A look at learning models in Three by Three Bimatrix games. *

Nicole Marie Bouchez
Royal Holloway
University of London
Egham Surrey TW20 0EX
N.Bouchez@rhul.ac.uk

20 November 2000

Abstract

Experimental data is used to test a variety of learning models using a model that extends several of the existing learning models. Generally, the parameter estimate are in the expected ranges. The individual agent parameter estimates indicate that there is considerable individual heterogeneity. Representative agent parameter estimates adequately predict the mode of the individual parameter estimates when the data is pooled across matrices. They are not very effective at predicting the mode of the disaggregated data. There is some evidence in favour of the restriction that the two discounts are equal. The restrictions that the agents equally weigh actions experienced and actions not experienced is rejected using both representative agent and individual agent parameter estimates)thought there is evidence that subjects put more weight on actions experienced over those not experienced). We reject the rote learning version of the model in favour of a (weak) belief learning parameterizations of the model.

*This research was supported in part by NSF Doctoral Dissertation Grant SRB-9617648

1 Introduction

This paper is about learning. We have all experienced learning but learning actually encompasses some very different things: we learn how to walk as children, we sit in classrooms and learn Math, and English, we learn how to drive and we learn how to invest money in the stock market. We learn when we acquire information, and we learn when we are confronted with a situation where we don't know what others are going to do.

We learn the rules of the road, but how we actually drive also depends on *who* is on the road. Every time we get into the car, we interact with the others on the road. A trip to the grocery store at midnight differs from a trip at rush hour; how you drive on the freeway depends on who else is on the road with you, if there is a patrol car ahead you may not choose to speed, if the traffic is going at 75mph you may decide to go with the traffic flow. If you don't know ahead to time who is going to be on the road with you how do you decide how to drive?

This type of learning also happens in sports, and in any situation where the outcome depends on the actions of others. If you don't know what they are going to do ahead of time, what is it that determines your actions?

The models I will be considering have roots in the behavioral school of psychology that was dominant in the first half of the century. Much of the psychology research on learning ended in the 1950's as the behavioral school fell into disfavour (and was replaced by cognitive approaches). Recently, there has been renewed interest in learning in psychology including work by Friedman, Massaro, Kitzis & Cohen (1995) and Kitzis, Kelley, Berg, Massaro & Friedman (1998).

Out of the pre-1950 tradition came the rote learning model (also known as reinforcement learning, stimulus response or the law of effect). In these models, successful strategies are reinforced and are more likely to be used again. First formalized by Bush & Mosteller (1955); recent work on reinforcement models has come from Harley (1981), Cross (1983), Börgers & Sarin (1996), Roth & Erev (1995),Erev & Roth (1998), and Tang (1996).

Belief models are another type of learning mechanism. Here, players form beliefs about the state of the world next period and optimally respond to these beliefs and individuals are allowed to take into account things they have not personally experienced. Examples of the simplest of this type of learning include Cournot best response (Cournot 1863) and fictitious play (Brown 1951). As mentioned previously, in Cournot learning, what happened last period is assumed to be the best predictor of what will happen next period. In fictitious play, the opponent's average play over time is used as the best predictor of their action next period. Other models of this type include Cheung & Friedman (1997) and Fudenberg & Levine (1998) which both allow for the weighting of past periods. Still more sophisticated are the "sophisticated belief" models that provide complex models of opponent behaviour (Selten 1991, Stahl 1993, Stahl 1996, Stahl 1999*a*, Stahl 1999*b*). Most recently, Camerer & Ho (1998*c*) show that the previously competing simple belief and rote learning models can be nested in their Experience Weighted Average (EWA) model, a hybrid model with belief and rote learning as special cases.

Empirical work so far has concentrated on comparing different models in terms of their fit and predictive ability with experimental data. Camerer & Ho (1998*b*) and (1998*a*) find that their hybrid model performs better than either of the other two approaches. Erev & Roth (1998) find that reinforcement models perform better and Feltovich (1998) finds that specific model performance depends both on the design of the experiment and on the comparison criterion.

This paper proposes a model that combines insights from Fudenberg & Levine (1995) and Camerer & Ho (1998*c*) which in turn was built on Cheung & Friedman (1998) and Roth & Erev (1995). The model is fitted to data from an experiment conducted at the University of California, Santa Cruz and the resulting parameter estimates are compared to the models' predictions. Both representative agent estimations and individual parameter estimations are conducted and their results compared. Three by three bimatrix games are chosen because they provide a crucial stepping stone from the two by two bimatrix world to the more flexible multi-choice world. Ultimately, we want to know which is the better model for a specific class of problems.

2 The Basic Model

The games I will be considering in this paper have payoffs that depend on both your own action and the action of your opponent (your opponent can be a single person or you can be playing against a group of opponents). These games are repeated for a number of periods. Every period you and your opponent simultaneously make choices among the available actions. Everyone gets to see the payoff consequences of these actions and you get to make your next period's choice[1]. The payoffs are used to build the propensities for each action which are then used to predict probabilities of play. The propensity $P^i_{t,j}$ of individual i at time t for each action j is:

$$P^i_{t,j} = \frac{\sum_{\tau=1}^{t} (\phi^i)^{(t-\tau)} \pi^i_{\tau,j}(\delta^i, c^i_\tau, s^i_\tau)}{\sum_{\tau=1}^{t} (\phi^i)^{\tau-1}}. \tag{1}$$

The numerator is a discounted sum of the stream of payoffs over time and $\phi^i \in (0,1]$ is the discount. The payoff at time t to action j is $\pi^i_{t,j}(\cdot)$. The denominator normalizes the propensity so that the propensities at different times can be directly compared. A discount near zero, would indicate a Cournot type player who only uses payoff information from the last period.

[1]The model can be applied more generally than this discussion indicates. The structure of this game is introduced here for illustrative purposes only. This model is analyzed in Bouchez (2000).

While a discount of $\phi^i = 1$, averages of the payoffs to each choice over time (this is also known as a Fictitious Play player).

The payoff, or profit, to action j is the jth element of the vector $\pi_t^i(\delta^i, c_t^i, s_t^i)$ and is a function of δ^i, a weight on the importance of the payoffs to actions not chosen relative to that chosen. As δ approaches 0, actions not chosen have a smaller and smaller weight (they are no longer used) which corresponds to reinforcement learning. At $\delta = 1$ we get pure belief learning and the propensities are updated with all the payoff information. Intermediate values of δ can be thought of as weak belief learning where you don't put as much weight on actions you have not experienced. The action chosen by i at time t is a vector c_t^i (and is assumed to be a pure strategy vector[2]). The state faced by i (what i's opponents are doing) at time t is s_t^i. Note that in both c_t^i and s_t^i, the elements sum to one since they represent distributions of play over possible actions.

The actual payoff function used in this analysis and in the experiments is a function of a payoff matrix M (with elements m_{kl}). For a three choice game, M is a 3x3 matrix and the payoff function is:

$$\pi_t^i(\delta^i, c_t^i, s_t^i) = \begin{bmatrix} \pi_{t,1}^i(\delta^i, c_t^i, s_t^i) \\ \pi_{t,2}^i(\delta^i, c_t^i, s_t^i) \\ \pi_{t,3}^i(\delta^i, c_t^i, s_t^i) \end{bmatrix} \quad (2)$$

$$= \left(diag \begin{bmatrix} c_1^i + \delta^i c_2^i + \delta^i c_3^i \\ \delta^i c_1^i + c_2^i + \delta^i c_3^i \\ \delta^i c_1^i + \delta^i c_2^i + c_3^i \end{bmatrix} \right) \cdot M \cdot \begin{bmatrix} s_1^i \\ s_2^i \\ s_3^i \end{bmatrix} \quad (3)$$

$$= \begin{bmatrix} (c_1^i + \delta^i c_2^i + \delta^i c_3^i)(m_{11} s_1^i + m_{12} s_2^i + m_{13} s_3^i) \\ (\delta^i c_1^i + c_2^i + \delta^i c_3^i)(m_{21} s_1^i + m_{22} s_2^i + m_{23} s_3^i) \\ (\delta^i c_1^i + \delta^i c_2^i + c_3^i)(m_{31} s_1^i + m_{32} s_2^i + m_{33} s_3^i) \end{bmatrix} \quad (4)$$

If an individual i had played $c_2^i = [1, 0, 0]'$ period 2 and had been faced

[2] This assumption is not necessary but simplifies the presentation of the model.

by as state $s_2^i = [\frac{1}{3}, \frac{1}{2}, \frac{1}{6}]'$, her payoffs that period would be:

$$\pi_t^i(\delta^i, c_t^i, s_t^i) = \begin{bmatrix} (1+\delta^i 0+\delta^i 0)(m_{11}\,\frac{1}{3}+m_{12}\,\frac{1}{2}+m_{13}\,\frac{1}{6}) \\ (\delta^i 1+0+\delta^i 0)(m_{21}\,\frac{1}{3}+m_{22}\,\frac{1}{2}+m_{23}\,\frac{1}{6}) \\ (\delta^i 1+\delta^i 0+0)(m_{31}\,\frac{1}{3}+m_{32}\,\frac{1}{2}+m_{33}\,\frac{1}{6} \end{bmatrix} \tag{5}$$

$$= \begin{bmatrix} 1\,(m_{11}\,\frac{1}{3}+m_{12}\,\frac{1}{2}+m_{13}\,\frac{1}{6}) \\ \delta^i\,(m_{21}\,\frac{1}{3}+m_{22}\,\frac{1}{2}+m_{23}\,\frac{1}{6}) \\ \delta^i\,(m_{31}\,\frac{1}{3}+m_{32}\,\frac{1}{2}+m_{33}\,\frac{1}{6}) \end{bmatrix} \tag{6}$$

If $\delta = 1$, this model is equivalent to a belief model where all possible payoffs are weighted equally. If $\delta = 0$, we have the reinforcement learning model where only the action chosen is used in the propensity. Relaxing the assumption that the discount values in the numerator and denominator are equal makes this model asymptotically equivalent to a continuous time version of the EWA model[3].

The model is a probabilistic model, the probability of an individual

[3]The EWA model, as presented in Camerer and Ho (1998a,b,c), consists of the observation equivalent of past experience N(t) and the propensity $P_i^j(t)$ for action j at time t. The initial N(0) and P(0) are updated after period 0 so that:

$$N(t) = \rho \cdot N(t-1) + 1, t \geq 1$$

and,

$$P_i^j(t) = \frac{\phi \cdot N(t-1) \cdot P_i^j(t-1) + [\delta + (1-\delta) \cdot I(s_i^j, s_i(t))] \cdot \pi_i(s_i^j, s_{-i}(t))}{N(t)}, t \geq 1.$$

This is equivalent to

$$N(t) = \rho^t \cdot N(0) + \sum_{\tau=1}^{t} \rho^{\tau-1}$$

and,

$$P_i^j(t) = \frac{\phi^t N(0) P(0) + \sum_{\tau=1}^t \phi^{t-\tau}[\delta + (1-\delta) \cdot I(s_i^j, s_i(t))] \cdot \pi_i(s_i^j, s_{-i}(t))}{\rho^t \cdot N(0) + \sum_{\tau=1}^t \rho^{\tau-1}}.$$

For $0 < \phi, \rho < 1$, this is asymptotically equivalent to

$$P_i^j(t) = \frac{\sum_{\tau=1}^t \phi^{t-\tau}[\delta + (1-\delta) \cdot I(s_i^j, s_i(t))] \cdot \pi_i(s_i^j, s_{-i}(t))}{\sum_{\tau=1}^t \rho^{\tau-1}}.$$

choosing an action increases as the propensity for that action increases. To map the propensities on to actions, the Logit probability response function is used so that the probability of i's jth action $c^i_{t,j}$ at time t is:

$$Prob(c^i_{t,j} = j) = \frac{e^{\lambda^i \cdot P^i_{t,j}}}{\sum\limits_j e^{\lambda^i \cdot P^i_{t,j}}} \tag{7}$$

where λ^i is the parameter. A λ^i near zero would indicate a misspecified model while a large λ^i would indicate that the propensities do a good job of explain the observed play. Negative λ^i values would suggest misspecification, possibly due to higher order reasoning and anticipatory play in the sense of Selten (1991).

The Logit has been used widely used in the literature on Learning (Mookherjee & Sopher 1994, McKelvey & Palfrey 1995, Fudenberg & Levine 1998, Camerer & Ho 1998*b*). Other mapping functions are possible, but work by Camerer & Ho (1998*c*) has shown little difference between the Logit and power response functions. Tang (1996) had similar results looking only at reinforcement models. The Logit has the additional benefit of allowing negative payoffs.

The Logit is invariant to an additive constant on the propensities. In order to estimate the model the probabilities are normalized relative to one of the actions. The use of the Logit assumes the independence of irrelevant alternatives: that the ratio of the probabilities of any two actions j and k, P_j/P_k, be independent of the remaining probabilities (Green 1993). This implies that adding an alternative to the model, or changing the characteristics of another alternative that is already included, will not change the odds between actions j and k (Davidson & Mackinnon 1993), a plausible constraint in this class of games.

3 Experiment

3.1 Lab procedures

In order to maintain the anonymity of the subjects, the experiments were conducted at visually isolated computer stations in the UCSC computer labs. The entire experiment is scripted to maintain consistency between sessions. The subjects were given written instructions which are then covered fully by the experiment monitor (see Bouchez (2000) for the experiment instructions). The subjects were told that the other subjects with whom they interacted do not necessarily have the same matrix. A brief (2 period) scripted practice run was conducted to familiarize the subjects with the computer interface. During the instruction period, subjects were free to ask for questions and all non-strategic questions were answered.

The actual experiment was conducted in silence (questions can be directed privately to the experiment monitor and are answered at her discretion). No strategic questions were answered. Eight to ten runs were conducted, and each new run was announced and was preceded by a brief pause. Subjects were not told the number of runs nor the exact ending time of the experiment to rule out any end of experiment effects.

The points that each subject accumulates over the course of the different runs were converted to cash payments at the end of the experimental session. The experiments lasted approximately 2 hours and the subjects averaged $ 10/hour (actual payments depended on their accumulated points and varied from $12 to $24).

3.2 Experiment design

The data used to fit the model comes from an experiment run with groups of 9 to 20 undergraduates at UCSC. Within each two hour session they experience between 8 to 10 runs of 12 to 24 periods each. The runs have a variety of different payoff bimatrices, but in any given session all runs have the same information environment.

The bimatrices in this analysis were chosen to illustrate cases when Nash Equilibrium theory was insufficient to explain the experimental evidence. Two single population games were chosen, one with only a pure strategy NE at (1,0,0) is from Stahl & Wilson (1994):

$$\text{Matrix 2} = \begin{bmatrix} 4 & 1 & 7 \\ 2 & 8 & 0 \\ 3 & 10 & 6 \end{bmatrix}. \tag{8}$$

The other is a Hawk Dove Bourgeois (HDB) game from Cheung & Friedman (1997) with both a pure strategy NE (0,0,1) and a mixed strategy NE (2/3,1/3,0):

$$\text{Matrix 4} = \begin{bmatrix} -2 & 8 & 3 \\ 0 & 4 & 2 \\ -1 & 6 & 4 \end{bmatrix}. \tag{9}$$

Four two population games are also used, one symmetric from Shubik (1996), a Shapley shimmy with the only NE at (1/3,1/3,1/3):

$$\text{Matrix 1} = \begin{bmatrix} 0 & 5 & 4 \\ 4 & 0 & 5 \\ 5 & 4 & 0 \end{bmatrix}. \tag{10}$$

The other two are non symmetric and come from Tang (1996). Matrix 17 satisfy's Selten's local stability criterion (Selten 1991) for his Anticipatory learning model and has a mixed NE at (1/6,1/3,1/2),(1/6,1/3,1/2). The other, Matrix 22, satisfies, Selten's local instability criterion and also has a NE at (1/6,1/3,1/2),(1/6,1/3,1/2):

$$\text{Matrix 17 group1} = \begin{bmatrix} 20 & 8 & 8 \\ 5 & 20 & 5 \\ 0 & 0 & 20 \end{bmatrix} \tag{11}$$

$$\text{Matrix 17 group2} = \begin{bmatrix} 0 & 12 & 12 \\ 16 & 4 & 12 \\ 16 & 10 & 8 \end{bmatrix} \quad (12)$$

and

$$\text{Matrix 22 group1} = \begin{bmatrix} 4 & 10 & 12 \\ 15 & 0 & 15 \\ 18 & 0 & 14 \end{bmatrix} \quad (13)$$

$$\text{Matrix 22 group2} = \begin{bmatrix} 0 & 15 & 10 \\ 12 & 6 & 12 \\ 16 & 10 & 8 \end{bmatrix}. \quad (14)$$

The runs were conducted under a "full history" environment where the subjects are able to see a complete history of every period played during the run. This history included the individual subject's choice and payoff each period, the choice distribution of all players in the opponent population for each period, and the time average choice distributions from the start of the run.

"Mean matching," where each individual is matched against the average action of all other players[4], is used throughout all runs. Although some information may be lost by not varying the matching protocol, Cheung and Friedman (1998) show no great difference between mean matching and randomly pairing participants beyond a slightly more rapid convergence.

A variety of other treatment variables are also available but not exploited: the number of subjects, sub-population versus full population, the experience of the players, and the number of periods played. In past work (Friedman 1996, Cheung & Friedman 1998), the number of subjects does not seem to make a great deal of difference unless very small populations

[4]Note that since the payoff function used in this work is linear, mean matching is equivalent to playing against each of the people in the opponent populations and receiving the average payoff.

are chosen. The number here (9-20 subjects per session) should be adequate for avoiding these problems. The number of subjects varies for practical, subject recruitment reasons, as well as because in asymmetric matrix treatments it is important to have sufficiently large half populations.

Since the interest in this study concerns "intermediate term" learning, the number of periods in most runs ranged between 10 and 20. Most runs were 12 periods. In one of the games (matrix 2), however, behaviour did not appear to be settling down, and subsequent runs were set at 24 periods. The majority of the players used in these experiments were inexperienced. They had never seen nor participated in a bimatrix experiment before.

The principal treatment variable is the payoff matrix. Six matrices are used in this analysis. Two of the matrices were single group matrices where the players are all in one group. The other four matrices were run under opposing group treatments, one symmetric and the other three asymmetric. Details of treatments and the individual experiment runs can be found in Bouchez (2000).

4 Overview of the experimental data

There are a minimum of 14 (and a maximum of 24) runs for each of the matrices used in this paper. Appendix A provides graphs of selected two period time averages of play. There is some variation between matrices, between groups of experimental subjects and between runs with the same subjects. There are however quite a few similarities between runs of the same matrix.

4.1 Starting points

Figures 1-4 show the first period averages of play for each matrix. The mean (and standard deviation) for each of the matrices are listed in Table 1

Table 1: First period of play averages -Single population Games (with standard deviations) *indicates that the mean is significantly different from random play (1/3) at the 10% level. ** at the 5% and *** at the 1% level)

Matrix	Runs	Top Action %	Middle Action %	Bottom Action %
2	18	0.22*** (0.14)	0.24*** (0.14)	0.53*** (0.19)
4	14	0.29 (0.18)	0.29 (0.16)	0.42* (0.17)

and Table 2. There is substantial variability between the starting points of different matrices and many are significantly different from random play (a probability of 1/3 on each of the three possible actions). The variation in starting points is of importance when deciding how to treat initial periods of play (the model does not attempt to fit the initial periods). This will be taken up again in the methods section (5.1) below. The starting points for matrix 2 are fairly dispersed but none are near the pure strategy NE (1,0,0). The first period actions for Matrix 4 diverge from the pure and the mixed strategy NE.

4.2 End points

Figures 5-8 show the last two period averages of play and Nash equilibria for each matrix. The graphs for Matrix 2 and Matrix 4 have a grouping of final period actions near (but not at) the single periods pure strategy Nash equilibrium. There is, however, no evidence that the pure strategy NE would be reached if the number of periods were extended. The number of periods of play was increased for Matrix 2 (to 24 period runs) and there was no substantial difference in observed behaviour (see the two period averaged figures in Appendix A). Runs in excess of 24 period were not attempted. The end points for Matrices 1,7,17 and 22 are much more diverse.

Table 2: First period of play averages -Two population Games (with standard deviations) *indicates that the mean is significantly different from 1/3 at the 10% level. ** at the 5% and *** at the 1% level)

Matrix	Group	Runs	Top Action %	Middle Action %	Bottom Action %
1	1	23	0.20*** (0.20)	0.37 (0.20)	0.43* (0.23)
1	2	23	0.26** (0.17)	0.34 (0.22)	0.41* (0.23)
7	1	21	0.44** (0.20)	0.35 (0.21)	0.21*** (0.19)
7	2	21	0.41 (0.25)	0.23*** (0.16)	0.36 (0.24)
17	1	24	0.52*** (0.19)	0.16*** (0.14)	0.32 (0.20)
17	2	24	0.19*** (0.17)	0.30 (0.16)	0.50*** (0.19)
22	1	18	0.28 (0.19)	0.20*** (0.16)	0.52*** (0.20)
22	2	18	0.25** (0.15)	0.21*** (0.15)	0.53*** (0.16)

Figure 1: First period population averages of play for matrix 2 (single population). The points are labelled with the experiment and run numbers.

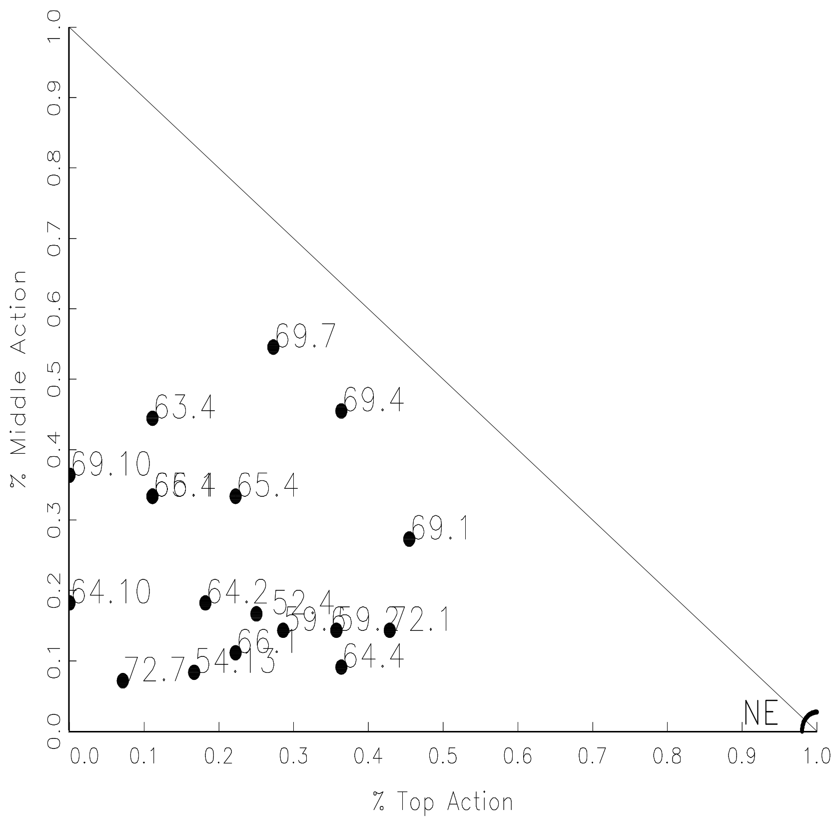


Figure 2: First period population averages of play for matrix 4 (single population)

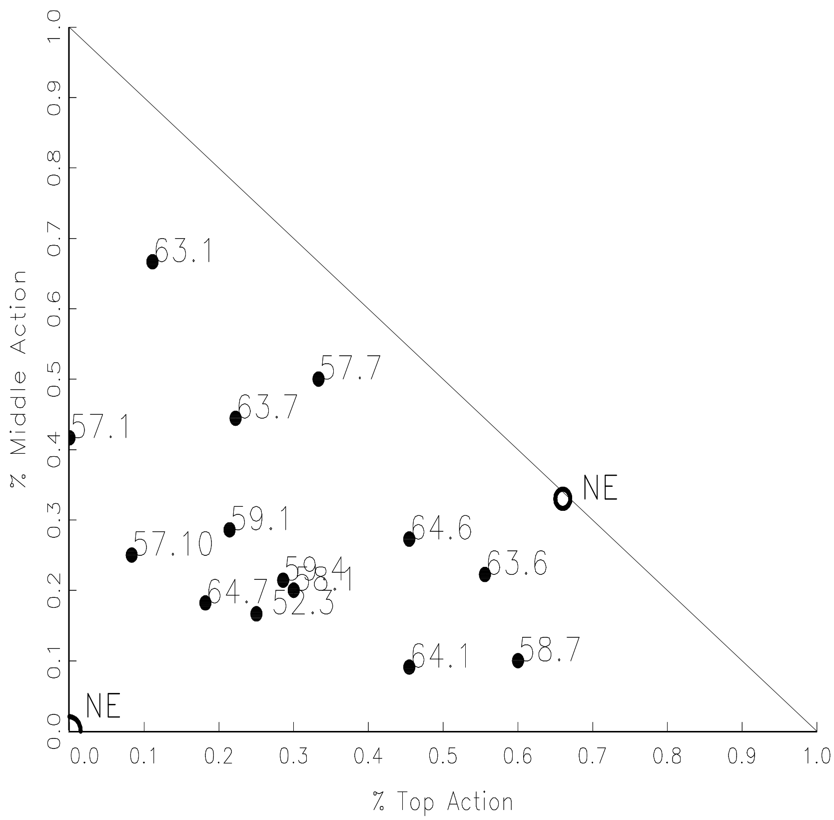


Figure 3: First period population averages of play for matrices 1 and 7 (opposing groups). The points are labelled with the experiment and run numbers.

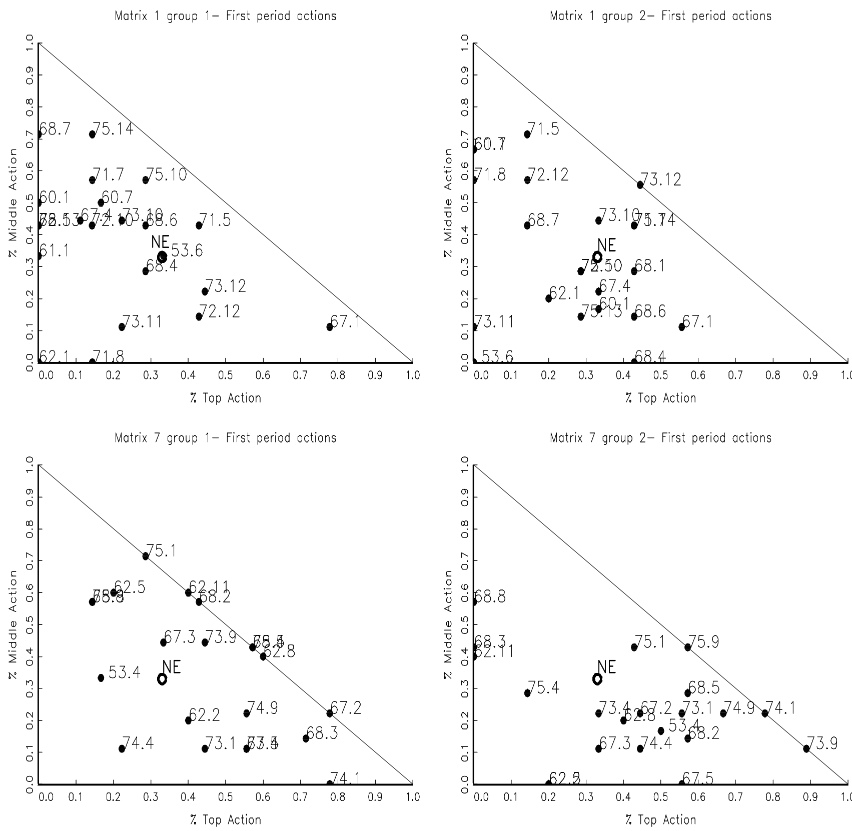


Figure 4: First period population averages of play for matrices 17 and 22 (opposing groups). The points are labelled with the experiment and run numbers.

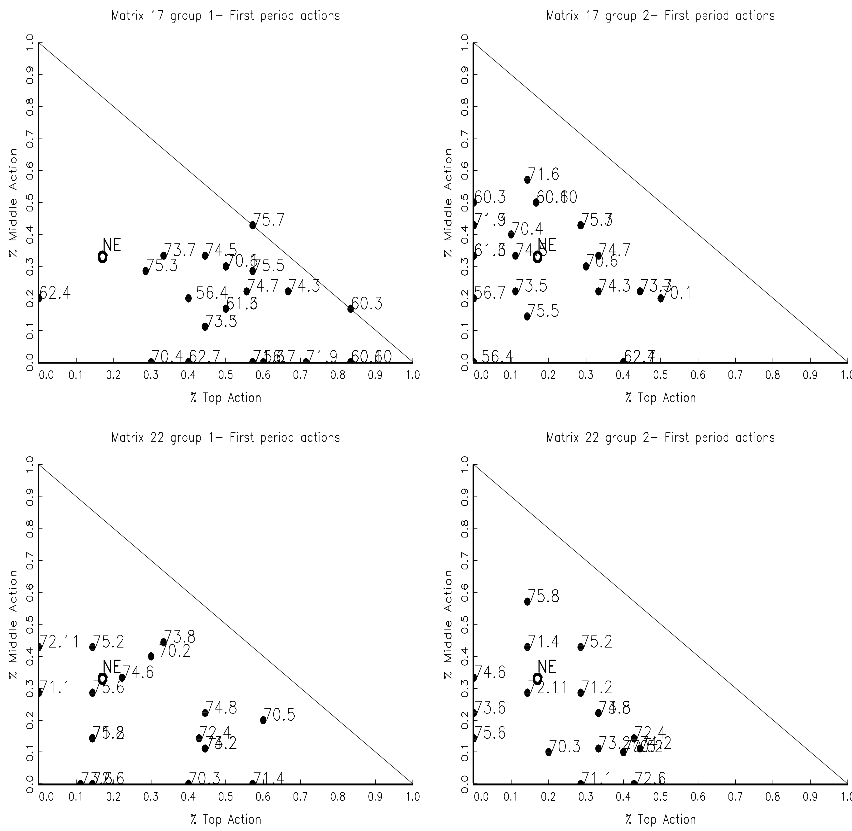


Figure 5: Last two period averages of play for matrix 2 (single population).

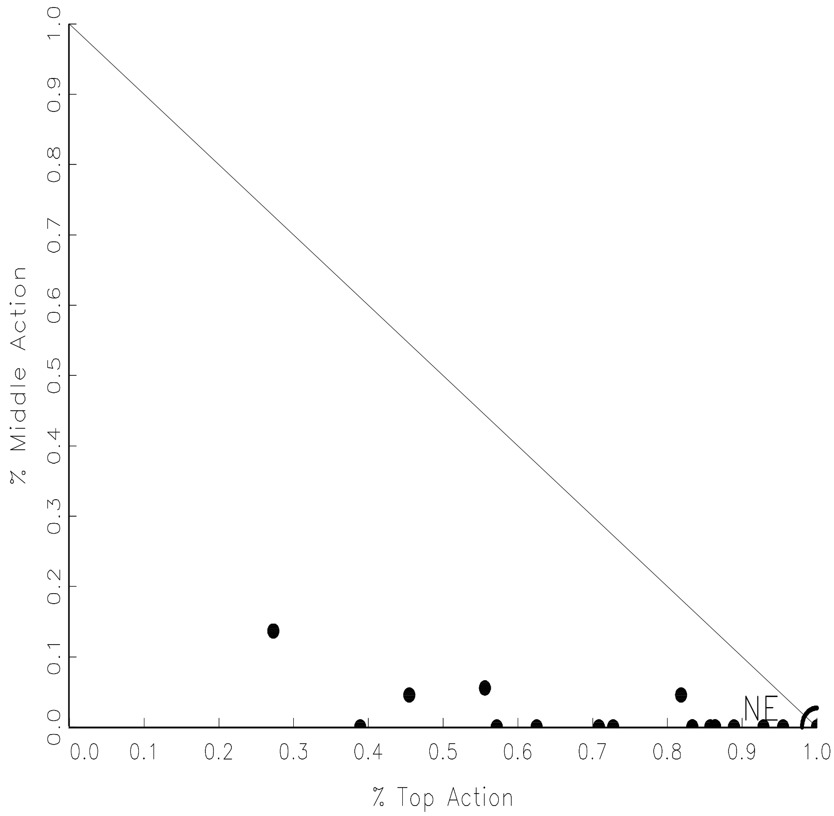


Figure 6: Last two period averages of play for matrix 4 (single population).

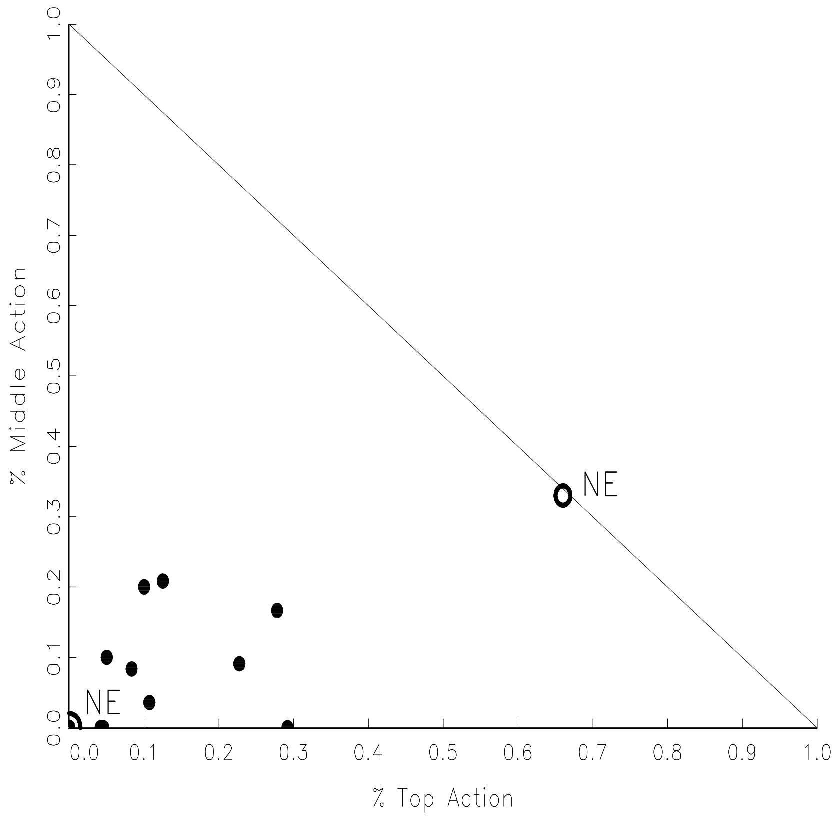


Figure 7: Last two period averages of play for matrices 1 and 7 (opposing groups).

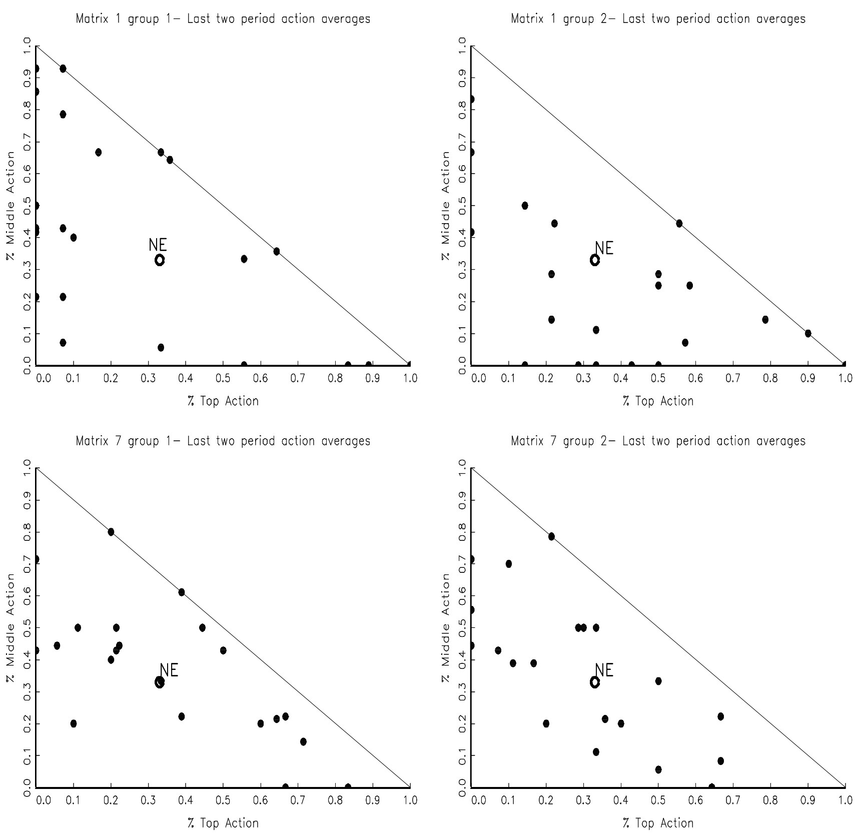
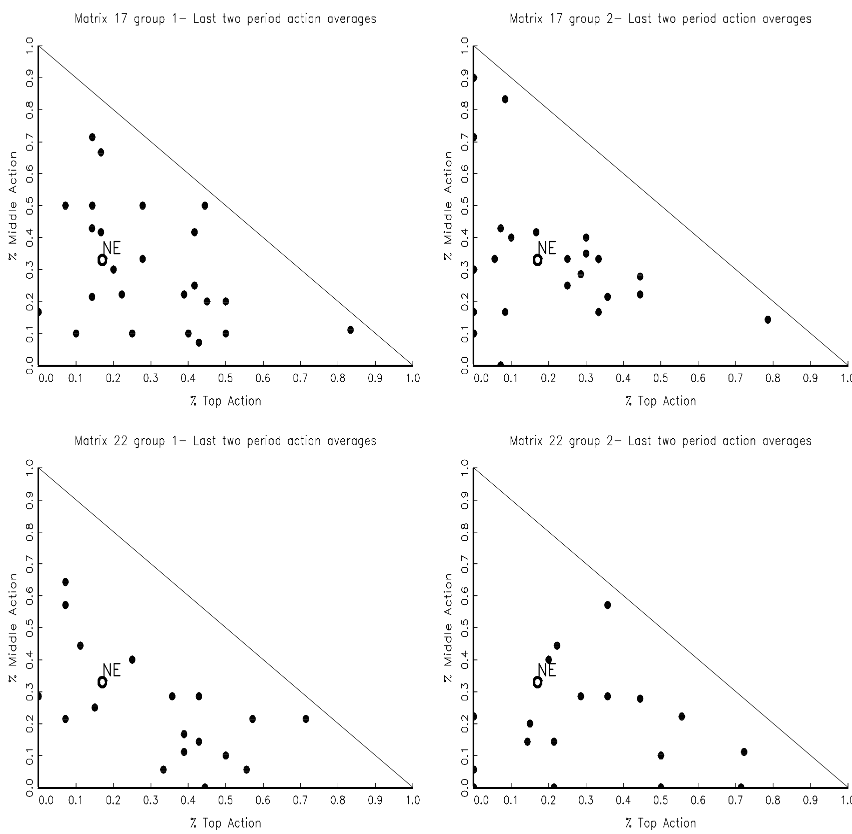


Figure 8: Last two period averages of play for matrices 17 and 22 (opposing groups).



5 Fitting the models

The model (Equation 1) is fitted using a Maximum likelihood estimation where the log likelihood function is

$$\ln L = \sum_{t} \sum_{j=T,M,B} c^{i}_{t,j} \ln Prob(c^{i}_{t,j} = j) \tag{15}$$

and $Prob(c^{i}_{t,j} = j)$ is the probability of i having chosen the jth pure strategy action $(c^{i}_{t,j})$ at time t (see Equation 7 for the logit probability mapping). The probability of play depends on three parameters in this specification: a discount ϕ, a logit parameter λ and δ, the weight of actions not chosen relative to actions chosen. A more general version of this model which is asymptotically equivalent to the Camerer & Ho (1998*a*) has two different discount rates with ρ substituted for ϕ in the denominator.

5.1 Methods

The Maximum likelihood equation was estimated using the CML procedure of GAUSS. The non-linear Broyden, Fletcher, Goldfarb and Shanno (BFGS) method is used to calculate the asymptotic covariance matrix of the parameter estimates as well as to calculate the Hessian (second derivative) approximations to be used during parameter iterations. This method produced virtually identical results to Newton's method (Bouchez 2000). BFGS is generally faster than Newton's method, though requiring more iterations. The line search method used is the polynomial or cubic fitting method STEPBT.

The parameter estimates are robust to starting points [5]. Thus, the starting points for all the estimated were set at $(\phi^{i}, \lambda^{i}\, \delta^{i}) = (0.5, 1.5, 0.5)$. All of

[5] This was determined in both the representative agent and the individual parameter estimates by varying the starting values. In the individual parameter estimates, the starting points had little to no effect. In the representative agent estimations, poorly chosen starting points can lead to non-convergence so non-convergent estimates are restarted at a different starting point.

Table 3: Convergence of Individual Parameter Estimates for the 3 parameter model (with initial propensities set the matrices' mean choice in the first period and the parameters are not estimated first period.)

Matrix	Group	# Indiv.	# Non Conv.	# At bounds	Estimable Individuals
2	-	54	5	0	49
4	-	32	0	0	32
1	-	74	0	1	73
7	1	41	0	1	40
7	2	41	2	5	36
17	1	44	3	6	38
17	2	44	2	3	40
22	1	49	2	5	44
22	2	49	2	5	42
Total		428	16	25	394

the estimates were loosely bounded (between 50 and -50 or 100 and -100) and non convergent estimates and estimates at the bounds were excluded. This procedure affected only one of the representative agent parameter estimations. This was more of an issue in individual parameter estimations where approximately 97% of the data was estimable (see Table 3[6]).

Since I am not estimating the initial propensities (P_0), the issue of starting points becomes important. Cheung & Friedman (1998) solve the problem by not estimating the first period (which is very effective with low ϕ values). I tried using all periods as well dropping the first and first two periods. Since the loss of power when dropping two periods is too substantial and leads to increased non-convergence, the estimates presented here have only one period of data dropped[7]. The fewer initial periods dropped, the larger the potential errors from poorly specified initial propensities. To minimize this problem, initial propensities were set at the appropriate matrices' first period

[6]Please note that an individual estimate could have been non-convergent *and* have parameters at the bounds as well.

[7]Individual parameter estimate distributions do not change much as the number of periods dropped is varied.

averages of play. This is not a problem for Camerer and Ho (1998a,1998b, and 1998c) since they explicitly estimate the initial propensities as part of their estimation.

5.2 Results

5.2.1 Representative Agent Estimations

The initial step of the analysis is to estimate representative agents for each matrix. This has been the predominant estimation method (Camerer & Ho 1998*c*) but constrains the parameters to be the same across individuals, something that is not suggested by theory or empirics.

Table 4 show the representative agent estimations by matrix. Note that for asymmetric matrices, the two matrices have been estimated separately. All of the estimates were convergent and were the same under BFGS and Newton's method (with the exception that 17-1 was at the bounds under BFGS and thus did not converge).

With the exception of matrix 17-1's negative ϕ estimate, all of the parameter estimates are in the expected ranges. The discounts ϕ are between -0.09 and 0.53, the δ estimates (the weight of actions not chosen relative to those chosen) are between 0.61 and 0.81, and the λ estimates are between 0.17 and 1.58. Using weak link games to estimate their model, Camerer & Ho (1998*b*) estimated $0.525 \leq \phi \leq .582$[8] and $\delta = 0.652$.

5.2.2 Individual Parameter Estimates

Individual parameter estimates were estimated in the same way as done in the representative agent model (e.g. initial propensities set at the matrix' first period averages of play and the first period data were not estimated). Only individuals who have a minimum of 36 periods of data are used (this

[8]The actual estimate of ϕ depends on the model estimated.

Table 4: Representative Agent Parameter estimates (and standard deviation) by Matrix. Estimates are the same using Newton's method or BFGS except for 17-1 which does not converge under BFGS.

Matrix	Periods	ϕ	λ	δ	LogL	LogL/# periods
Pooled	16020	0.24	0.45	0.70	-12900.30	.8053
		(0.00)	(0.00)	(0.00)		
2	1932	0.38	1.20	0.79	-1079.42	.5587
		(0.00)	(0.00)	(0.00)		
4	1152	0.53	1.58	0.63	-656.57	.5699
		(0.01)	(0.01)	(0.00)		
1	3832	0.23	1.01	0.61	-2262.38	.5904
		(0.00)	(0.00)	(0.00)		
7-1	1560	0.09	0.44	0.62	-1285.56	.8241
		(0.00)	(0.00)	(0.00)		
7-2	1560	0.22	0.565	0.66	-1268.83	.8133
		(0.00)	(0.00)	(0.00)		
17-1	1728	-0.09	0.17	0.65	-1539.16	.8907
		(0.00)	(0.00)	(0.00)		
17-2	1728	0.10	0.35	0.81	-1407.71	.8146
		(0.00)	(0.00)	(0.00)		
22-1	1764	0.28	0.29	0.64	-1444.68	.8190
		(0.00)	(0.00)	(0.00)		
22-2	1764	0.17	0.34	0.70	-1390.56	.7883
		(0.00)	(0.00)	(0.00)		

corresponds to three runs of 12 periods or two runs of 24 periods). Individual parameter distributions allow the parameter values to be tested for model miss-specification, but more importantly, they attempt to answer the questions of how much individual variation there is.

Figures 9-11 show the kernel density plots for the individual parameter estimates by matrix while Figures 12-14 show the individual parameter estimates pooled across matrices. Table 5 provides the summary statistics. The kernel density estimates use the GAUSS library for kernel estimation by Konning (1996). The Epanechnikov (1969) kernel was used in the analysis[9].

The δ parameter distributions (Figures 9 and 12) are primarily between 0 and 1, with the modes between 0.69 and 0.93. It is, however, interesting to note that several of the matrices (1, and 22 for both groups) have some weight at or near zero, as is the case when pooled across matrices (figure

[9]The kernel density estimates use the GAUSS library for kernel estimation by Konning (1996) and implement the methods described by Härdle (1990) and Silverman (1986). The Epanechnikov kernel was used; Gaussian, rectangular and triangular kernels were checked but did not alter the results. This is consistent with Wand & Jones (1995) who show that most unimodal densities perform about the same when used as kernel densities. It is however important to note that as the degree of skewness, kurtosis and multimodality increase, the estimation becomes more difficult. The Epanechnikov kernel is defined as:

$$K(u) = \frac{3}{4\sqrt{5}}(1 - \frac{1}{5}u^2)1_{|u| \leq \sqrt{5}} \qquad (16)$$

with its derivative

$$K'(u) = \frac{3}{10\sqrt{5}}u \ 1_{|u| \leq \sqrt{5}} \qquad (17)$$

where u is the vector of data points and $1_{|u| \leq \sqrt{5}}$ is an indicator which is equal to 1 when $|u| \leq \sqrt{5}$ and 0 when not . It is the most efficient kernel when compared to the optimal kernel, the asymptotic mean integrated squared error-AMISE (Wand and Jones 1995). The bandwidth was calculated following Silverman (1986) so that the bandwidth h is:

$$h = \frac{0.9 \ min(s, IQR/1.34)}{n^{1/5}} \qquad (18)$$

with s the sample standard deviation of the data points and IQR the inter-quartile range of the data points.

Figure 9: The distribution of individual subject estimates of the δ parameter. The dashed vertical lines are at zero and one. The solid line indicates the representative agent estimate.

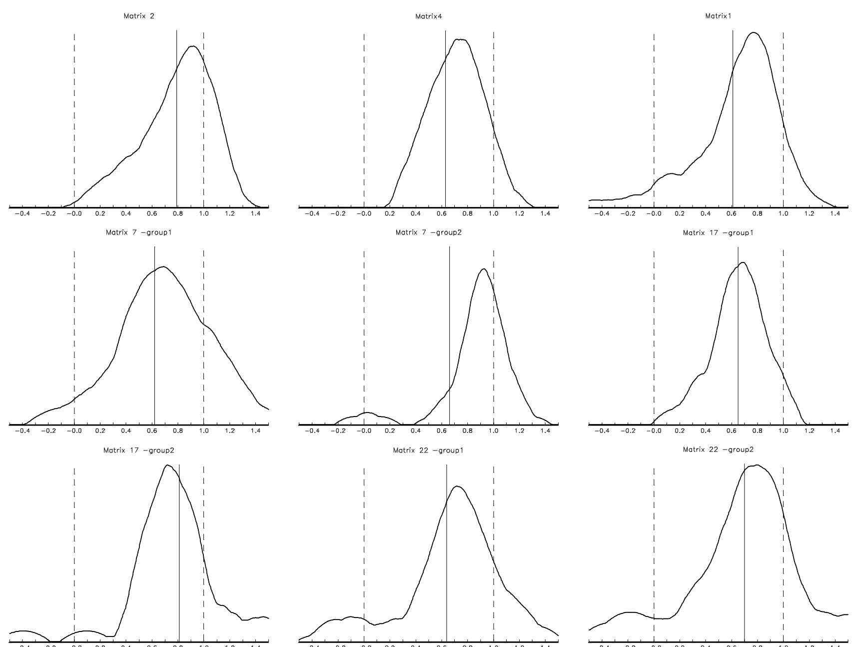


Figure 10: The distribution of individual subject estimates of the Logit parameter (λ). The dashed vertical lines are at zero and one. The solid line indicates the representative agent estimate.

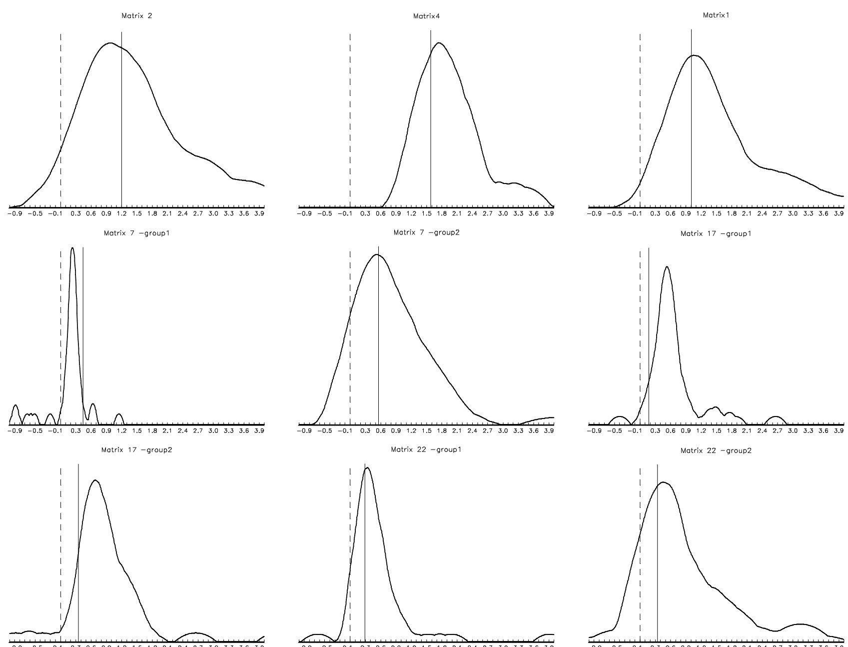


Figure 11: The distribution of individual subject estimates of the discount parameter (ϕ). The dashed vertical lines are at zero and one. The solid line indicates the representative agent estimate.

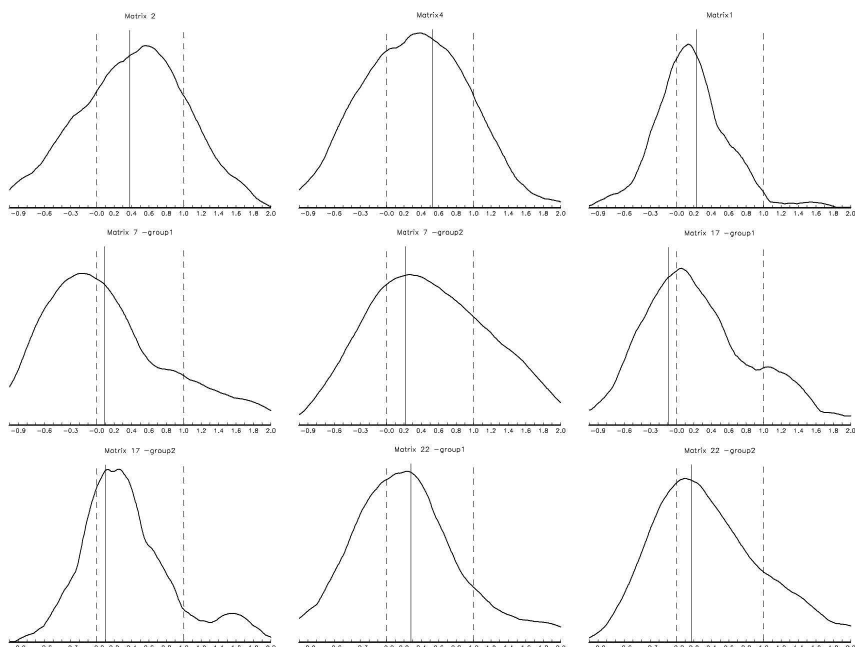


Figure 12: Individual subject δ parameter estimates pooled across matrices. The dashed vertical lines are at zero and one. The solid line is at the representative agent parameter estimate.

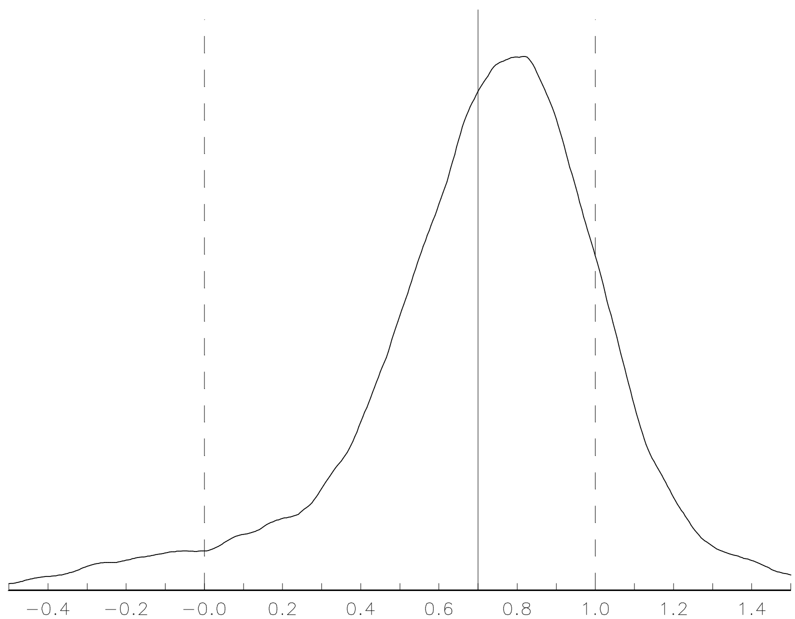


Figure 13: Individual subject λ parameter estimates pooled across matrices. The dashed vertical lines are at zero and one. The solid line is at the representative agent parameter estimate.

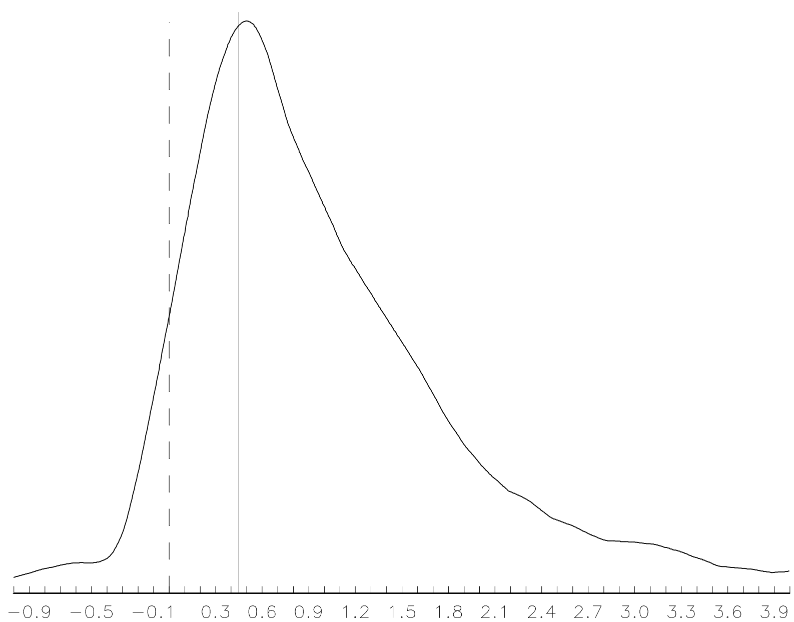
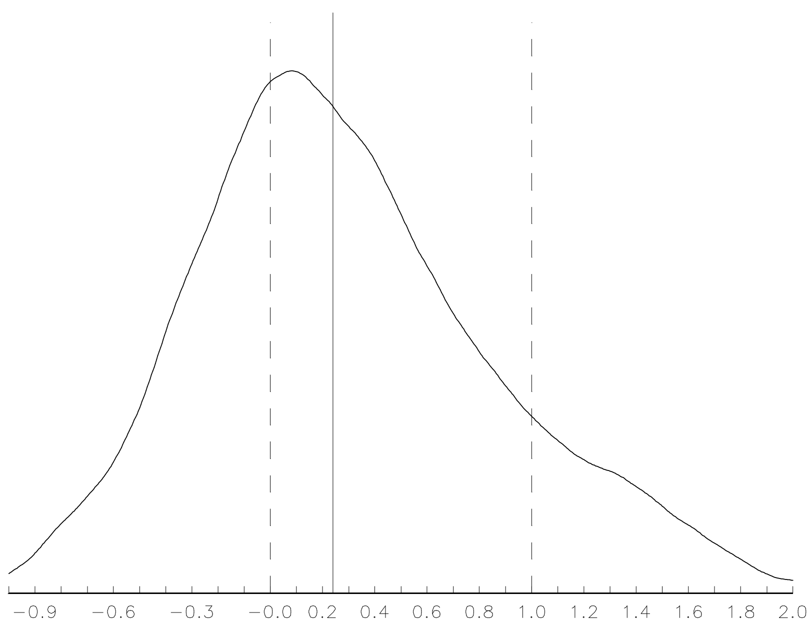


Figure 14: Individual subject ϕ parameter estimates pooled across matrices. The dashed vertical lines are at zero and one. The solid line is at the representative agent parameter estimate.



14). Since belief learning has implicitly assumed a $\delta = 1$, I will be testing this constraint in the next section.

The parameter estimates for the logit parameter λ (Figures 10 and 13) are very dispersed in some matrices (2,4,1,17-2,22-2) and less so in others (7,17-1,22-1). The mass of the distributions is mostly above zero, which is consistent with the model. The value of λ depends strongly on the matrix and is related to the scale of the matrix (i.e.: scaling up the payoff matrix up would lower λ), and thus the distribution of parameters pooled across matrices (Figure 13) provides little information beyond there being little weight below zero.

The discount parameter ϕ estimates (Figures 11 and 14) are very dispersed in all of the matrices, and have much of their distributions below zero. It is important to note that this is not unexpected given the high values of δ. The simulations in Bouchez (2000) clearly indicate that values of δ close to 1 lead to more dispersed ϕ parameter estimates. It is also important to note that the matrices with the most period to period movement (matrices 7, 17, 22 -see figures in Appendix A) have the lowest discounts. In most of the matrices, there is very little weight above one. Players with discounts greater than one can be interpreted as imprintable: early periods have a greater impact than subsequent periods. The pooled parameter estimates have a mode near 0.1, but here again have a wide distribution.

Comparing the individual to the representative agent parameter estimates, the latter are fairly good at identifying the mode of the individual estimates. This is especially the case with the pooled estimates (figures 4-6). It may be rational to "forget" past periods when faced with a rapidly changing environment. When looking at the parameter estimates by matrix, there is a little more variability. The representative agent estimates of λ are very close to the mode of the distributions, as are the estimates of δ. The ϕ estimates are more dispersed, and especially in the case of matrix 7-1, the representative agent estimation is not a very good indicator of the mode of the distribution.

6 Hypothesis tests

6.1 The discount parameters

The EWA model proposed by Camerer & Ho (1998*a*) has two different discounts (which have been collapsed into one in the model presented here). The discount ρ is separated from ϕ in the denominator of the model so that

$$P_{t,j}^i = \frac{\sum_{\tau=1}^t (\phi^i)^{(t-\tau)} \pi_{\tau,j}^i(\delta^i, c_\tau^i, s_\tau^i)}{\sum_{\tau=1}^t (\rho^i)^{\tau-1}}. \tag{19}$$

in the discrete version of the model[10]. To test the 3 parameter restriction, I use the likelihood ratio test. The estimates of the 4 parameter model were done in the same way as the estimates for the three parameter model with initial propensities set to the first period averages of play and the first period is excluded from the estimates. There are more cases of non-convergence than before –40% versus 8%– and non-estimable individuals[11] are excluded.

The likelihood ratio test statistic LR under the null hypothesis that the constraint, is true is

$$LR = 2(\ln L_u - \ln L_r) \sim \chi^2_{\alpha,k} \tag{20}$$

where $\ln L_u$ and $\ln L_r$ are the unrestricted and restricted log likelihoods, and the degrees of freedom k is equal to the number of restrictions (1 in this case).

Using the representative agent data (table 6), we fail to reject the restriction that the two discounts are equal at the 10% confidence level in all

[10]The EWA model also includes initial propensities which are also estimated so this model is still not identical to Camerer & Ho (1998*a*) but is asymptotically equivalent.

[11]Individuals with non-convergent estimates or estimates at the bounds in either the 3 or 4 parameter models.

Table 5: Individual Parameter estimates summary statistics for the 3 parameter model (with initial propensities set at the first period averages of play and not estimating the first period)

Matrix		ϕ	λ	δ
2	Mean	0.41	2.41	1.54
	(SD)	(0.57)	(2.97)	(4.73)
	Median	0.46	1.34	0.89
	Mode	0.56	0.98	0.92
4	Mean	1.08	2.47	0.65
	(SD)	(4.26)	(1.91)	(0.39)
	Median	0.38	1.97	0.71
	Mode	0.39	1.74	0.72
1	Mean	0.26	2.63	0.61
	(SD)	(0.65)	(7.57)	(0.38)
	Median	0.16	1.28	0.72
	Mode	0.14	1.05	0.77
17-1	Mean	0.54	0.02	0.70
	(SD)	(3.49)	(1.00)	(0.79)
	Median	-0.07	0.23	0.66
	Mode	-0.17	0.23	0.69
17-2	Mean	1.18	2.33	0.88
	(SD)	(3.34)	(8.49)	(0.57)
	Median	0.45	0.66	0.91
	Mode	0.27	0.52	0.93
7-1	Mean	0.26	0.66	0.55
	(SD)	(0.63)	(0.52)	(0.69)
	Median	0.16	0.55	0.66
	Mode	0.06	0.53	0.69
7-2	Mean	0.98	1.01	0.82
	(SD)	(3.24)	(1.04)	(0.41)
	Median	0.26	0.73	0.79
	Mode	0.26	0.68	0.72
22-1	Mean	0.99	0.62	0.80
	(SD)	(1.94)	(1.28)	(1.19)
	Median	0.25	0.40	0.73
	Mode	0.24	0.34	0.72
22-2	Mean	1.39	0.84	0.69
	(SD)	(6.27)	(0.87)	(0.47)
	Median	0.27	0.61	0.77
	Mode	0.10	0.45	0.80

Table 6: Testing $\phi = \rho$ with representative agent parameter estimates and likelihood ratio tests. The 4 parameter model estimates failed to converge in some cases using BFGS (indicated by *) so those estimates were done using Newton's method (the other results were same using BFGS or Newton's method).

Matrix	$\ln L_r$ ($\phi = \rho$)	$\ln L_u$ ($\phi \neq \rho$)	χ^2 Stat. (p-value)
Pooled	-12900.39	-12892.57	15.63 (0.00)
2	-1079.42	-1072.39	14.06 (0.00)
4	-656.57	-655.32	2.50 (0.11)
1	-2262.38	-2260.03	4.69 (0.03)
7-1	-1285.56	-1285.53	0.07 (0.79)
7-2	-1268.83	-1268.05	1.55 (0.21)
17-1	-1539.16*	-1539.07	0.18 (0.67)
17-2	-1407.71	-1407.53*	0.37 (0.53)
22-1	-1444.68	-1443.9*	1.58 (0.21)
22-2	-1390.56	-1390.21*	0.71 (0.40)

Table 7: Testing if $\delta = 1$. Representative agent likelihood ratio tests.

Matrix	$\ln L_r$ ($\phi = 1$)	$\ln L_u$ ($\phi \neq 1$)	χ^2 Stat. (p-value)
Pooled	-14103.76	-12900.39	2406.74 (0.00)
2	-1194.46	-1079.42	230.09 (0.00)
4	-755.32	-656.57	197.49 (0.00)
1	-2720.28	-2262.38	915.81 (0.00)
7-1	-1398.01	-1285.56	224.90 (0.00)
7-2	-1388.20	-1268.83	238.74 (0.00)
17-1	-1604.80	-1539.16	131.28 (0.00)
17-2	-1485.13	-1407.71	154.84 (0.00)
22-1	-1631.14	-1444.68	372.91 (0.00)
22-2	-1562.76	-1390.56	344.38 (0.00)

but two cases (Matrices 1 and 2) the results are not strongly in favour of the alternate, especially in the case of the asymmetric two population matrices (matrices 7, 17, and 22).

When looking at individual parameter estimates, the conclusions are similar. Of 214 estimable individuals, we fail to reject the null hypothesis that the two discounts are equal (at the 10% level) in 29 individuals (13% of the estimable individuals) while at the 25% level that climbs to 80 individuals (37%). There is little or not loss in assuming $\rho = \phi$.

6.2 Testing δ

The other restriction tested is whether $\delta = 1$. This is the assumption of pure belief learning models (as opposed to rote models where $\delta = 0$ –which is clearly rejected). From the parameter distributions (Figures 9, 12 and Table 5) the modes of the distributions are between 0.69 and 0.93, with medians between 0.66 and 0.91.

In the case of the representative agent estimations (Table 7), there is no evidence to support the restriction $\delta = 1$. This is similar when testing the individual parameter estimates where we fail to reject the null at the 10% level in 3 out of 245 individuals (1% of estimable individuals). These conclusions are not consistent with the parameter distributions (Figure 9) which show δ values near one. I conclude that $\delta = 1$ is a decent first approximation in many cases, but it does not hold strictly.

7 Conclusion

Overall, the parameter estimates are in the expected ranges in both the representative agent estimations and the individual estimations. The representative agent parameter estimates are fairly good at estimating the mode of the individual estimates in many (but not all) cases.

Mixed evidence is found when testing $\phi = \rho$. Using representative agent data we reject the hypothesis at the 20% level. Using individual parameter data, the evidence is less clear and mildly in favour of the restriction (we find evidence in favour of $\phi = \rho$ in 13% of the estimates at the 10% level and in 37% of the estimates at the 25% level).

Both representative and individual agent estimates agree. The hypothesis tests reject the restriction that $\delta = 1$. The pure rote learning restriction that $\delta = 0$ is even more clearly rejected and the δ parameters are generally around 0.6-0.7. This means that we find strong evidence that individuals are closer to belief learners than rote learners and $\delta = 1$ is a decent first approximation in many cases.

The discrepancy between the two types of estimates in the hypothesis tests and the dispersion of the parameter estimates clearly point to the value of individual parameter estimates despite their data requirements. Representative agent estimates do not do a very good job testing restrictions on the model.

References

Börgers, T. & Sarin, R. (1996), Naive reinforcement learning with endogenous aspirations, University College London and Texas A&M University.

Bouchez, N. M. (2000), Learning models: what they tell us to expect and what we actually see, PhD thesis, University of California Santa Cruz.

Brown, G. (1951), Iterative solution of games by fictitious play, *in* T. C. Koopmans & A. Alchian, eds, 'Activity analysis of production and allocation : proceedings of a conference', number 13 *in* 'Monograph - Cowles Commission for Research in Economics'.

Bush, R. & Mosteller, F. (1955), *Stochastic models for learning*, Wiley, New York.

Camerer, C. & Ho, T. H. (1998*a*), 'Experience-weighted attraction learning in coordination games: Probability rules, heterogenity, and time-variation', *Journal of Mathematical Psychology.* **42**, 305–326.

Camerer, C. & Ho, T. H. (1998*b*), Experience-weighted attraction learning in games: Estimates from weak link games, *in* I. E. David V. Budescu & R. Zwick, eds, 'Games and human behavior: Essays in honor of Amnon Rapoport's 60th birthday', Erlbaum.

Camerer, C. & Ho, T. H. (1998*c*), 'Experience weighted attraction learning in normal-form games', *Econometrica.*

Cheung, Y.-W. & Friedman, D. (1997), 'Individual learning in games: Some laboratory results.', *Games and Economic Behavior* **19:1**, 46–76.

Cheung, Y.-W. & Friedman, D. (1998), 'An empirical comparison of learning and replicator dynamics', *Journal of Economic Behavior and Organization* **35**, 263–280.

Cournot, A. (1863), *Recherches sur les principes mathmatiques de la thorie des richesses*, Paris : L. Hachette.

Cross, J. G. (1983), *Theory of adaptive economic behavior*, Cambridge University Press, New York/London.

Davidson, R. & Mackinnon, J. G. (1993), *Estimation and Inference in Econometrics*, Oxford University Press.

Erev, I. E. & Roth, A. E. (1998), 'Predicting how people play games reinforcement learning in experimental games with unique, mixed strategy equilibria', *The American Economic Review.*

Feltovich, N. (1998), Reinforcement-based vs. beliefs-based learning models in experimental asymetric-information games, University of Houston.

Friedman, D. (1996), 'Equilibrium in evolutionary games: some experimental results', *The economic journal* **106**, 1–25.

Friedman, D., Massaro, D. W., Kitzis, S. N. & Cohen, M. M. (1995), 'A comparison of learning models', *Journal of mathematical psychology* **39**(2), 164–178.

Fudenberg, D. & Levine, D. K. (1995), 'Consistency and cautious fictitious play', *Journal of Economic Dynamics and Control* **19**, 1065–1089.

Fudenberg, D. & Levine, D. K. (1998), *The Theory of Learning in Games*, Cambridge: MIT Press.

Green, W. H. (1993), *Econometric Analysis*, 2 edn, Macmillan Publishing Company.

Harley, C. B. (1981), 'Learning in the evolutionary stable strategy', *Journal of theoretical biology* **89**, 611–633.

Kitzis, S., Kelley, H., Berg, E., Massaro, D. & Friedman, D. (1998), 'Broadening the tests of learning models', *Journal of mathematical psychology* **42**, 327–355.

McKelvey, R. D. & Palfrey, T. R. (1995), 'Quantal response equilibria for normal form games', *Games and Economic Behavior* **10**, 6–38.

Mookherjee, D. & Sopher, B. (1994), 'Learning behavior in an experimental matching pennies game', *Games and economic behavor* **7**, 67–91.

Roth, A. & Erev, I. (1995), 'Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term', *Games and Economic Behavior* **8**(1), 164–212.

Selten, R. (1991), Anticipatory learning in two-person games, *in* R. Selten, ed., 'Game Equilibrium Models I', Springer-Verlag.

Shubik, M. (1996), 'Why equlibirium? a note on noncooperative equilibria of some matrix games', *Journal of Economic Behavior & Organization* **29**, 537–539.

Stahl, D. O. (1993), 'Evolution of smart players', *Games and Economic Behavior* **5**, 604–617.

Stahl, D. O. (1996), 'Boundedly rational rule learning in a guessing game', *Games and Economic Behavior* **16**(2), 303–330.

Stahl, D. O. (1999*a*), 'Evidence based rule learning in symmetric normal-form games', *International Journal of Game Theory*.

Stahl, D. O. (1999*b*), Sophisticated learning and learning sophistication, Technical Report 99002, University of Texas.

Stahl, D. O. & Wilson, P. W. (1994), 'Experimental evidence on players' models of other players', *Journal of Economic Behavior and Organization* **25**, 309–327.

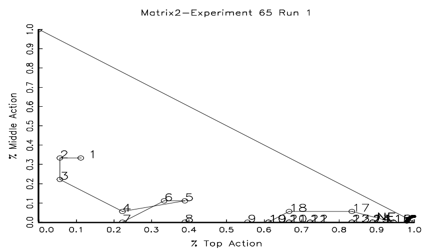
Tang, F.-F. (1996), Anticipatory Learning in Two-Person Games: An experimental study., PhD thesis, University of Bonn. Discussion Paper B362-B364.

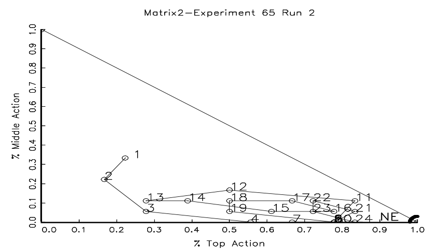
Wand, M. P. & Jones, M. C. (1995), *Kernel Smoothing*, Chapman & Hall, London.

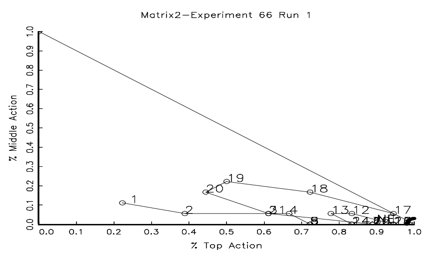
A Experimental Data - Selected Figures

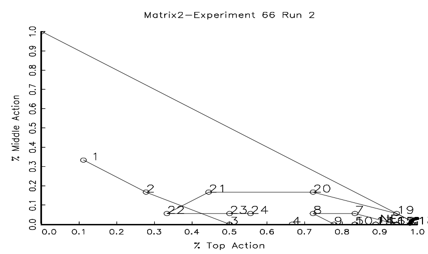
The following are the graphs of the subject choices over time (with two period smoothing) for some of the experiments. Two population games are shown by population.

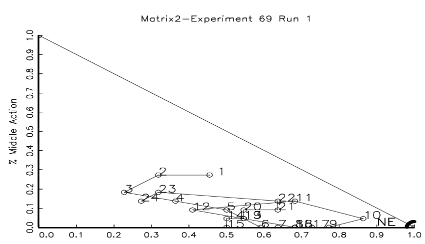
A.1 Matrix 2 -Single population 24 period sessions

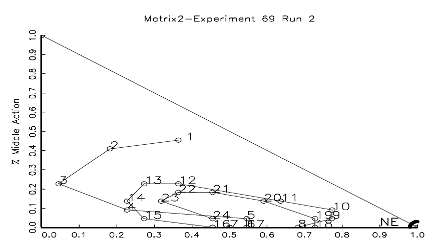


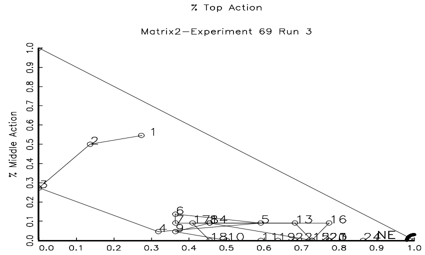


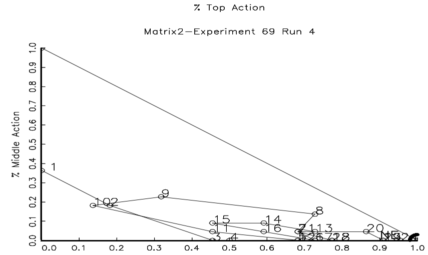




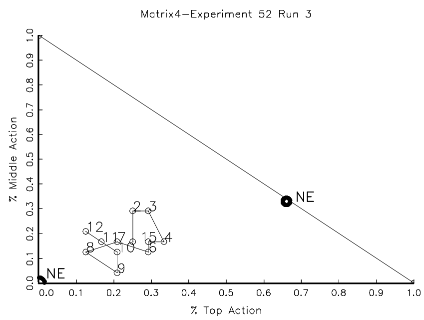


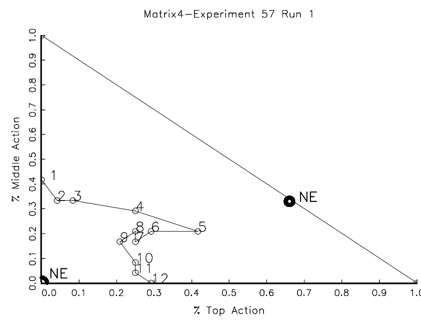


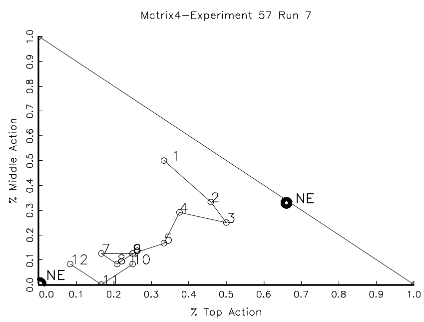


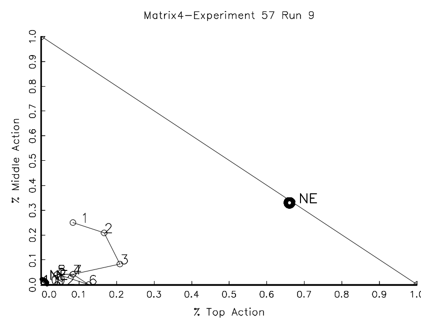


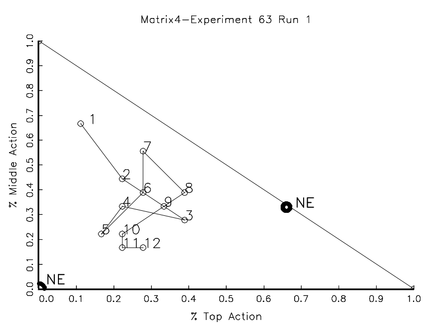
A.2 Matrix 4 -Single population

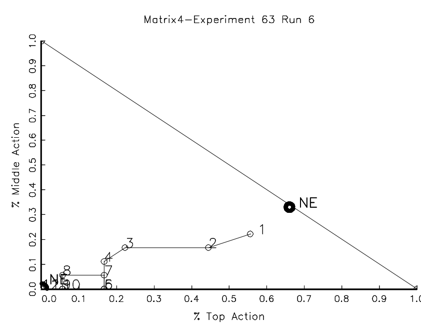


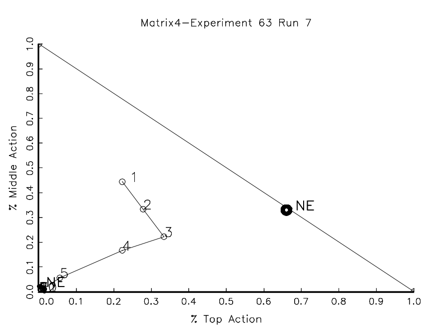


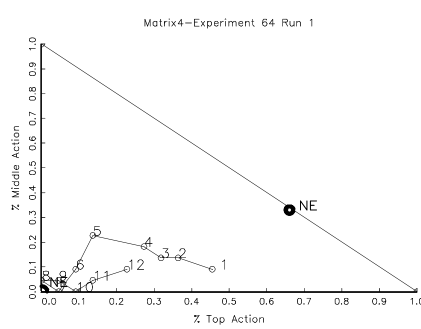




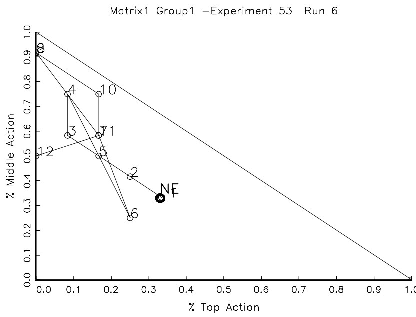


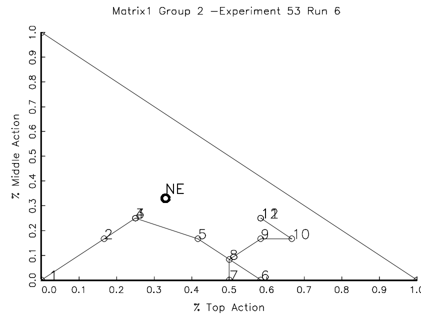


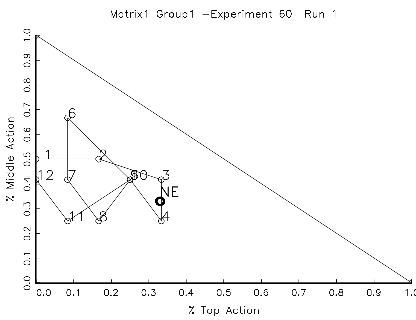


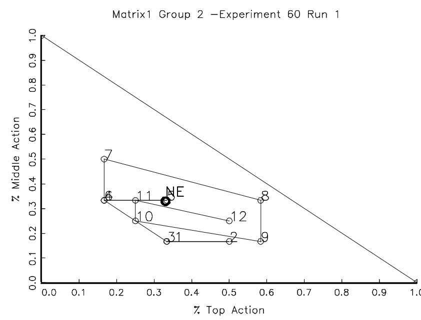


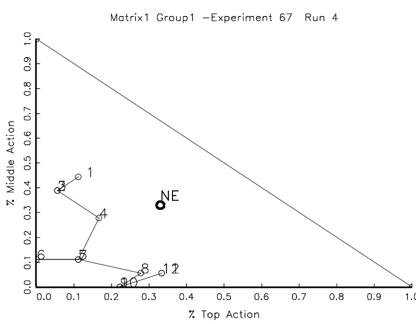
A.3 Matrix 1 -Two population

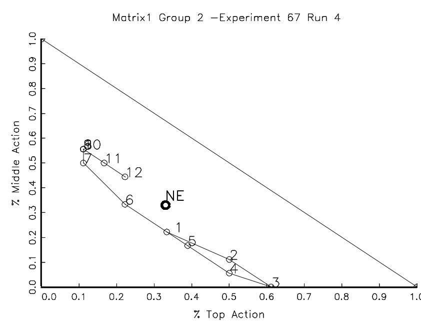


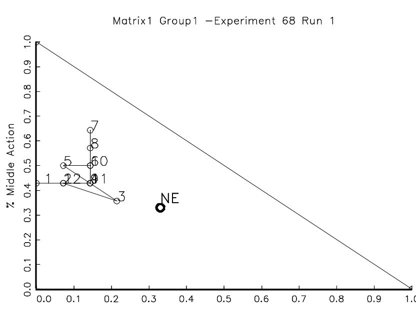


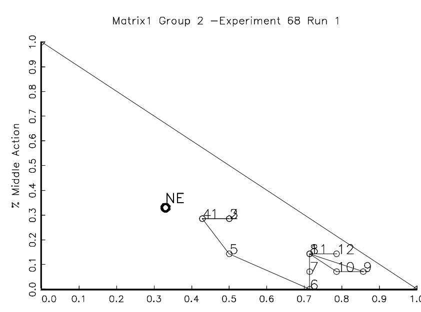




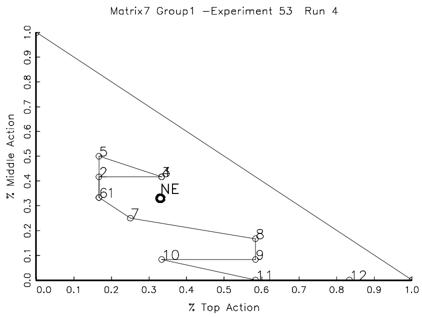


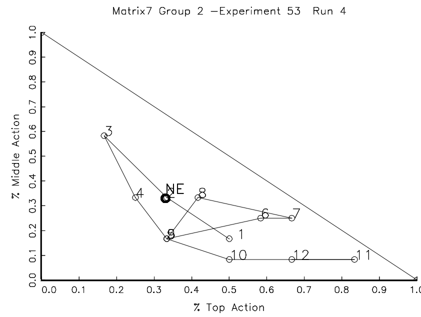


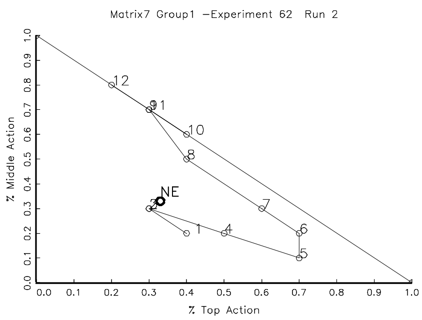


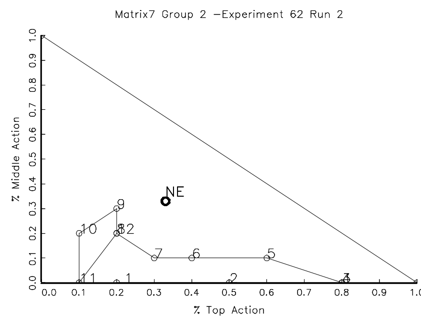


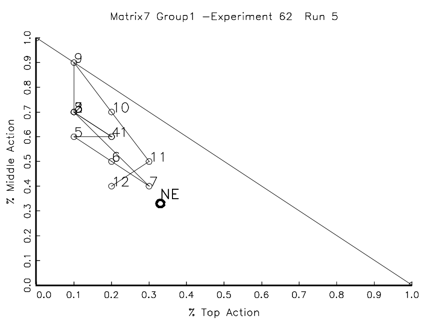
A.4 Matrix 7 -Two population

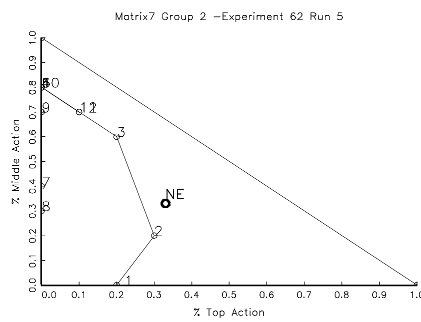


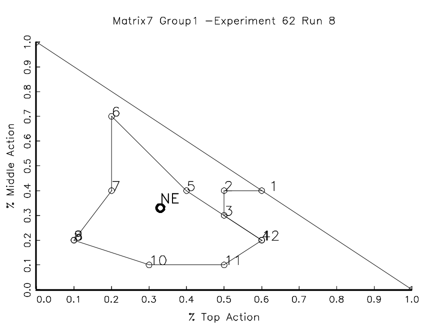


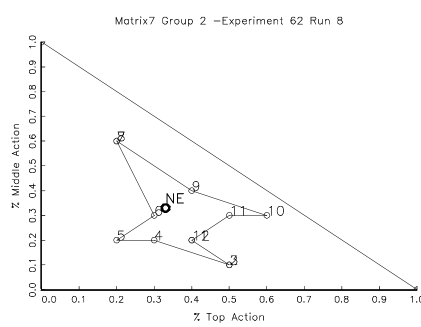




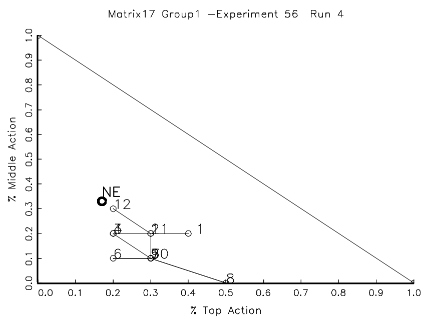


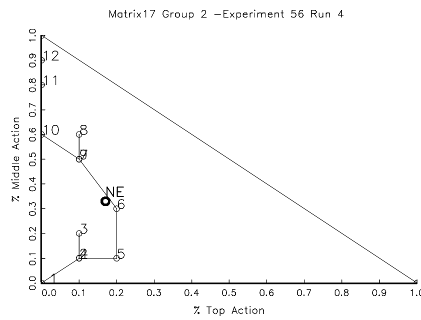


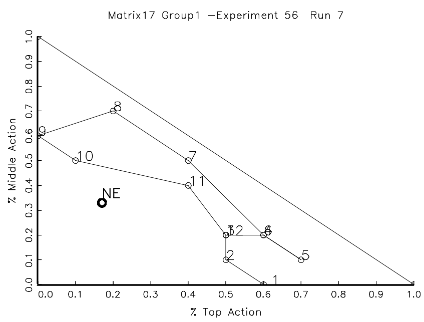


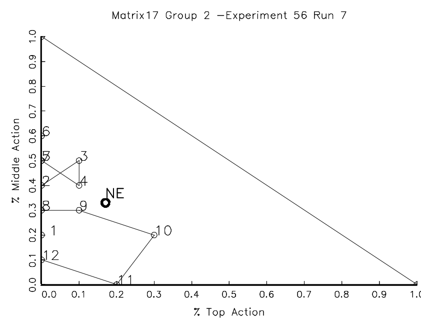


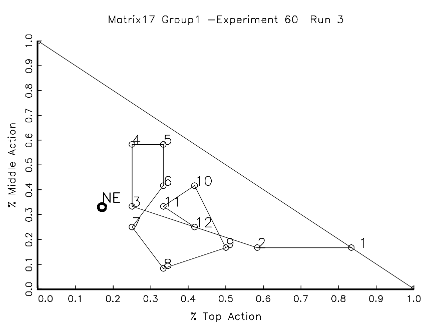
A.5 Matrix 17 -Two population

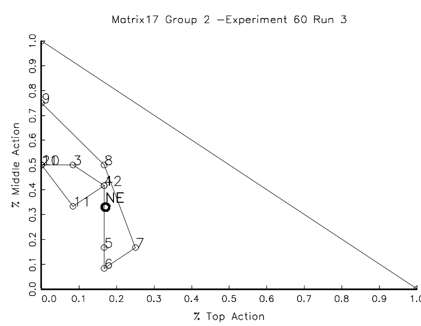


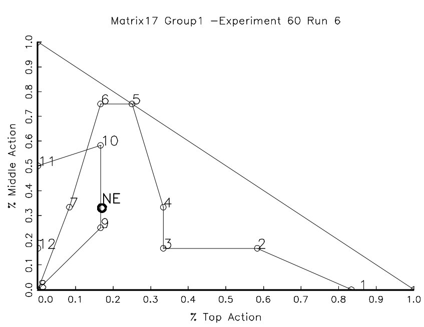


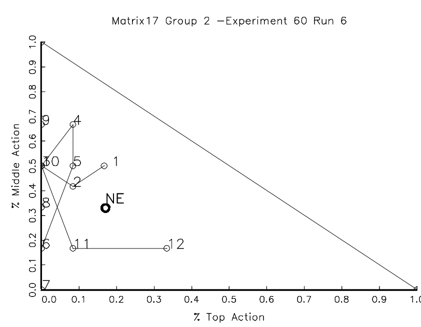












A.6 Matrix 22 -Two population