based on the confidence regions calculated via a series of likelihood ratio tests. The

$\gamma^{FOL} < \gamma^{NC}$ for both the dominance solvable and non-dominance solvable game, and that this does not depend upon whether the full or partial sample is used. In terms of statistical significance, using the full sample (which, for γ, we have argued is the appropriate sample to use), for the dominance solvable game, the 95% confidence interval for followers is $\gamma^{FOL} \in [0.325, 0.665]$, while for non-convergers it is $\gamma^{NC} \in [0.73, 1.095]$; therefore, $\gamma^{FOL} < \gamma^{NC}$. For the non-dominance solvable game we do not have a statistically significant difference; the respective confidence intervals are: $\gamma^{FOL} \in [0.5, 0.705]$ and $\gamma^{NC} \in [0.63, 0.825]$.

3.2.5 Does it Pay to Teach?

From our earlier discussion, it should be clear that teaching is very prevalent in the data, and the presence of a teacher substantially facilitates the process of convergence. But this begs the question: "Does it actually pay to teach?" Note well that the answer to this question need not be in the affirmative: teaching represents an investment. For a number of periods, the teacher is explicitly not best responding to her beliefs, and, therefore, behaving sub-optimally in a static sense. However, by teaching, she may hope to converge to an action pair (e.g., Nash equilibrium) which gives a higher total payoff than would cyclic non-convergence. Therefore, we would expect that the payoffs to teachers is higher than the payoffs to non-convergers; that is, $\pi^T > \pi^{NC}$. However, to the extent that teachers are not best responding during teaching phases, while followers may well, one might conjecture that $\pi^F > \pi^T$. These considerations define the following null hypothesis.

Teacher Payoff Hypothesis: *There should be no difference in the payoffs of teachers, followers and non-convergers, i.e.,* $\pi^T = \pi^F = \pi^{NC}$.

To examine this hypothesis consider Table 5, which presents the average payoff (in experimental points) that players obtained over the 20 periods of the game that they played.

BHHH method uses takes the outer product of the score function (calculated with numerical derivatives) to approximate the Hessian. In small samples, these two approximations may not give us very reliable estimates of the standard errors, hence our use of LR tests. See Greene (2003) for more details.

As can be seen, teachers generally earn substantially more than those who do not converge. In the case of the dominance solvable game, this difference is highly statistically significant ($t_{DSG} = 7.10$), while for the non-dominance solvable game, the difference is significant at the 5.7% level ($t_{nDSG} = 1.61$). Regarding our second conjecture, followers make slightly more than teachers in both games; however, in no case is this difference statistically significant ($t_{DSG} = 0.26$; $t_{nDSG} = 0.21$). Therefore, it seems that teachers do not suffer greatly, relative to followers. This may be due to the fact that, for those pairs that converged, the follower picked up relatively quickly on the Nash equilibrium and so relatively few periods of teaching were needed.[10]

Table 5: A Comparison of Payoffs: Teachers, Followers & Non-Convergers

		Teachers	Followers	Non-Convergers
DSG Game:	Mean Payoff	1434.04	1451.38	1065.34
	N	21	13	30
nDSG Game:	Mean Payoff	1344.64	1361.10	1206.53
	N	18	14	32

3.3 The OP and RM Treatments

Perhaps the most striking evidence that teaching is essential for convergence can be seen in our RM and OP treatments. If teaching helps convergence then if we change the environment in a way to make teaching more difficult, we would expect less convergence. This was the reason for our RM and OP treatments. In the RM experiment, subjects were rematched with a different opponent each period so the incentives to teach are substantially reduced since the probability of facing the same opponent again is small. In the OP treatment subjects did not know their opponent's payoffs so that they were unable to calculate the Nash equilibrium. Clearly it is very hard to teach here since a subject does not necessarily know what to teach

[10]Note well that the observed patterns are *not* generated by the fact that teachers are more likely to be row (or column) players: in our data, row and column players are, in fact, equally likely to become teachers. In the non-dominance solvable game, half the teachers are row players and the other half are column players. In the dominance solvable game there is one more row teacher.

(*i.e.,* the Nash equilibrium) or how to interpret his opponent's response.

Observe that to the extent that teaching is essential for convergence, these two treatments should lead to significantly lower convergence rates while if teaching and non-myopic play actually gets in the way, then convergence rates could conceivably increase in the RM and OP treatments.[11] Before considering Tables 6 and 7, which show the convergence rates in our experimental treatments, note that our definition of convergence must change in the RM treatment. This is so because, we have defined convergence as a property of a pair and not an individual. However, in the RM treatment we do not have fixed pairs so given the choices of subjects in any period, the outcomes we observe is a function of how subjects were matched in that period. Instead we take the average frequency that the Nash action was chosen by the players in 10 period intervals.

Table 6 offers compelling support for our hypothesis that teaching is an essential ingredient for convergence. With the exception of the dominance solvable game in the last 40 periods of the RM treatment, which we will discuss more below, convergence rates are far lower in the RM and OP experiments than in the AP experiment where teaching is most easy. As we can see, convergence is greatest in that treatment where teaching is relatively easy and lower when it is relatively difficult. For example, for the nDSG game, while the 20-period convergence rate is 50% in the AP treatment, it is only 5.5% in the OP treatment and 7.0% in the RM treatment. A test of proportions indicates that the proportion of pairs that converge are different in the AP as compared to the OP treatment for both games ($Z_{DSG} = 3.17$ and $Z_{nDSG} = 4.15$; in both cases, $p < 0.01$).

Recall that we ran our OP and RM treatments for an additional 40 periods to see if we might find convergence at a later period. In other words, it might take subjects longer to

[11]One might wonder whether differences in initial beliefs could also play a role in the different convergence rates that we observe. The distributions of initial beliefs for both row and column players in the AP treatment and in the RM treatment are not statistically distinguishable. While the distributions in the AP and OP treatments are different, it is due to the fact that initial beliefs in the OP treatment are concentrated around $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. Moreover, given the parameter estimates for the γ/λ-model, we would not expect convergence rates to be different, regardless of any differences in initial beliefs.

Table 6: 20 Period Convergence Rates Across Treatments

	AP	OP	RM‡ (Per 1-10)	RM‡ (Per 11-20)
DSG	51.3%	16.6%	28.5%	36.0%
nDSG	50.0%	5.5%	14.5%	7.0%

‡ Average frequency that the Nash action was chosen in time range.

converge when teaching is made difficult. These convergence rates are presented in Table 7. Note that in the OP treatment, the 60 period convergence rates were greater than the 20 period convergence rates, but were still below the corresponding 20 period convergence rates found in the AP treatment. For the non-dominance solvable game in the RM treatment, increasing the length of play had no effect on the convergence rate. However, in the RM treatment, the DSG game is remarkably different: in the final 10 periods of play, the Nash action was chosen 93% of the time.

Table 7: 60 Period Convergence Rates Across Treatments

	AP†	OP	RM‡ (Per 21-30)	RM‡ (Per 31-40)	RM‡ (Per 41-50)	RM‡ (Per 51-60)
DSG	51.3%	31.6%	52.0%	72.5%	89.5%	93.0%
nDSG	50.0%	23.5%	4.5%	7.0%	3.5%	6.0%

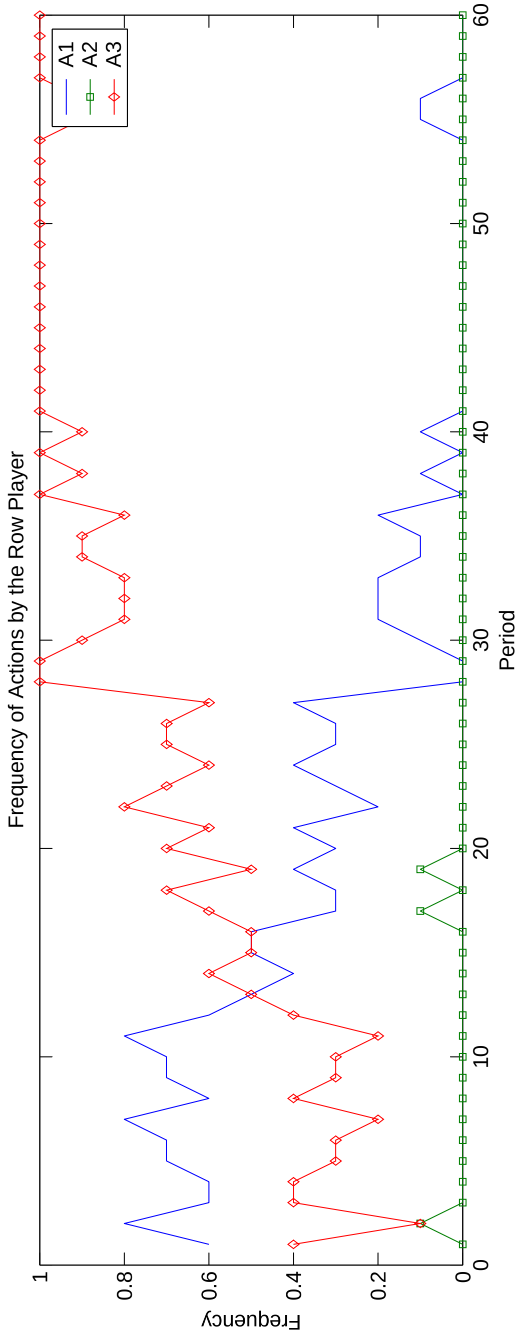
† 20 period convergence rate.

‡ Average frequency that the Nash action was chosen in time range.

While the high convergence rates in the last 40 rounds of the RM treatment appear to contradict our view that convergence is rare when teaching is difficult, we will demonstrate that the high convergence rates found in those periods of the RM treatment were due to subjects learning over time how to iteratively delete dominated strategies.[12] To illustrate the iterative dominance principle at work, consider Figures 7 and 8. These figures present the frequency of actions taken by the row and column players period by period in the DSG game as well as the mean beliefs of subjects. The belief data is presented in terms of which

[12]While we believe, given the low incentives to do so, that teaching is not the primary reason for the high 60 period convergence rate, that is not to say that teaching does not play any role. If you examine the correlation between the period of convergence and whether a subject's beliefs converged before or after his/her actions, you see that subjects who converged relatively earlier were more likely to have actions converge before beliefs. This could be considered weak evidence in favour of teaching.

Figure 7: History of Play (Actions): Random Matching, DSG Game



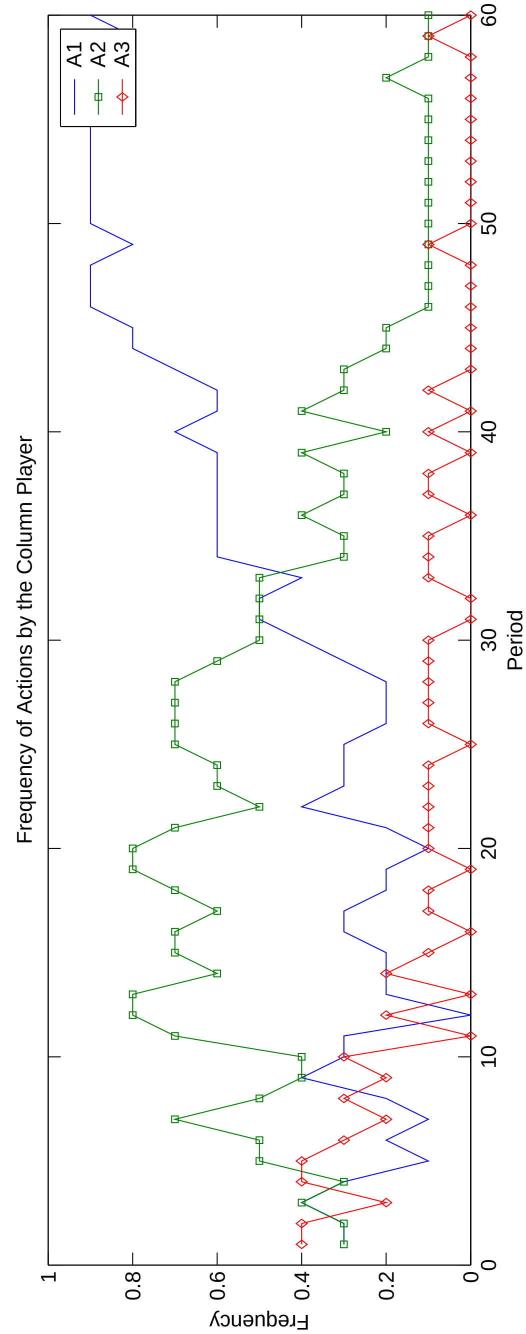
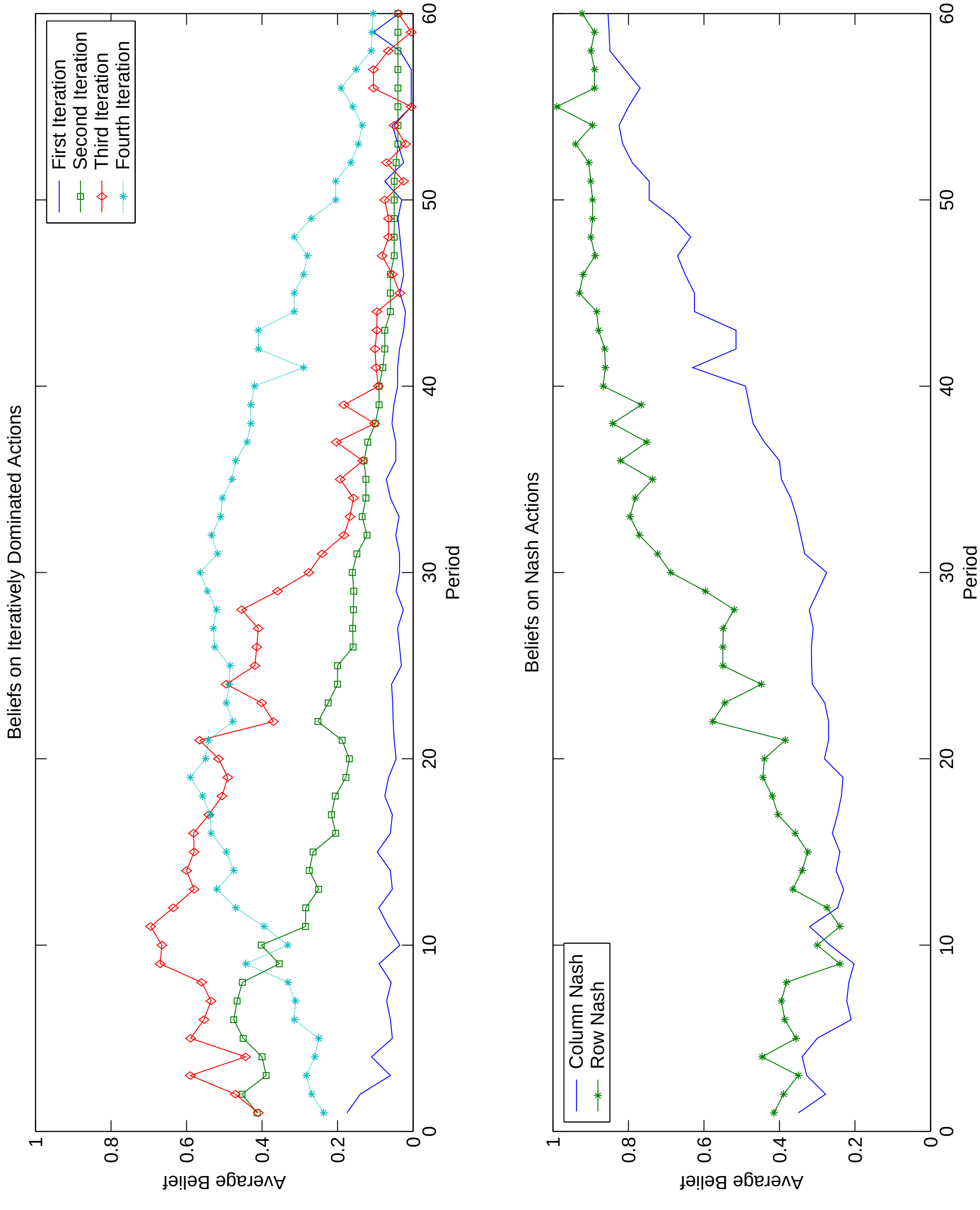


Figure 8: History of Play (Beliefs): Random Matching, DSG Game



strategy is supposed to be eliminated first, second, third, *etc.* In this game action A_2 for row is the first strategy to be iteratively eliminated, followed by A_3 for column and then A_1 for row and finally A_2 for column.

Looking at Figure 7 we see that, in fact, this sequence is actually followed over time. For example, as just mentioned A_2 for row is supposed to be the first strategy eliminated. As we see in Figure 7 row players virtually never play that strategy either at the beginning or later on in experiment. Correspondingly, if we look at the mean beliefs of column players about row's use of A_2 (the thin solid line in Figure 8 indicating the belief on the first strategy to be eliminated) we see that the column player does not expect to see A_2 played either at the beginning or during the course of the game. So strategy A_2 for row is eliminated from the very beginning and that meets the expectations of the column player.

The next strategy that should be eliminated is A_3 for column. As we see from Figure 7, by period 13 the mean use of strategy A_3 for column drops below all the other strategies and stays there throughout the rest of the game. Clearly it is the second strategy eliminated. In Figure 8 we see that the beliefs of row concerning the use of A_3 by column are in fact second lowest and eventually fall to close to zero. If we continued this process we would see that the next strategy whose use falls to zero is A_1 for row and finally A_2 for column with the corresponding beliefs falling in a complementary fashion. Finally consider Figure 8 which shows the beliefs that row and column place on the Nash actions of their opponents over time and notice that these beliefs are converging toward 1. It appears, therefore, that while play did converge in the RM treatment toward the Nash equilibrium, it did so following a process of iterative elimination of dominated actions rather than a teaching process.

Finally, one could ask whether the play of the nDSG game with random matching has converged to one of the two mixed strategy equilibria for that game. To answer this question, we computed the average frequency that each action is played over the 60 periods of the game. The results of this exercise are presented graphically in Figure 9. In the figure, the

$\bullet$'s indicate each player's 60 period action choice frequencies, the $\blacksquare$ denotes the average over all players. The two $\blacklozenge$'s indicate the two mixed strategy Nash equilibria. Finally the $-.$ line depicts the Quantal Response Equilibrium choice probabilities as a function of λ and the $+$ shows a specific QRE at $\lambda \approx 0.8$, which is the estimate of the γ/λ-model for this data.

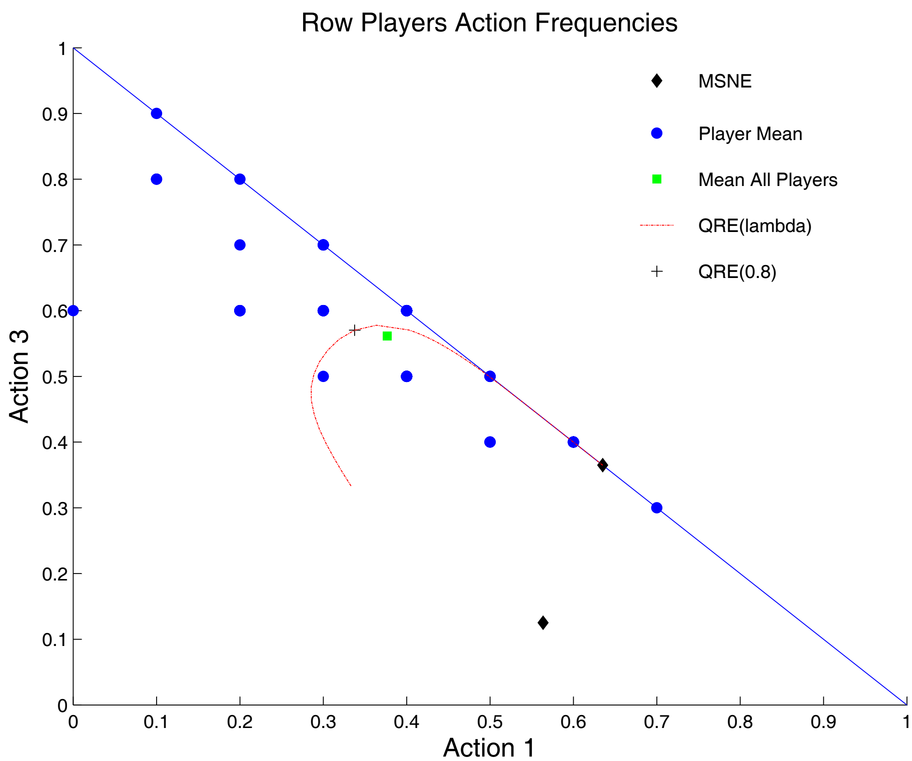
As the reader can see, the players are very far from both of the mixed strategy Nash equilibria. Interestingly, however, the average frequency of play for all players is very close to the estimated quantal response equilibrium for both row and column players.

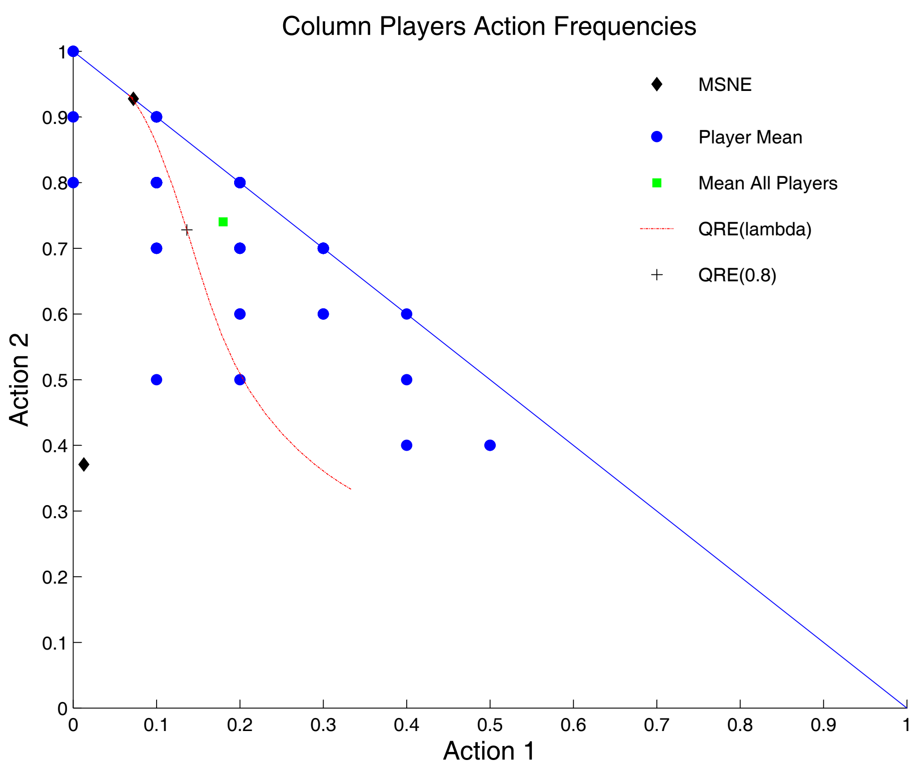
3.4 A Belief Formation Model

As we have indicted before, in most belief learning models beliefs are formed using the historical frequencies of past actions of one's opponent. In other words, as in (1) — the Cheung-Friedman belief model — beliefs are historical and take into account only the actions of one's opponent and not his payoffs. This backward looking focus is true of other non-belief learning models such as reinforcement learning (Erev and Roth, 1998) and the EWA model of Camerer and Ho (1999), as well as evolutionary models (Samuelson (1997) and Weibull (1995)). None of them take into account an opponent's payoffs in determining the likelihood that he or she will use any particular strategy. We now turn to this question and specifically ask whether one's payoffs and the payoffs of one's opponent matter in the belief formation process.

The reasons why we believe taking into account an opponent's payoff is important are twofold. First and foremost, when the game is common knowledge, there is hope for successful teaching, since by playing her part of the Nash equilibrium, the teacher can signal to the follower that he has an interest in also playing the Nash equilibrium. Beyond that, however, consider the following: Say that a row player and a column player have just finished a round of play in a repeated game and assume that the row player has done very well (received a high payoff) but the column player has done terribly (received a very low payoff). Clearly,

Figure 9: Average Frequency of Action Choices





you might think that the column player would want to avoid this mistake in the next period and hence the row player might believe that the column player will alter his or her play. The question is, how does the row player think the column player will respond?

Suppose that the row player believes that the column player is a 'Level 1' player. That is, she believes that column will best respond to the action she chose last period (as in a Cournot model). Under this assumption, the row player clearly needs to know the payoffs to the column player in order to formulate her beliefs; that is, she needs to know what the best response is. Alternatively, the row player may believe that the column player is a 'Level 2' player. That is, she believes that the column player will play a best response to the row player's best response to *his* last period's chosen action. Finally, she may believe that the column player is a 'Level 0' player, in which case she would assume that her opponent will choose each strategy with equal likelihood. Clearly, in these situations (except for the Level 0 assumption), both players' payoffs must be known in order to formulate such a belief.[13]

We incorporate these concerns into a model of belief formation by positing the following model:

$$B_i^k(t+1) = \alpha\Gamma_i^k(t+1) + (1-\alpha)\Lambda_i^k(t+1). \tag{4}$$

In this model the belief of agent i about agent j's use of strategy k in period $t+1$ is a weighted average of two things. The first, $\Gamma_i^k(t+1)$, is the player's Cheung-Friedman beliefs which, as we know, is a belief measure based solely on the historical play of one's opponent. α is the weight placed on this term. Clearly, if $\alpha = 1$, our player is merely a person who uses history to update beliefs as in the Cheung-Friedman model. The second component is there to capture the more forward looking or sophisticated aspect of agent i's belief formation process. More precisely, $\Lambda_i^k(t+1)$is defined as:

$$\Lambda_i^k(t+1) = \Pr[L_0][\frac{1}{3}] + \Pr[L_1][l_t(BR(L_1(t)) = k)] + \Pr[L_2][l_t(BR(L_2(t)) = k)] \tag{5}$$

[13]We restrict attention to two levels of sophistication since it is the simplest model in which the payoffs of both players must be known in order to accurately form beliefs.

Here the probability that agent i thinks that his opponent will use strategy k depends on what level of sophistication she assumes for her opponent. $l_t(BR(L_m(t)) = k)$ is an indicator function which takes a value of 1 if strategy k is the best response of a level m ($m = 1, 2$) opponent to her action in period t, and zero otherwise. If her opponent is a level 0 player, which he is assumed to be with probability $\Pr[L_0]$, he will simply randomize and choose each strategy with probability $\frac{1}{3}$. If, her opponent is a level 1 player, which he is assumed to be with probability $\Pr[L_1]$, and if k is the best response of a level 1 player to her opponent's last period's action, he will take action k for sure. Finally, a player's opponent could be choosing strategy k because he is a level 2 player and k is the level 2 best response to last period's outcome. The probability that k is chosen for any of these these reasons is the expected value of these considerations, which is $\Lambda_i^k(t+1)$.

The question now arises as to how we calculate $\Pr[L_m]$. For each player, we calculated the fraction of times that he/she took an action that was consistent with Level 1 play only, that was consistent with level 2 play only, that was consistent with both level 1 and level 2 play and the fraction of times that he/she took an action that was consistent with neither level 1 nor level 2 play. This latter fraction we took to be $\Pr[L_0]$. We then take $\Pr[L_1]$ ($\Pr[L_2]$) as the fraction of times the player took an action consistent with *only* level 1 (level 2) play, plus half the fraction of times the action was consistent with both level 1 and level 2 play. Hence, these probabilities do not change as the game progresses.[14] An alternative, and perhaps more satisfactory, method of calculating our probabilities would have been to estimate them; however, this would introduce two more parameters to estimate, which would make estimation rather difficult since we are interested in estimating the model with individual level data.

This model is then a two parameter model of belief formation characterized by α and γ.

[14]Obviously, these fractions are not known until the game is over, yet our subjects form their beliefs period by period. Hence, we must consider this an approximation. However, it should be noted that having subjects form beliefs based on the empirical distribution of "types" shares some similarities to the *sophisticated* types of Stahl and Wilson (1995), Costa-Gomes *et al* (2001) and Costa-Comes and Crawford (2006), among others, who best respond to the empirical distribution of types.

Given the stated beliefs of the subjects reported in the experiment, we can estimate these parameters by minimizing the sum of squared prediction errors. That is, given a sequence of choices, $\{A_{i,t}\}_{t=1}^{20}$ and stated beliefs $\{b_{i,t}\}_{t=1}^{20}$ for each player i, our criterion function is:

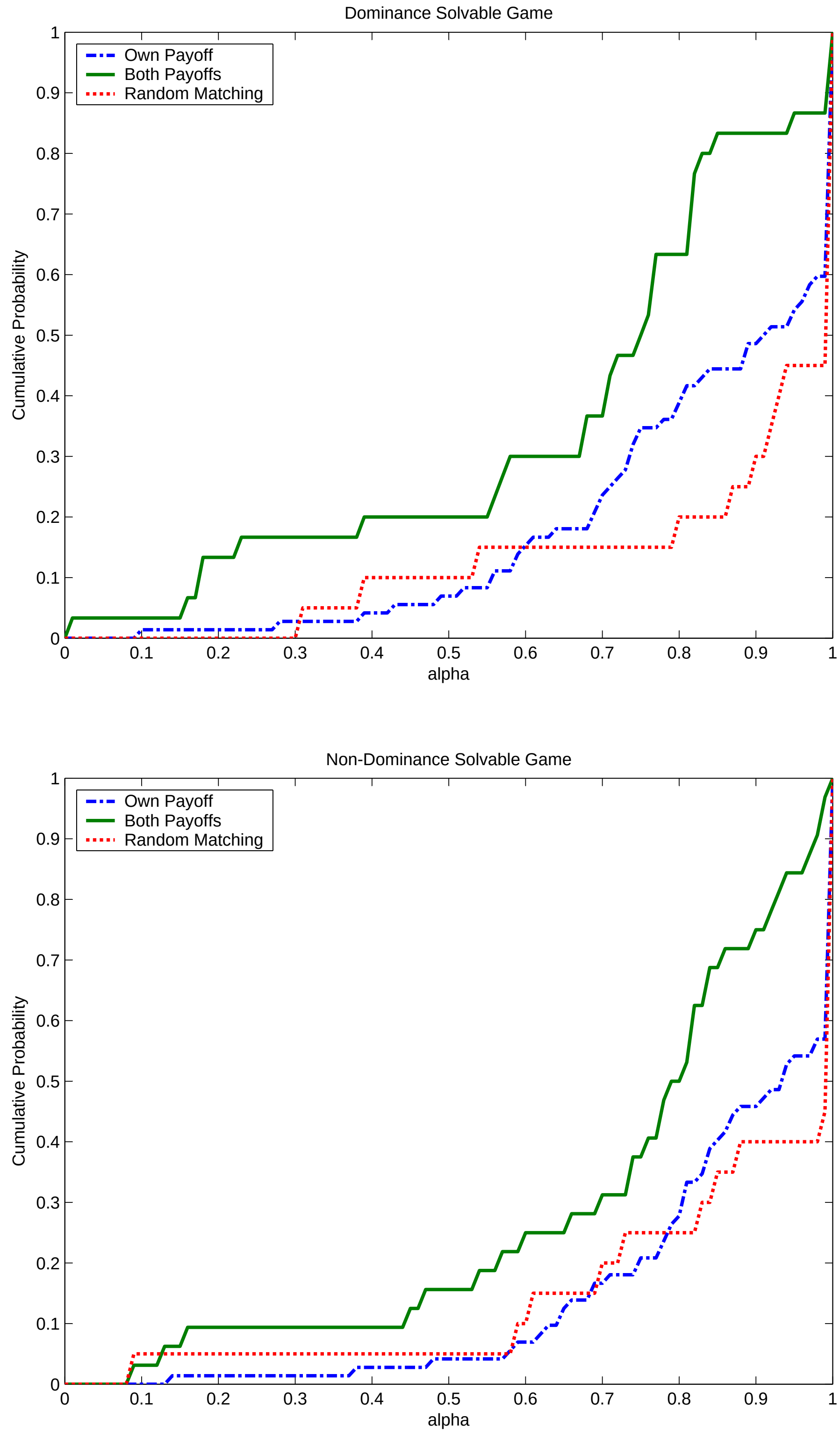
$$SSE(\alpha, \gamma) = \sum_{t=1}^{20} \sum_{k=1}^{3} (B_i^k(t) - b_{i,t}^k)^2 \tag{6}$$

where $B_i^k(t)$ is the prediction of our belief model about the use of strategy k by subject i in period t and $b_{i,t}^k$ is the stated belief of that subject. We sum over all periods $t = 1, \ldots, 20$ and all possible actions $k = 1, 2, 3$. We then search for the $(\alpha, \gamma) \in [0, 1]^2$ pair which minimizes (6).

If we estimate this belief formation process on individual data and find $\alpha < 1$, then this provides evidence that subjects are sophisticated in their belief updating process in the sense that they use more than mere historical data to form beliefs. They also use the payoffs of their opponent last period. Performing this exercise subject by subject on the data from our AP treatment, we plot our empirical distribution of $\hat{\alpha}$ in Figure 10. In this figure, the solid lines plot the empirical distributions of the estimated $\hat{\alpha}$ for the DSG and nDSG games. It is apparent that, in both games, almost all subjects have $\hat{\alpha} < 1$ and many subjects have $\hat{\alpha}$ significantly below 1. Moreover, compared to a straight-forward Cheung-Friedman belief model, where α is restricted to 1, the fit is significantly improved: on average the sum of squared errors is reduced by 14% in the DSG game and by 10% in the nDSG game.

As another check of sophistication, we can actually conduct the same exercise using data from the treatments RM and OP to obtain estimates of α. In the RM treatment, since opponents are constantly being rotated and the incentives for sophisticated play are diminished, we expect to find higher values of α. In the OP treatment, since subjects only see their own payoff, they cannot actually be forming beliefs in this manner. Therefore, we should expect that $\alpha = 1$ and any $\alpha < 1$ is simply spurious. However, we still believe that estimating this model on data from our OP treatment can serve as a useful diagnostic tool

Figure 10: Empirical Distributions of $\hat{\alpha}$



for comparing behavior with the AP treatment, which is our primary goal. That is, to the extent that the distribution of α for the AP treatment is stochastically smaller than the distribution for the OP treatment, we can say that subjects are more sophisticated in the AP treatment. Our conjectures can be stated more succinctly as:

Sophistication Hypothesis: *For both the RM and OP treatments and in both games, the empirical distributions of $\hat{\alpha}$ should first-order stochastically dominate the empirical distribution for the AP treatment. Moreover, for the OP treatment, we should see $\hat{\alpha} = 1$ a significant fraction of the time.*

One can see from Figure 10, which also plots the empirical distributions for treatments OP and RM, that our sophistication hypothesis is verified: there is a clear First-Order Stochastic Dominance relationship between AP and OP, *and* between AP and RM, with the $\hat{\alpha}$'s in the OP and RM treatments substantially higher than in the AP treatment. Kolmogorov-Smirnov tests also support the apparent difference visible from the figure: for both games, we easily reject the null hypotheses that the distribution $F_{AP}(\alpha)$ is identical to *either* $F_{OP}(\alpha)$ or $F_{RM}(\alpha)$ (for the dominance solvable game, we have: $D_{AP\,vs.\,OP} = 0.39$ and $D_{AP\,vs.\,RM} = 0.63$, while for the non-dominance solvable game, we have: $D_{AP\,vs.\,OP} = 0.40$ and $D_{AP\,vs.\,RM} = 0.57$; in all cases $p < 0.01$). Moreover, in the OP treatment, we see that $\hat{\alpha} = 1$ for over 40% of our subjects; in the RM treatment, the same is true for nearly 60% of our subjects.

Thus, the belief formation process varies substantially between treatments; in particular, when information regarding payoffs of one's opponent is available and relevant (as in treatment AP), they take advantage of it. This suggests that, in order to fit behavior, one must use learning models which allow for the payoffs of one's opponent to influence both their action choices and their belief formation process.

4 Conclusions

This paper has attempted an investigation of the process through which people playing games converge to an equilibrium — a state where their beliefs about the actions of their opponents are confirmed. What we have uncovered is that in the class of games we have looked at here, teaching is an essential element for convergence. More precisely, in the two person 3×3 games we used in our experiments, those pairs that were successful in converging did so through a process quite different from the backward looking myopic and mechanical (BLM&M) process described in the learning literature. Rather, convergence seems to be an action led process where one player, the teacher, takes it upon herself to lead the way to the Nash equilibrium by repeatedly choosing her Nash action despite the fact that it is not a best response to her beliefs. Successful convergence matches such a teacher with a fast learner, *i.e.,* someone who places sufficient weight on recent history. Non-convergence appears to be the result of a faulty, too sluggish, belief updating process rather than an inability to best respond. Hence the telltale sign of convergence is the fact that actions (at least of the teacher) reach an equilibrium state before beliefs.

To test these findings we ran a set of experiments where teaching is made difficult, *i.e.,* where players know only their own payoffs (OP treatment) or are randomly rematched each period (RM treatment). What we found is that when teaching becomes difficult convergence becomes rare. This is interesting because these treatments are expected to have either no impact (in the OP treatment) on the convergence of BLM&M processes or actually enhance it (in the RM treatment).

In the last section of our paper we investigate the belief formation process used by subjects by positing our own belief formation model. What we try to capture here is that in forming one's beliefs about one's opponent at some time in their interaction, a player should place some weight on the payoffs that their opponent has received and not just the history of their actions. This is in contrast to most belief and attraction models which look only at the previous history of actions in forming beliefs. We find that subjects do in fact take these payoffs into account by assessing the level of sophistication of their opponent, which the model requires them to do. In the OP and RM treatments, where this type of belief formation model is hard to use, we find that subjects in fact revert to relying more on history and this is partially responsible for the lack of convergence in these experiments.

There is still a lot of work to be done. If what we have uncovered here is replicated and makes sense investigators may want to span a wide variety of classes of games and environments to see if teaching is relevant in all, some or none of them. For example, in multi-person games the ability of a player to teach is diluted by the actions of others. Still environments with different feedback or communication rules may foster teaching and hence convergence. Monotone games with information about the choices of others should also foster convergence (see Chen and Gazzale (2004)) and, in fact, there is evidence (see Chen and Plott (1996)) that truthful revelation mechanisms defining monotone games have sharp convergence. Is this because of teaching?

References

[1] Boylan, R. T., and M. A. El-Gamal (1993), "Fictitious Play: A Statistical Study of Multiple Economic Experiments," *Games and Economic Behavior*, 5(2), 205–222.

[2] Camerer, C. and Ho, T-H. "Experience-Weighted Attraction Learning in Normal Form Games" vol. 67, issue 4, *Econometrica*, 1999, pp 827-874.

[3] Camerer, C., T. Ho, and J. Chong (2002), "Sophisticated EWA Learning and Strategic Teaching in Repeated Games," *Journal of Economic Theory*, 104, 137-188.

[4] Chen, Y. and Plott, C. (1996) "The Groves-Ledyard Mechanism: An Experimental Study of Institutional Design," *Journal of Public Economics* vol. 59, Issue 3, 335-364.

[5] Chen, Y. and Gazzale, R.(2004) "When Does Learning in Games Generate Convergence to Nash Equilibrium? The Role of Supermodularity in an Experimental Setting," *American Economic Review,* vol. 94, No. 5: 1505-1535.

[6] Cheung, Y.-W., and D. Friedman (1997), "Individual Learning in Normal Form Games," *Games and Economic Behavior*, 19, 46-76.

[7] Costa-Gomes, Miguel A. and Weizsäcker, Georg, "Stated Beliefs and Play in Normal-Form Games" (May 22, 2004)

[8] Costa-Gomes, Miguel A. and Vincent Crawford, "Cognition and Behavior in Two-Person Guessing Games: An Experimental Study," The American Economic Review, *forthcoming.*

[9] Costa-Gomes, Miguel, Vincent Crawford, and Bruno Broseta (2001), "Cognition and Behavior in Normal-Form Games: An Experimental Study," *Econometrica*, 69, 1193-1235.

[10] Crawford, V. "Adaptive Dynamics in Coordination Games," Econometrica, 63 (1995), 103-143.

[11] Dufwenberg, Martin, and Uri Gneezy (2000), "Measuring Beliefs in an Experimental Lost-Wallet Game," *Games and Economic Behavior*, 30, 163-182.

[12] Erev, I. and A. Roth, "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed-strategy Equilibria," *The American Economic Review*, 88 (1998), 848-881.

[13] Feltovich, Nick (2000), "Reinforcement-based vs. Belief-based Learning Models in Experimental Asymmetric-information Games," *Econometrica* 68 (3), 605-641.

[14] Fischbacher, Urs (1999) "z-Tree - Zurich Toolbox for Readymade Economic Experiments — Experimenter's Manual," Working Paper Nr. 21, Institute for Empirical Research in Economics, University of Zurich.

[15] Fudenberg, D. and D. Kreps (1993), "Learning Mixed Equilibria," *Games and Economic Behavior*, 5, 320-67.

[16] Fudenberg, D. and D. K. Levine (1998), "The Theory of Learning in Games," Cambridge, MA: MIT Press.

[17] Greene, William H., (2003) *Econometric Analysis,* 5$^{\text{th}}$ edition. Prentice-Hall: Upper Saddle River, New Jersey.

[18] Hart, Sergiu, and Andreu Mas-Colell (2005), "Stochastic uncoupled dynamics and Nash equilibrium," *Games and Economic Behavior*, 57(2), 286–303.

[19] Haruvy, Ernan (2002) "Identification and Testing of Modes in Beliefs," *Journal of Mathematical Psychology*, 46, 88–109.

[20] Hopkins, Ed (2002), "Two Competing Models of How People Learn in Games." *Econometrica* 70 (6), 2141-2166.

[21] Huck, Steffen, and Georg Weizsäcker (2002), "Do Players Correctly Estimate What Others Do? Evidence of Conservatism in Beliefs," *Journal of Economic Behavior and Organization*, 47, 71–85.

[22] Mason, Charles, and Owen Phillips (2001), "Dynamic Learning in a Two-Person Experimental Game," *Journal of Economic Dynamics and Control*, 25, 1305–1344.

[23] McKelvey, Richard and Thomas Palfrey (1995), "Quantal Response Equilibrium for Normal Form Games," *Games and Economic Behavior*, 10, 6–38.

[24] Milgrom, P. and J. Roberts (1991), "Adaptive and Sophisticated Learning in Normal Form Games," *Games and Economic Behavior,* 3, 82–100.

[25] Mookherjee, D., and B. Sopher (1994), "Learning Behavior in an Experimental Matching Pennies Game," *Games and Economic Behavior*, 7(1), 62–91.

[26] Nyarko, Yaw and Andrew Schotter (2002a), "An Experimental Study of Belief Learning Using Real Beliefs," *Econometrica*, 70, 971–1005.

[27] Nyarko, Yaw and Andrew Schotter (2002b), "Comparing Learning Models Using Ideal Micro-data Sets," mimeo.

[28] Offerman, Theo and Joep Sonnemans (2001), "Is the Quadratic Scoring Rule Behaviorally Incentive Compatible?" Working paper.

[29] Offerman, Theo, Joep Sonnemans and Arthur Schram (1996), "Value Orientations, Expectations and Voluntary Contributions in Public Goods," *Economic Journal* 106, 817–45.

[30] Partow, Z., and Andrew Schotter (1993), "Does Game Theory Predict Well for the Wrong Reasons: An Experimental Investigation," Working Paper 93-46, New York University.

[31] Rutström, E. Elisabet and Nathaniel T. Wilcox (2004), "Learning and Belief Elicitation: Observer Effects," Working paper.

[32] Salmon, Timothy C. (2001), "An Evaluation of Econometric Models of Adaptive Learning," *Econometrica*, 69 (6), 1597–1628.

[33] Samuelson, L. (1997), *Evolutionary Games and Equilibrium Selection,* The M.I.T. Press, Cambridge, MA.

[34] Selten, R., and J. Buchta (1999), "Experimental Sealed-Bid First Price Auctions with Directly Observed Bid Functions." In D. Budescu, I. Erev, and R. Zwick (eds.): *Games and Human Behavior: Essays in the Honor of Amnon Rapoport*. Lawrenz Associates Mahwah NJ.

[35] Selten, R., and R. Stoecker (1986), "End Behavior in Sequences of Finite Prisoners' Dilemma Supergames: A Learning Theory Approach." *Journal of Economic Behavior and Organization*, 7, 47-70.

[36] Stahl, Dale, and Paul Wilson (1995), "On Players' Models of Other Players: Theory and Experimental Evidence," *Games and Economic Behavior*, 10, 218-254.

[37] Terracol, Antoine and Jonathan Vaksmann (2006), "Dumbing Down Rational Players: Adaptive Learning and Teaching in an Experimental Game," Working Paper.

[38] Von Neumann, J. and O. Morgenstern (1944), *Theory of Games and Economic Behavior.* Princeton, N.J., Princeton University Press.

[39] Weibull, J. (1995), *Evolutionary Game Theory.* Cambridge: MIT Press.

[40] Weizsäcker, Georg (2003), "Ignoring the Rationality of Others: Evidence from Experimental Normal Form Games," *Games and Economic Behavior*, 44, 145-171.

[41] Wilcox, Nathaniel, and Nick Feltovich (2000), "Thinking like a Game Theorist: Comment," mimeo, University of Houston.

[42] Young, H. Peyton (2006), "The Possible and the Impossible in Multi-Agent Learning," Department of Economics, University of Oxford, May 2006.

[43] Young, H. Peyton (2004), *Strategic Learning and its Limits*, Oxford University Press: Oxford, UK.

A Instructions Used In The AP Treatment

The following instructions were used for the AP Treatment. The instructions were only minimally changed for the other treatments.

General Instructions

Welcome and thank you for coming today to participate in this experiment. The purpose of this experiment is to learn how people behave in certain very simple settings.

After this experiment, another experiment will take place. The precise details of that experiment will be explained to you at the appropriate time. At the beginning of this experiment, you will be randomly paired with another participant and you will remain paired with this person for the duration of this **first** experiment. However, at no point in time will it ever be revealed to you with whom you are matched. Depending on your choices and those of your partner you will earn money, which will be paid at the end of the experiment. The exact method of calculating your final payment will be described below.

We ask that you remain silent throughout the experiment. If, at any time, you have a question, please ask the session coordinator. Failure to comply with these instructions means that you will be asked to leave the experiment and all earnings will be forfeited.

In the experiment it is more convenient to work with points rather than dollars. At the end of the experiment, the total number of points earned will be converted to dollars. The exact conversion factor is the following:

200 points = $1.00

Decision Problem

In this experiment, you and your partner will play a game for a total of 20 periods. You will be asked two things. In each period, you will be asked to choose an action. Your action choice and the action choice of your partner will determine the bulk of your payoff for each period. However, before making your action choice, you will be asked to *predict* the action choice that your partner will make. We now explain, in detail, both of the decisions you must make.

The Game

On your computer screen, you will be presented with the following representation of the game that you and your partner will play. The game is exactly the same for all 20 periods.

	$A1$	$A2$	$A3$
$A1$	$12, 83$	$39, 56$	$42, 45$
$A2$	$24, 12$	$12, 42$	$58, 76$
$A3$	$89, 47$	$33, 94$	$44, 59$

In each period, both you and your partner will simultaneously choose one of three actions, labeled $A1$, $A2$ and $A3$. The actions that you and your partner take in each period, as well as your position as either the **row** or **column** player, determine the payoffs for that period. Each of the nine boxes above represent the nine possible action combinations. In each box, the **first** entry represents the payoff for the **row** player, while the **second** entry represents the payoff for the **column** player.

For example, suppose that you are the *row* player and chose action $A3$ in the current period; suppose also that your partner, the *column* player, chose action $A1$. Then you will earn 89 experimental points for the current period and your partner will earn 47 experimental points. As another example, suppose that you are the column player and took action $A2$, while your partner, the row player, took action $A1$. In this case, you, the column player, will earn 56 points and your partner, the row player, will earn 39 points.

*On the computer screen it will be clearly marked whether you are the **row** player or the **column** player. Moreover, your position will not change in **any** of the 20 periods which comprise this first experiment. That is, if you are the row player in period 1, you will continue to be the row player in each of periods 2 through 20.*

Predicting Other People's Choices

Prior to choosing an action in each period, you will be given the opportunity to earn additional money by predicting the choices of your partner in the game. On your computer screen, each of you will be asked the following three questions:

• On a scale from 0 to 100, how likely do you think it is that your partner will take action $A1$?

• On a scale from 0 to 100, how likely do you think it is that your partner will take action $A2$?

• On a scale from 0 to 100, how likely do you think it is that your partner will take action $A3$?

Your response to each question must be a number between 0 and 100. Moreover, the sum of the three numbers that you provide ***must*** *be* ***exactly*** *100.*

For example, suppose that you think there is a 30% chance that your partner will take action $A1$, a 25% chance that your partner will take action $A2$ and a 45% chance that your partner will take action $A3$. In this case, you will enter 30 in the first box on the screen in the bottom left corner and 25 and 45 in the second and third, respectively. The exact computer screen you will see is given below.

You will earn experimental points for your predictions according to a specific payoff function, which we now explain. Suppose your predictions are as in the above example. Furthermore, suppose that in the current period your partner actually chose $A2$. In that case your payoff for predicting your partner's action will be:

$$\text{Payoff} = 5[2 - (\frac{30}{100})^2 - (1 - \frac{25}{100})^2 - (\frac{45}{100})^2]$$

In other words, we will give you a fixed amount of 10 points from which we will subtract an amount which depends on how inaccurate your prediction was. To do this, we find out what choice your pair member has made. We then take the number you assigned to that choice, in this case 25% on $A2$, subtract it from 100%, square it and multiply by 5. Next, we take the number you assigned to the choices not made by your pair member, in this case the 30% you assigned to $A1$ and the 45% you assigned to $A3$, square them and multiply by 5. These three squared numbers will then be subtracted from the 10 points we initially gave you to determine your final point payoff. Your point payoff will then be converted into dollars at the same conversion factor as given above.

Note that since your prediction is made before you know what your partner has actually chosen, the best thing you can do to maximize the expected size of your prediction payoff is to simply state your true beliefs about what you think you partner will do. Any other prediction will decrease the amount you can expect to earn as a prediction payoff.

Note also that you cannot lose points from making predictions, you can only earn more points. The worst thing that could happen is that you predict that your partner will choose one particular action (*e.g.,* $A2$) with 100% certainty *but* it turns out that your partner actually chose a different action (*e.g.,* $A3$). In this case, you will earn 0 points. In all other situations, you will earn a strictly

positive number of points.

The Computer Screen

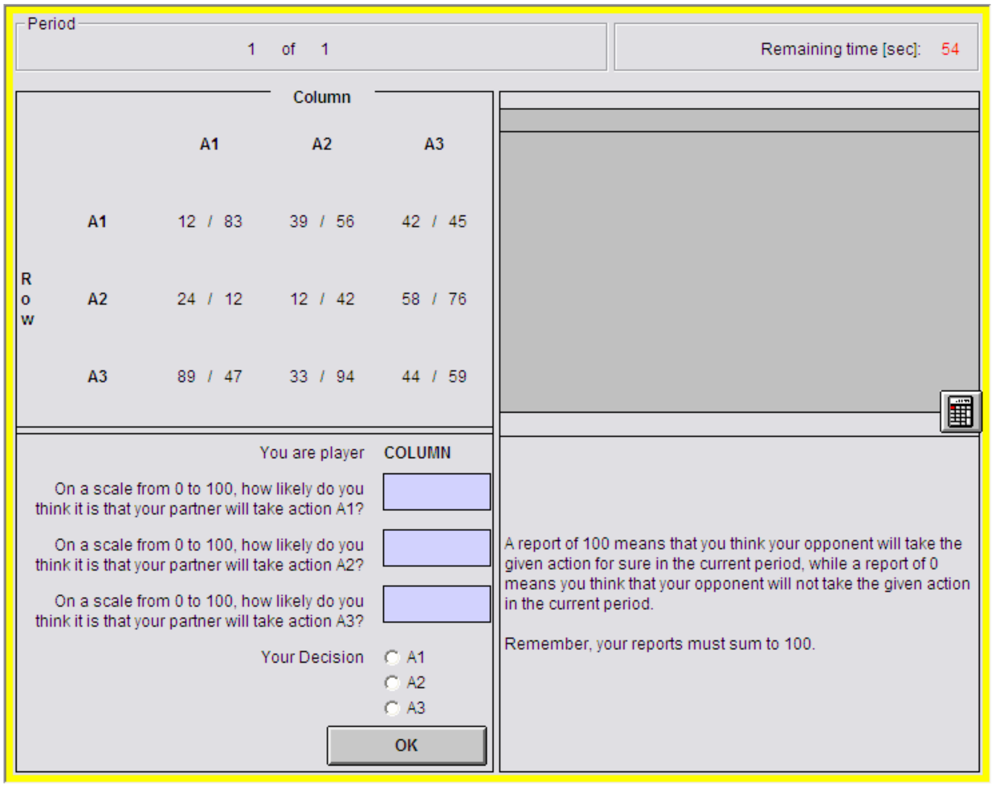
In each period, you will see the following computer screen. At the very top of the screen, you will see which period you are in, how many periods in total there are and the time remaining to make your decision. In the experiment, you will have 1 minute to state your predictions regarding what you think the other player will do and choose an action.

In the top left portion of the screen, you will see the game that you are playing and the payoffs to both you and your partner for each combination of actions taken. Recall that the first number in every box gives the payoff to the **row** player and the second number gives the payoff to the **column** player for each of the nine possible action combinations.

The top right portion of the screen shows the outcomes from all previous periods. In particular, you are able to see what action you took in all previous periods and what action your partner took in all previous periods.

The bottom left corner of the screen is where you will make your decisions. The first thing that you will see is whether you are the row player or the column player. Below that are the three questions asking you to predict the action choice of your partner. **Your responses to these three questions must each be numbers between 0 and 100 and the three numbers must sum to 100. Your response may contain at most 1 number after the decimal point.** The bottom right portion of the screen has a couple of reminders that you may wish to refer to during the experiment. You will also see a calculator button. By pressing this button, the computer's calculator appears, which can be used as a check that your predictions add up to 100.

After all of your decisions have been made, click on the **OK** button. Once both you and your partner have pressed OK, you will be taken to a new screen where you may review the action that you took, learn the action taken by your partner and find out your payoff from your action choices for that period. In the bottom right corner of this screen, you may press **continue**. Once both you and your partner have done so, you will be returned to the main screen, where a new period, exactly the same as the previous, will begin. There are 20 periods in total.



Final Payment

Your final payment for the experiment will be determined as follows. We will sum up the number of experimental points earned in each period for your action choices as well as for your predictions regarding your partner's behavior. The total number of points will then be converted back into dollars at the rate of \$1 = 200 experimental points. This will be combined with your \$7 participation fee to come up with your final payment. Payments will be made privately at the conclusion of the two experiments.