ECONOMIC RESEARCH REPORTS

An Experimental Study of Belief Learning Using Real Beliefs

By

Yaw Nyarko and Andrew Schotter

RR# 98-39 **November 1999**

C.V. STARR CENTER FOR APPLIED ECONOMICS

NEW YORK UNIVERSITY
FACULTY OF ARTS AND SCIENCE
DEPARTMENT OF ECONOMICS
WASHINGTON SQUARE
NEW YORK, NY 10003-6687

An Experimental Study of Belief Learning Using Real Beliefs

Yaw Nyarko and Andrew Schotter*

November 6, 1998

Abstract

This paper investigates belief learning. Unlike other investigators who have been forced to use observable proxies to approximate unobserved beliefs, we have, using a belief elicitation procedure (proper scoring rule), elicited subject beliefs directly. As a result we were able to perform a more direct test of the proposition that people behave in a manner consistent with belief learning. What we find is interesting. First to the extent that subjects tend to "belief learn" the beliefs they use are the stated beliefs we elicit from them and not the "empirical beliefs" posited by fictitious play or Cournot models. Second, we present evidence that the stated beliefs of our subjects differ dramatically, both quantitatively and qualitatively, from the type of empirical or historical beliefs usually used as proxies for them. Third, our belief elicitation procedures allow us to examine how far we can be led astray when we are forced to infer the value of parameters using observable proxies for variables previously thought to be unobservable. By transforming a heretofore unobservable into an observable we can see directly how parameter estimates change when this new information is introduced. Again, we demonstrate that such differences can be dramatic.

Key Words: Belief Learning, Game Theory, Experimental Economics

JEL Classification: D83, C91, C73

1 Introduction

In recent years game theorists and experimental economists have focused a great deal of attention on the question of how people learn when repeatedly playing a simple

*New York University - The authors would like to thank the C.V. Starr Center for Applied Economics for its financial support. In addition, we owe a great deal of thanks to Sangeeta Pratap for all her assistance as well as that of Gautam Barua and Allan Corns. Finally, we would like to thank the participants of the Cal Tech/UCLA Experimental Economics Workshop and the New York University Microeconomcis Seminar for their helpful suggestions.

matrix game. While some, e.g., Roth and Erev (1995) and Arthur (1991), focus on reinforcement learning in which people learn by looking back at their experience and seeing what has been successful in the past[1], others, Cheung and Friedman(1997), Boylan and El-Gamal (1993), Sopher and Mookherjee (1994, 1997)[2], Van Huyck, Battalio and Rankin (1997) and Fudenberg and Levine (1998) focus on belief learning and look to the past to update beliefs about their opponent's future action. Still others, Camerer and Ho (1997) select the best features of both of these models (among other things) in an approach that has proven to be remarkably successful.

In all of this research, however, there is an assumption that while past actions and payoffs are observable, beliefs are unobservable and therefore must be represented by proxies and inferred. For example, in the two most common belief learning models, the Cournot and fictitious play models, beliefs are either equivalent to the last period action of one's opponent, or an average of the previous actions of one's opponent. Some authors also use what we shall subsequently refer to as $\gamma - weighted$ empirical beliefs (or simply empirical beliefs). Here a weighted average of past actions is taken as a proxy for beliefs, where the weights decline geometrically at rate γ. (See Van Huyck et al.(1997), Cheung and Friedman (1997) and Fudenberg and Levine's (1998) model of smooth fictitious play). At various points in our paper below we will focus on fictitious play models because, given their widespread use, both experimentally and theoretically, they form a natural baseline from which to measure our results

This paper presents the results of a series of two-person constant-sum game experiments in which we directly elicited the beliefs of subjects using a "proper scoring rule" which provided subjects with an incentive to report their beliefs truthfully. We call these beliefs the subjects' "stated" beliefs. In addition, in two of the three experiments reported on here, we also allowed our subjects to choose mixed strategies. As a result, this paper presents, we think for the first time, an investigation of belief learning in which all relevant variables are observable; i.e., we study belief learning using real beliefs.[3,4,5] Our experimental results lead to two broad conclusions. First, under the assumption that people best-respond to some beliefs, we find that stated beliefs fit the data better than the other beliefs we study (Cournot, fictitious play and more generally, the class of $\gamma - weighted$ empirical distributions). Second, under

[1]Reinforcement learning is actualy an outgrowth of the psychology literature (see, Thorndike (1898) and Bush and Mosteller (1955)) whose main unifying theme is the "law of effect" which states that actions that have been successful in the past should be used more often on the future.

[2]Sopher and Mookhejee actually investigate both belief and reinforcement learning

[3]We would like to thank Jason Shachat for supplying us with his mixed stratetgy laboratory program.

[4]Shachat (1996) and Noussair (1997) allow for the use of mixed strategies, but neither allow for observable beliefs.

[5]Others have elicited beliefs in the study of public goods problems most notably Offerman (1997) and Offerman, Sonnemans, and Schram (1996). See also McKelvey and Page (1990). However, this paper presents an attempt to integrate this belief solicitation procedure into the study of belief learning, which, to our knowledge, has not been attempted before.

the assumption that people choose strategies via a logistic belief learning rule generated by some beliefs, we again find that stated beliefs fit the data the best. These conclusions validate our belief elicitation procedure.

Specifically, we ask and attempt to answer three questions:

Question 1: Are fictitious play beliefs (or, more generally, $\gamma - weighted$ empirical beliefs) a good proxy for stated beliefs.

Question 2: If subjects best respond, what is it that they best respond to? I.e., do they best respond to their stated or their empirical beliefs?

Question 3: If, as the experimental learning literature leads us to believe, subject behavior can best be described by a logistic belief learning model, which beliefs, when employed in such a model, provides the best fit for our data?

What we find is quite revealing.

First, we find little support for the idea that the process of forming fictitious play or empirical beliefs is descriptive of how subjects (or perhaps people in general) form their true or stated beliefs. Fictitious-play beliefs define a very stable time path while the stated beliefs of our subjects vary greatly from period to period. This leads us to question the validity of all belief-based models which rely on history alone to define beliefs, such as fictitious play or Cournot Best Response models.

Our study of Question 2 indicates that it is stated beliefs that they best respond to the most often. More specifically, in our Experiment 1 where only pure strategies are permitted for subjects, the strategy choices of subjects are consistent with best responses to their stated beliefs almost 75% of the time, while the comparable percentages for Cournot and Fictitious play beliefs don't exceed 50%. Finally, in studying Question 3, we use a logit model to predict choice behavior of individuals, and use this to compare three belief formation models — Cournot, Fictitious play, and Stated — in an effort to see which explains our data best. What we find is that the logit model using Stated beliefs does a far better job of explaining our choice data than do any of the other belief formation models we examined. Our results, through goodness of fit measures, tend to support the view that people do, in fact, follow a belief learning method of play. The only caveat, of course, is that they appear to use their heretofore unobserved stated beliefs to do so and not the fictitious play beliefs so often referred to in the literature or various other $\gamma - weighted$ empirical beliefs. It is this discovery that we feel provides one of the main lessons to be learned from this paper. Furthermore, because we are able in this work to measure beliefs directly and compare them to the types of empirical beliefs so frequently used in the literature, our experimental design provides a perfect setting within which to investigate how far off parameters estimates derived using only observable action data can be when compared to the those estimated using true or at least stated beliefs.

In this paper we will proceed as follows: In Section 2 we will explain the experiments performed and present our experimental design. In Sections 3 and 4 we will discuss our results, while in Section 5 we will discuss what we feel we have learned from these experiments and present some conclusions.

2 Experimental Design and Procedures

2.1 Experimental Design

Three different sets of experiments were run using the experimental laboratory of the C.V. Starr Center for Applied Economics at New York University during the Fall of 1997 and Spring of 1998. Subjects were recruited from undergraduate economics courses and reported to the lab for experiments that took between $1\frac{1}{2}$ and 2 hours. No subjects had any training in game theory. In these experiments subjects played a 2x2 game 60 times with the same opponent under various treatments. Payoffs were denominated in experimental dollars and converted into U.S. dollars at a rate of 1 pt. =$.05 for row choosers and 1 pt. = $ 0.05 for column choosers. Subjects, on average, earned approximately $15.00 for their participation which was paid to them at the end of the session. They were paid $3.00 simply for showing up.

The program used to run the experiments was generously supplied to us by Jason Shachat and the Experimental Science Lab of the University of Arizona. In using this program, subjects are able, if they wish, to actually choose mixed strategies by specifying the exact probability mixture to use in any given round.[6]

In the three experiments the identical 2x2 constant sum game was run under different informational and strategic conditions. In Experiment 1, as is true in most of the literature, subjects were constrained to only use pure strategies. In Experiments 2 and 3, subjects played this game with the ability to use both pure and mixed strategies. In both Experiment 1 and Experiment 2, after each round the subject was informed only of the action chosen by his or her opponent and the payoffs of each subject. They were not informed about the actual mixture used. In Experiment 3, at the end or each round subjects could view the actual mixture used by their opponent. Experiment 1 will be called the Pure strategy experiment, Experiment 2 will be called the Low information experiment, and Experiment 3 the High information experiment. The game used in Experiments 1-3 is presented below:

Game Played in Experiments 1-3:
Payoff Matrix

		Player 2	
		Red	*Green*
Player 1	*Red*	6, 2	3, 5
	Green	3, 5	5, 3

This game has many features we desired in our design. First, we wanted a game that was easy to understand, with an equilibrium that was not too difficult to either

[6] The instructions were computerized and are available upon request from the author.

Table 1: Experimental Design

	Feasible Strategies	Information	No. of Rounds	No.of Subjects	No. of Pairs
Experiment 1	pure	low	60	28	14
Experiment 2	mixed	low	60	28	14
Experiment 3	mixed	high	60	20	10

calculate or learn deductively. We wanted the equilibrium to be a mixed one, however, since we did not want equilibrium beliefs to be degenerate. These features were provided by a 2x2 constant sum game since a 2x2 game is as simple a game as one can find and the equilibria of such games are supported not only by the logic of best responses but also the entire weight of the mini-max theorem. Further, because our objective was to study learning, we wanted subjects to play against the same partner repeatedly but in a setting where the repeated game equilibrium prescription is unambiguous. This criterion was also met nicely by the 2x2 game we employed.

Finally, an important feature of the 2x2 game is that there are large portions of the unit interval, the domain of beliefs, over which the best response is constant. For example, in our experimental game whenever stated or empirical beliefs predict that Red will be chosen with a probability p $\in [0.4, 1]$,these beliefs prescribe the same best response for our subjects. Such a best response function stacks the deck against observing differences across our belief models so that if we do observe statistically significant differences our results are that much more persuasive.

The full experimental design is described in Table 1:

2.2 Eliciting Beliefs

Before subjects chose their mixed or pure strategies in any round, they were asked to write down on a worksheet a probability vector that they felt represented their beliefs or predictions about the likelihood that their opponent would use each of his of her pure strategies.[7]

In Experiment 1, where only pure strategies are allowed, we rewarded subjects for their beliefs as follows: First subjects report their beliefs by writing down a vector $\mathbf{r} = (r_{Red}, r_{Green})$ indicating their belief about the probability that the other subject will use the Red or the Green strategies[8]. Since in this experiment only one such strategy will actually be used, the payoff to player i when the Red strategy is chosen

[7]See Appendix 1 for the instructions concerning this part of the experiment.

[8]In the instructions m_j and r_{ij} are expressed as numbers in [0,100], so are divided by 100 to get probabilities.

by a subject's opponent and $\mathbf{r}$ is the reported belief vector of subject i will be:

$$\pi_{Red} = 0.10 - \frac{1}{20}\left\{(1 - r_{Red})^2 + (r_{Green})^2\right\}. \quad (1)$$

The payoff to subject i when the Green strategy is chosen is, analogously,

$$\pi_{Green} = 0.10 - \frac{1}{20}\left\{(1 - r_{Green})^2 + (r_{Red})^2\right\}. \quad (2)$$

The payoffs from the prediction task were all received at the end of the experiment.

Note what this function says. A subject starts out with \$0.10 and states a belief vector $\mathbf{r} = (r_{Red}, r_{Green})$. If their opponent chooses *Red*, then the subject would have been best off if he or she had put all of their probability weight on Red. The fact that he or she assigned it only r_{Red} means that he or she has made a mistake. To penalize this mistake we subtract $(1 - r_{Red})^2$ from the subject's \$0.10 endowment. Further, the subject is also penalized for the amount he or she allocated to the Green strategy, r_{Green} by subtracting $(r_{Green})^2$ from his or her \$0.10 endowment as well. (The same function applies symmetrically if Green is chosen). The worst possible guess, i.e. predicting a particular pure strategy only to have your opponent choose another, yields a payoff of 0 (and explains the normalization constant (1/20) which appears in the formula). It can easily be demonstrated that this reward function provides an incentive for subjects to reveal their true beliefs about the actions of their opponents. Telling the truth is optimal.

In Experiments 2-3, in order to induce subjects to report truthfully in any period t, we rewarded them for their predictions in that period using the following reward function:

$$\text{Prediction Payoff for Subject } i = 0.10 - \frac{1}{20}\Sigma_{j=1}^{2}(m_j - r_{ij})^2 \quad (3)$$

where, r_{ij} is the probability weight reported by subject i as to i's prediction about his opponent's use of strategy j, and m_j is the actual weight assigned to strategy j in the mixed strategy chosen by i's opponent.

As you can see, this function offers them a \$0.10 payment if they predict correctly and then subtracts an amount proportional to the squared distance of their stated belief vector from the actual mixed strategy used by their opponent in that period. Obviously, if a subject predicts correctly, the loss term would be zero and the subjects would keep all of his \$0.10 for that round. Again, the worst possible outcome is a payoff of 0.

As is true of all scoring functions, while payoffs are maximized by truthful revelation of beliefs, there are other beliefs that could be stated which are more secure in the sense that they guarantee a higher minimum payment. For example, reporting equal probability for each strategy would guarantee the largest minimal payment. [9]

[9] See Camerer (1995) and Allen (1987) for a discussion of this point.

If subjects were risk averse, such an action might be desirable. As can be seen in the data, there is little indication that such equi-probable vectors were used.

We made sure that the amount of money that could potentially be earned in the prediction part of the experiment was not large in comparison to the game being played. (In fact, the maximum earnings that could be earned in the prediction part of Experiments 1, 2 and 3 was only \$6.00 as opposed to the average payoffs in the game of \$15.00). The fear here was that if more money could be earned by predicting well rather than playing well, the experiment could be turned into a coordination game in which subjects would have an incentive to co-ordinate their strategy choices and play any particular mixed or pure strategy repeatedly so as to maximize their prediction payoffs at the expense of their game payoffs. Again, absolutely no evidence of such coordination exists in the data. In fact, we offer quite a bit of evidence that supports the view that the beliefs we elicited were truthful.

2.3 Defining Beliefs

Given any γ in $(-\infty, \infty)$, we define, using the notation of Cheung and Friedman (1997), player i's γ-weighted empirical beliefs (or, for simplicity, γ- empirical beliefs) to be the sequence defined by

$$b_{it+1}^{j} = \frac{1_t(a^j) + \Sigma_{u=1}^{t-1} \gamma_i^u 1_{t-u}(a^j)}{1 + \sum_{u=1}^{t-1} \gamma_i^u} \tag{4}$$

where b_{it+1}^{j} is player i's belief about the likelihood that the opponent will choose action a^j in period t+1, $1(a^j)$ is an indicator function equal to 1 if a^j was chosen in period t and 0 otherwise, and γ_i^u is the weight given to the observation of action a^j in period t-u. Fictitious play beliefs are those as above for the special case of $\gamma = 1$. For experiments 1 and 2, where players only observe the actions of their opponents (and not their mixed strategies) we define the Cournot beliefs to be those which assign probability one to opponent's previous period play. This is the special case of (4) for $\gamma = 0$. In experiment 3, the high information experiment where in each period a player observes the previous period mixed strategy of her opponent, we define the Cournot belief in that period to be the previous period mixed strategy of her opponent.

Since there are only two actions, we represent all beliefs in terms of the probability assigned to the action RED. Let BS_t and $b_t(\gamma)$ denote player i's date t stated beliefs and player i's date t $\gamma - \textit{weighted}$ empirical belief (where t$\in$ {1,...,T})(and, for emphasis, these are probabilities the player assigns to the event that the opponent will choose action RED). We define γ^* to be the value of γ which minimizes the distance between the stated beliefs and the $\gamma - \textit{weighted}$ empirical beliefs in a mean squared error sense. That is, γ^* is the value of γ which solves $\text{Min}_\gamma \sum_{t=1}^{T} \mid BS_t - b_t(\gamma) \mid^2$. A subject's γ^*- empirical belief is $b_t(\gamma^*)$.

3 Results

We will structure the discussion of the results of our experiment by answering a series of questions which originally motivated our research.

3.1 Question 1: Are empirical beliefs a good proxy for stated beliefs.

To demonstrate the relationship between stated $\gamma - weighted$ empirical and fictitious play beliefs we present Figures 1a-1f, 2a-2d and Table 2. Figures 1a-1c which presents three histograms of the distributions of absolute difference between the stated, $\gamma - weighted$ empirical, and fictitious play belief that the Red strategy will be chosen by one's opponent. These differences are presented, subject-by-subject for the first, second, and third 20 round segments of each experiment. That is, we divide the data set into three twenty round periods and for each period we present a histogram of the absolute differences between the beliefs subjects reported to us (their stated beliefs about the probability of Red being chosen) and the fictitious play beliefs we calculated and aggregate these differences in 20 round segments. This is done for each experiment. Table 2 presents a set of descriptive statistics about these histograms. In addition, we present in Figures 1d-1f the mean Euclidean distance between stated, $\gamma - weighted$ empirical, and fictitious play beliefs for each round. Finally, to give some insight as to how the two time series differed on the individual level, Figures 2a-2d present some representative belief time series graphs of four subjects taken from Experiment 2.

Figures 1a-1f, Figures 2a-2d, and Table 2 Here

Looking at Figures 2a-2d first, we see that while fictitious play beliefs soon become stable, stated beliefs are quite variable over the full horizon of the experiment. (These figures are more than typical of the population of subjects). Even Figure 2d, which presents an individual whose stated beliefs were relatively stable, the variability in stated beliefs is far greater than that of the fictitious play belief series it is compared to.

With respect to Figures 1a-1c, if there is a great deal of agreement between stated and fictitious play beliefs, then we would expect that the histogram of the absolute value of these differences would be concentrated around 0 with a small variance around that point and a mode close to 0 as well. If stated and fictitious play beliefs tended to differ, then most of the observations would be spread over then full support of the distribution and represent large positive or negative differences.

As we can see, there is little support for the hypothesis that the absolute value of the differences between subjects' stated and fictitious play beliefs is zero. To characterize these histograms we calculated the mean and median absolute difference as well as the inter-quartile range[10] of the distribution. These are presented in Table 2. In general, the mean absolute difference between stated and fictitious play beliefs of choosing Red varies from a low of 0.220 in rounds 21-40 of Experiment 2 to a high of 0.254 in rounds 21-40 of Experiment 1 with the median varying from a low of 0.195 rounds 21-40 of Experiment 2 to a high of 0.256 in rounds 1-20 of Experiment 3. The inter-quartile range of these distribution start from a low of 0.1554 in rounds 20-40 of Experiment 1 to a low of 0.2141 in round 1-20 of Experiment 1. The fact that the lower endpoints of the inter-quartile ranges tend to be substantially above zero indicates that in general stated and fictitious play beliefs differ.

To demonstrate that these differences do not change or decrease over time, we performed a set of Kolmogorov-Smirnov tests on the data to test whether the distribution of these absolute difference changes over time, i.e., whether the distribution of absolute differences is the same in the first as in the final 20 round period. What we find is that one can not reject the hypothesis that these distributions are identical within each experiment. In other words, the distribution of absolute differences within the first 20 rounds of any experiment is not significantly from that same distribution say in the last 20 rounds. [11]

Finally, Figures 1d-1f presents the round-by-round mean Euclidean distance between stated and fictitious play beliefs. As we can see, on average over the 60 rounds of the experiment the mean Euclidean distance between stated and fictitious play beliefs is 0.351, 0.323, and 0.338 in Experiments 1, 2, and 3 respectively.

Table 3 presents data on the γ^*-empirical beliefs for each subject.

[Table 3 here]

Note that these γ^*'s are clustered around 1 with a relatively small variance. This is interesting since it would, on the face of it, indicate that fictitious play beliefs are about as good as we can get as an approximation to stated beliefs using the belief formation model (3.1). This does not imply that the fit is very good, however, as is evidenced by the large sum of squares terms in the table. In fact, as we have seen in Figures 2a-2d, by choosing $\gamma's$ near 1 γ^* -empirical beliefs are, in many instances, attempting minimize the distance between empirical and stated beliefs by passing a relatively stable straight empirical belief series through the middle of a cycling stated

[10]The interquartile range is the interval between the first and third quartile of a distribution.

[11]In the results below, D is the calculated test statistic defined by the Kolmogorove-Smirnov test. Critical value for D at the 5% level is 8.

Stated vs empirical Experiment 1, rounds 1-20 vs rounds 40-60 D=7

Experiment 2, rounds 1-20 vs rounds 40-60 D=6

Experiment 3, rounds 1-20 vs rounds 40-60 D=4

Stated vs $\gamma*$- empirical

Experiment 1, rounds 1-20 vs rounds 40-60 D=7

Experiment 2, rounds 1-20 vs rounds 40-60 D=5

Experiment 3, rounds 1-20 vs rounds 40-60 D=4 beliefs series. With only one parameter, this may be the best we can do but that still may not be very good.

A pair-wise Kolmogorov-Smirnov test of these distributions indicate that there is no difference in the distribution of γ_i^* 's across these experiments at the 5% level.[12] This is interesting, especially when comparing Experiments 2 and 3 since, in Experiment 3 subjects are informed of their opponent's last period mixture after every round. This extra information does not seem to affect stated beliefs in such a way as to alter the fit between them and a subject's γ_i^*-empirical beliefs.

The entries in Table 2 and Figures 1a-1f and 2a-2d labeled γ^* - empirical beliefs replicate the calculations we have performed for fictitious-play beliefs using our now more sophisticated γ^* - empirical belief measure. As you can see, while there is a closer relationship between γ^* - empirical and stated beliefs than there was between fictitious play and stated beliefs, qualitatively all of the conclusions stated before carry through here. For example, while the mean absolute difference between fictitious play and stated beliefs over the 60 rounds of the experiments 1, 2, and 3 were 0.351, 0.323, and 0.338, respectively, they were 0.331, 0.298, and 0.331 using γ^*- empirical beliefs. The histograms of absolute differences (as summarized in Table 2) show the exact same features as those for empirical beliefs and the Kolmogorov-Smirnov test run to investigate whether there was a tendency for the differences between stated and γ^*- empirical beliefs to converge over time could also detect no significant difference between any two 20 round distributions in any experiment.

In short, as these descriptive statistics indicate, stated and fictitious play beliefs show a great tendency to differ within each of our three experiments and theses differences show no tendency to diminish as the experiment progresses over its 60 round horizon.

Even if fictitious play beliefs are a poor proxy for true or stated beliefs, however, it does not mean that fictitious play beliefs are not a useful model since operationally all that matters is that the two sets of beliefs prescribe the same best-response action at each (or most) point in time. In the 2x2 games used in our experiments this might be quite likely since, as we stated above, there are broad ranges of beliefs over which the same response action is prescribed so there is a great deal of room available for fictitious play and stated beliefs to differ and yet prescribe the same action. For example, in Experiments 1-3, any belief on the part of the row player that their

[12] D is the test statistic defined by the
Kolmogorov Smirnov test.
Experiment 1 vs Experiment 2
D= 5, $\chi^2 = 1.78$
Experiment 2 vs Experiment 3
D= 0.164, $\chi^2 = 1.26$.
Experiment 1 vs Experiment 3
D= 0.2, $\chi^2 = 1.87$.
Critical value for χ^2
is 5.99 at the 5% level.

opponents are likely to use the Red strategy with a probability greater than 0.40 will lead them to chose Red as a best response. For column players, just the opposite is true. Any belief that the row player will use Red with a probability grater than .4 will lead the column player to choose Green with probability 1. Hence, if both stated and fictitious play beliefs spend the majority of their time appropriate regions, then no mater how different they might be, they would be observationally equivalent with respect to prescribed actions.

This conjecture is easily tested on the individual level by taking the time series of best responses to fictitious play beliefs and comparing it to that predicted as best responses to the time series of stated beliefs. We do this by constructing a "counting" index defined as follows. In each round of each experiment there are N subjects. Each subject in each round has a stated belief and a fictitious play belief. In addition, if they are maximizers, they would have a best response to those beliefs which, except when they hold equilibrium beliefs, prescribes a pure strategy. From these N subjects count in each period the number of subjects whose prescribed best response under fictitious play beliefs is the same as that under their stated belief. Hence, if fictitious play and stated beliefs were strategically equivalent, they would prescribe the same actions in each period and we should observe all N subjects choosing the same action. If the beliefs always prescribed different best responses, our index should be zero. In particular, our index is a measure of how close the best-response prescriptions of the two time series of beliefs are.

In Figures 3a-3d we plot our index, the fraction of agreements between the best responses to these different beliefs, period by period for Experiments 1-3.

Figures 3a-3d here

As we can see, looking at the line describing the difference between prescribed best responses for fictitious play and stated beliefs, in all of these figures, (and the line describing the difference between stated and Cournot beliefs in Figure 3c) there is quite a bit of similarity between the prescribed best responses of all of our belief time series. On average in any period the stated and fictitious play beliefs prescribe the same behavior approximately 66% of the time in Experiments 1, 2, and 3. (The actual averages are 0.6464, 0.6642,0.6683, for the comparison of stated and fictitious play beliefs in Experiments 1, 2, and 3 and 0.6900 for the comparison of stated and Cournot beliefs in Experiment 3.) Whether you consider this to be a large or small amount of agreement depends on how you model the process of belief formation for individuals. If you were to think that subjects simply draw their beliefs uniformly over the interval [0,1], then the probability that the beliefs of two subjects playing each other in an experimental game will both simultaneously lie in the sub-intervals [0, 0.4] or [0.4,1] (and hence prescribe the same best response), would be 0.52, which does not appear greatly different than 0.66. However, given the fact that empirical beliefs define a very flat time series, such beliefs are likely to be in only one of these sub-intervals almost all of the time. Hence, if stated beliefs were to remain in the other sub-interval for the majority of their lifetime, we would expect relatively little agreement in best responses. In such a situation 0.66 would seem large.

As we can see in Figures 3a-3d, there is also no tendency for this difference to disappear as time goes on so that there does not appear to be much learning over time.

Finally, as Figures 3a -3c show, using γ^*-empirical beliefs does not change the correspondence between the prescribed best response to stated and γ^*-empirical beliefs. On average they correspond about 66% of the time within each experiment. In conclusion, we have attempted to answer the question of whether, despite the observed difference between stated and fictitious play (or γ^*- empirical) beliefs, it is possible that they still prescribe basically the same best responses. What we find is a remarkably strong correspondence between these prescribed best responses (about 66% in the case of both empirical and γ^*- empirical beliefs). It is important to point, however, that this strong correspondence is merely a correspondence in the prescriptions of a theory which may or may not be revealed in the actual behavior of subjects. For example, it would fail completely if subjects did not behave so as to best-respond to their beliefs. These considerations prompt our third question.

3.2 Question 2: If subjects best respond, what is it that they best respond to?

Up to this point we have spoken very little about subject behavior, i.e., their actions. As a first cut at the data one might refute the best response hypothesis immediately since in Experiments 2 and 3, where mixed strategies are available, subjects consistently use them despite the fact that for almost all beliefs on the part of the their opponent (in particular for all beliefs other than the equilibrium beliefs) they have a pure strategy best response. More precisely, mixed strategies are used 75.8% and 80.3% of the time in Experiments 2 and 3 respectively, despite the fact that subjects hold equilibrium beliefs only 9.3% and 7.4% of the time in these experiments.

To investigate the best response behavior of subjects let us perform the following exercise. Given fictitious play, Stated, and Cournot beliefs we can predict, for any individual and any time during the experiment, what his or her best response should be to each of these. Hence, we can count the number of times the strategy choices of our subjects (either in Experiment 1 where pure strategies were chosen or in Experiments 2 and 3 where mixed strategies were used) were consistent with the best responses dictated by these different beliefs. When the chosen strategy of the subjects is consistent with two or even three beliefs (or none) we count them separately.

The results of this exercise are presented in Table 4 which present calculations of Experiments 1, 2 and 3, respectively.

[**Table 4 here**]

In this table we have placed Cournot, fictitious play and Stated beliefs along the first three rows and columns. (We could have added other $\gamma - weighted$ empirical beliefs to this table as well, but for ease of exposition we have decided to limit our comparisons to the most familiar empirical models of belief formation – Fictitious Play and Cournot). Any cell in the matrix presents the number of times that the actions of subjects were consistent with the best response suggested by one of these beliefs notions either alone or in conjunction with other beliefs. For example, along the diagonal of the initial 3x3 matrix, i.e. the cells (CC), (HH) and (SS) we present the number (fraction) of times that the strategy chosen was consistent with that prescribed by one and only one belief notion. Hence in the (CC) cell of the table relating to Experiment 2 we see that in that low information experiment there were 23 instances in which the observed behavior of subjects was consistent with the best response dictated by Cournot beliefs while in the HH cell there were 33 instances where behavior was consistent with only best responses to fictitious play beliefs. The off diagonal entries, such as CS, indicate when observed behavior was consistent simultaneously with the best-response prescriptions of two belief notions (in this case Cournot and Stated). If a mixed strategy was chosen, then that strategy would correspond to a best response to none of the beliefs (except when the subject held equilibrium beliefs). It would also be possible for a pure strategy to correspond to a best response to none of our three beliefs if, for example, the subject chose Green when all beliefs simultaneously indicated that Red would be best.

There are several things to note. First notice that in Experiments 2 and 3 there are a rather large number of cases where people do not best respond to any beliefs. This was mostly due to the fact that subjects used mixed strategies so often. Next, note that when they do best respond, they are much more likely to best respond to their stated beliefs, either in isolation or jointly with some other belief. For example, in the pure strategy experiment, Experiment 1, subjects best responded to their stated beliefs 800 (47.6%) times while best responding to their fictitious play and Cournot beliefs only 359 and 462 times respectively. It is rather remarkable, in fact, that while they best respond to their stated beliefs alone in Experiment 1 302 times, they do so with respect to their Cournot and fictitious play beliefs only 92 and 67 times respectively. A similar pattern occurs in Experiment 2 which is the low-information mixed-strategy experiment. There they best respond to their stated beliefs alone 216 times while they do so to their Cournot and fictitious play beliefs only 23 and 33 times respectively. In Experiment 3, the high-information mixed-strategy experiment, these differences are not as dramatic. Here they only best respond to their stated beliefs alone 88 times but do so to their Cournot beliefs 54 times. This is not so strange since it is in this experiment that they are given information about their opponent's previous period mixed strategy and hence are able to best respond to them if they wish.

In conclusion, it would appear that Stated beliefs are far more likely than Cournot or fictitious play beliefs to be the beliefs that subjects best respond to. This result, to some extent, tends to validate our beliefs elicitation procedure since it would appear that the beliefs we had subjects report to us were ones that they acted upon when money was on the line in the experiment. This finding, we feel, is important if such scoring rules are to be used in the future in experiments.

3.3 Question 3: If subject behavior can best be described by a logistic belief learning model, which beliefs provide the best fit for our data?

Our question here differs from that asked in question 2 since there we were interested only in best response choices and predictions while here we are interested in which of the beliefs we have, when employed in an appropriate discrete choice model of behavior, best explains the choices of our subjects. In such a model, the best response function is a continuous function of beliefs and prescribes a probability with which a subject should choose a given pure strategy rather than, as is true in deterministic fictitious play, having a point of discontinuity at which pure strategy prescriptions change. We will actually consider the model where, at time period t, the probability that any subject, i, chooses the Red strategy (in a 2x2 game with a Red and a Green Strategy available) is a function of the expected payoff difference between these two strategies. To calculate such expected payoffs we must use some set of beliefs and in our experiments we have at our disposal a number of different ones to choose between. After we have settled on the beliefs we expect to use, we must choose some form for the behavior rule.

In our analysis below we will use the frequently employed logistic function presented as:

$$\begin{aligned}\text{Probability of Red in period t} &= \frac{e^{\beta_0+\beta_1(E(\pi_t^d))}}{1+e^{\beta_0+\beta_1(E(\pi_t^d)}}, \\ \text{Probability of Green in period t} &= 1-\frac{e^{\beta_0+\beta_1(E(\pi_t^d))}}{1+e^{\beta_0+\beta_1(E(\pi_t^d)}}\end{aligned} \tag{5}$$

where $E(\pi_t^d)$, is the expected payoff difference to be derived from using the Red strategy instead of the Green strategy in period t given the beliefs that the subjects holds at that time, and β_0 and β_1 are constants to be estimated. When fictitious play beliefs are used to compute the expected payoff differences in this function, we obtain what Fudenberg and Levine (1998) call "smooth fictitious play".

There are two different interpretations of the logistic rule in (5), with subtle differences in the implications for the behavior of players. The first interpretation is that of a random utility model, henceforth the R.U.M. In that interpretation, the payoff accruing to a strategy, say Red, is composed of the deterministic part given by the expected stage game payoffs as in previous sections, plus a random term ε_{Red}

(assumed to be i.i.d. over time). The expected payoff to Red, given this player's beliefs about her opponent, is therefore equal to $\mathrm{E}\pi_{Red}$ plus the random term. In any round subjects act after seeing the realizations of ε_{Red} and ε_{Green} for that round. The Red strategy is therefore chosen whenever $E\pi_{Red} + \varepsilon_{Red} > E\pi_{Green} + \varepsilon_{Green}$. Hence, generically, the player will choose a pure strategy at each date. It is however assumed that these shocks are not observed by the experimenter. The ex ante probability, or the probability assessed by the experimenter, that the Red strategy is chosen is then $\Pr[(\varepsilon_{Red} - \varepsilon_{Green}) > (E\pi_{Green} - E\pi_{Red})]$, i.e. the probability that the difference in the random shocks attached to the utility of each strategy is greater than the difference of the expected utility of these strategies given the subject's beliefs. Given this formulation, strategies with higher expected payoffs are given higher probabilities of being chosen, yet the probability of choosing any such strategy depends upon how much greater its expected utility is from that of other strategies. When the distribution of the disturbances is of the extreme value type[13], these probabilities result in the logistic behavioral rule (5).

The second interpretation is based on the work of Luce (1959). Luce studied some axioms of choice and showed that they implied that subjects choose actions probabilistically. The probability each action is chosen is some function of the expected payoff to that action. By appropriately choosing that function we obtain the specification in (5). We shall call this interpretation the Luce Model. Under the Luce Model players will generically be choosing mixed strategies at every date.

If players are constrained to choose only pure strategies (as in Experiment 1) there is no way of distinguishing the R.U.M from the Luce model. On the other hand, if players are able to choose mixed and pure strategies (as in Experiments 2 and 3), then one would expect pure strategies to be used in the R.U.M and mixed strategies to be used in the Luce model. This provides a method of distinguishing the two models. Since in practically all rounds of experiments 2 and 3 players choose mixed strategies, this would suggest that the R.U.M interpretation of behavior fails.

Most of the learning literature constrains players to choose only pure strategies. The literature then estimates a R.U.M using maximum likelihood techniques. To compare our results to those of the literature we also perform this exercise. In particular, we will ignore mixed strategies (and the fact that, contrary to the R.U.M, they actually do mix). We will then use the actions realized from their mixed strategies, and assume that these were chosen pure. We then perform the maximum likelihood exercises used in the literature.

We estimate five such models each run on individual outcome observation data generated by our three experiments. These models are estimated on the individual level as well as on the aggregate level using pooled data. These differ only according to the belief formation process we posit for the subjects. In model 1, we use the stated

[13]For a more complete dervation of such a logistic function see McKelvey and Palfry (1995) where they derive the existience of a quantal-response equilibrium in which subjects play probabilistically as described here but where the beliefs defined are equlibrium or self-fulfilling beliefs.

beliefs of subjects to calculate expected payoffs while in model 2 we use fictitious play beliefs. In model 3 we estimate what we will call the $\hat{\gamma}$- empirical beliefs model where the $\gamma's$ themselves are estimated using maximum likelihood techniques simultaneously with β_0 and β_1. Model 4 uses Cournot beliefs. Finally, in model 5 we use our γ^*-empirical beliefs (as defined in section 2.3) as our belief proxy.

Next, in Model 6 we estimate the Luce model. We use the observed mixed strategy as our left hand or dependent variable. Instead of using maximum likelihood techniques to estimate β_0 and β_1,we find the β_0 and $\beta_1$that determines (using stated beliefs) a best fit, in the sum-of-squares sense, between the observed mixtures and the predictions of these probabilities using the logit function. Call these estimates of β_0 and β_1 our Luce or **"minimum sum of square estimates"** to indicate that they are derived to minimize a sum of squared errors and not to maximize a likelihood function. In particular, if m_t denotes the date t mixed strategy probability of Red, we find parameters β_0 and β_1 to solve the problem

$$Min_{\beta_0,\beta_1} \sum_{t=1}^{T} [m_t - \frac{e^{\beta_0+\beta_1(E(\pi_t^d))}}{1+e^{\beta_0+\beta_1(E(\pi_t^d)}}]^2 \tag{6}$$

where T is the total number of periods in the experiment. This exercise is equivalent to estimating the parameters β_0 and β_1 through non-linear least squares regression $m_t = G(\beta_0, \beta_1, E(\pi_t^d))+ \nu_t$ where $G(\beta_0, \beta_1, E(\pi_t^d))=\frac{e^{\beta_0+\beta_1(E(\pi_t^d))}}{1+e^{\beta_0+\beta_1(E(\pi_t^d)}}$ and ν_t is a white noise term with the standard properties.

All of these models, 1 - 6, were estimated individual by individual. In addition, we have estimated a set of aggregate regressions, one for each experiment, using the same specification along with dummies to represent the fixed effects present across individuals. Table 5 presents the estimates of our aggregate logit models and our Luce or minimum sum of squares model. In these tables we present the number of observations, the estimated β_0, $\beta_1$coefficients, (in model 3 the maximum likelihood estimates of γ are also presented) along with the standard errors of the estimates and their significance levels for each model and each experiment. In addition we present for each model the maximized likelihood.

[Table 5 here]

As we can see, there are some similarities across each of these experiments. First, in all regressions the β_1 coefficient was positive and significant at at least the 5% level. Obviously, we expected the positive sign since the model is predicated on the notion that strategies expected to yield higher payoffs should be used more often. The constant term varied both in sign and significance across models and experiments. For example, in Experiment 1, where pure strategies are used, the constant term was positive and significant in four out of the five models (all except the stated belief model), while in the high information experiment it was negative and insignificant (at the 5% level) in four out of the five experiments. More interesting, however, are the estimates of the γ parameter (i.e., the $\hat{\gamma}'s$). Here, we have statistically significant (5%) estimates for γ in the low information and pure strategy experiments (Experiments 1 and 2) of 0.6672 and 0.6098, respectively, while in the high information experiment, Experiment 3, γ was insignificantly different from zero. This is interesting because when, in our belief formation rule (4), γ takes on a value of 0, beliefs equal opponent's previous period pure strategy. It is our suspicion that because the high information experiment focused subjects attention to the previous round by giving them the mixed strategy information about their opponent, the previous outcome of their opponent might have become focal and led them to give it excessive importance.

Finally, at the more micro level, it is interesting to note how different the $\gamma's$ estimated in our individual model 3 regressions are from those calculated earlier when we defined our γ^* - empirical beliefs. These $\gamma's$ are presented in tabular form in Table 6 and graphically in Figures 4a, 4b, and 4c.

[**Table 6 and Figure 4a-4c here**]

Looking at Figures 4a-4c notice how dramatic the difference between the estimated $\gamma's$ of Model 3 and our calculated $\gamma^{*\prime}s$ is. For example, in all experiments every γ which was calculated from our γ^*-empirical series is greater than their counterparts estimated in model 3. A Wilcox two-tailed test indicates that in each experiment these distributions are different at the 1% percent level.[14] Further, while the γ^*-empirical estimates are centered around 1, those estimated from Model 3 tend to be centered around zero with nine of the 28 being negative.[15]

We consider this comparison important since it demonstrates exactly how far off parameter estimates can be when we attempt to use maximum likelihood techniques on data constructed from observable proxies for unobservable data (as most economic data is). More precisely, standard empirical analysis as conducted by economists is most like our Model 3 where γ is estimated using discrete (0-1) data using historical proxies for unobserved variables. Because we are able to observe both beliefs and true mixed strategies, we can calculate γ directly by finding that γ which best fits our stated belief series (our γ^*-empirical. beliefs). Hence, this paper offers a controlled experiment enabling us to measure how far off economists and policy makers may be

[14]Experiment 1, Pure Strategies
T = 0, z = -4.622, p(z) < .00005
Experiment 2, Low Information
T = 0, z = -3.919, p(z) < .00005
Experiment 3, High Information
T = 5, z = -4.508, p(z) < .00005

T is the test statistic of the Wilcoxon test. z is a transformation of T with a standard normal distribution.

[15]These results are strikingly simmilar to those of Cheung and Friedman (1997) in their estimates of γ.

when they are forced to use historical proxies for unobservable variables. Because these differences are so dramatic in our work here, we take these results as a warning urging us to be careful when we too quickly accept parameter estimates estimated in that manner.

4 Model Selection among Logistic Models

We now select among models 1 -6 of section 3.4 in terms of their goodness of fit. We will proceed in two ways. First, because we have data on the actual mixed strategies used by each subject in Experiments 2 and 3, we are able to use this data to construct direct measure of the goodness of fit of each of our models. More precisely, for these experiments we can simply take each individual's mixture in each round and compare it to the probabilities predicted by each of our logit regression models to see which fits the mixture data best. We can also, of course, compare the goodness of fit of these models by performing a set of model selection tests on our aggregate regressions which we will do shortly.

To explain our first approach more completely, consider the following: For each individual and for each of our logit models (i.e. Stated, fictitious play, $\hat{\gamma}$- empirical, Cournot, γ^* - empirical), we have an estimated[16] β_0 and β_1 coefficient. Hence, for any round if we were to plug in one of our belief measures in to the logit equation, we would get a predicted probability of Red for that round. Calling the difference between the predicted probability of Red in round t and the true probability of Red that the subject used in his mixed strategy the model's prediction error for that round, we can use the average prediction error round by round in each model as a goodness of fit measure.

More precisely, assume there are N people in an experiment and let p_{it}^j be person i's round t predicted probability of playing the Red strategy using logit model j, and m_{it} be the observed mixed strategy probability that subject i actually used for Red. The goodness of fit measure for model j in round t is then equal to the mean prediction error averaged across all N subjects in round t or:

$$MPE_t^j = \Sigma_{i=1}^N \mid p_{it}^j - m_{it} \mid /N.$$

Obviously, at any round t when there is a perfect match between the predicted and observed probabilities this measure will be 0 (i.e. the error will be minimized). It will take on a value of 1 at time t when the predicted and each observed belief differ by their maximum amount. Tables 7 and 8 present these calculations for Experiments 2 and 3 respectively using the Minimum Sum of Squares, Stated, fictitious play, Cournot, and $\hat{\gamma}$ - empirical and γ^* - empirical models while Figures 5a-5b present them graphically.

[Table 7 and 8 Figures 5a and 5b here]

[16] $\hat{\gamma}$ is estimated jointly with β_0 and β_1.

As can be seen from Tables 7 - 8 and Figures 5a-5b, in both the high and low information experiments, Experiments 2 and 3, the stated belief model fits the data better than any of our other models (except the minimum square error model). For example, in comparing the mean prediction errors round by round between the stated and empirical belief models, we see that the stated belief model has a smaller mean prediction error in 58 of the 60 rounds in both the Low and High Information Experiments. It outperforms the Cournot model in 58 of 60 rounds in the High Information Experiment as well and 55 of the 60 rounds in the Low Information Experiment.

On average, however, the model which predicts mixed strategies the best is the Minimum Sum of Squares model using stated beliefs (mean $\text{MPE}_t^i = 0.1720$ in the low and 0.2273 in the high-information experiments) which is not surprising since its objective is to minimize an objective function not very different from our goodness of fit measure. The model using stated beliefs, Model 1, comes in a close second, however (mean 0.2045 in the low and 0.2379 in the high information experiment) and outperforms the model using either fictitious play (mean 0.2632 in the low and 0.2940 in the high information experiment), Cournot (mean 0.2529 in the low and 0.2928 in the high information experiment)) or the γ-empirical model (mean 0.2544 in the low and 0.2931 in the high information experiment) beliefs. The fact that the Cournot model outperforms both the fictitious play and the $\hat{\gamma}-$ empirical model may be due to the fact that subjects may be somewhat myopic in their period to period reactions and hence events in the more distant past may have little if any influence on their actions today. Presumably, since the Cournot model is nested in the $\gamma-$ empirical model (i.e. $\gamma = 0$) we should expect that just the opposite might be true. Obviously for our goodness of fit metric this presumption was not born out.

A series of pair-wise Mann-Whitney U-tests performed on individual data from both experiments to test whether the sample of MPE_i^j's (i. e., the mean prediction error for individual i averaged over all 60 periods of the experiment) for the stated belief model in Tables 7- 8 come from populations with the same means as the other models strongly rejects that hypothesis in favor of the alternative that our goodness of fit measure was higher in the stated belief model at the 5% or less level for each comparison. [17] In summary, it appears as if having to rely on empirical proxies (like

[17] Experiment 2: Low Information
Stated vs. γ- empirical: U=264, z = -2.097
Stated vs. empirical: U=256, z = -2.228
Stated vs. Cournot: U = 268, z = -2.031
Stated vs.γ*- empirical: U=262, z = -2.130
Empirical vs. γ*- empirical: U=384, z = -0.131
Empirical vs. γ- empirical: U=368, z = -0.393
Gamma- Empirical vs. γ*- empirical: U=410, z = -0.294
Experiment 3: High Information
Stated vs. γ- empirical: U=137, z = -1.704
Stated vs. empirical: U=130, z = -1.893
Stated vs. Cournot: U = 126, z = -2.463
Stated vs. γ*- empirical:U=149, z = -1.379

fictitious play) to infer the hitherto unobservable stated beliefs can lead to results which, according to our metric, greatly inhibit the accuracy of our learning models.

A more conventional approach toward model selection would be to run a set of maximum likelihood ratio tests on our aggregate regressions to test if, pair-wise, any of these models fit the data better than any others. Because our models are not nested in a parametric sense, however, we can not employ the classical Maximum Likelihood Ratio Tests. Rather, we use a test of Vuong (1989) for non-nested models. As Vuong (1989) demonstrates, for any two such models, f and g, with maximized log likelihoods $\log \pounds_f$ and $\log \pounds_g$ and n observations, the test statistic

$$T = \frac{\frac{1}{\sqrt{n}}[\sum \log \pounds_f - \sum \log \pounds_g - k(f,g)]}{\frac{1}{\sqrt{n}}[\sum(\log \pounds_f - \log \pounds_g)^2]}$$

under the null hypothesis that models f and g are identical, tends asymptotically in distribution, to a standard normal random variable N(0,1). $k(f,g) = (\frac{p}{2}\log n - \frac{q}{2}\log n)$,where p is the number of parameters in model f and q is the number of parameters in model g, is a correction factor for models with different numbers of parameters. [18]The results of these tests are presented in Table 9 below:

Table 9 here

In this table each entry presents the test statistic (asymptotically a standard normal random variable (see Vuong (1989)) used to test the null hypothesis that there is no difference in the goodness of fit between any two of our five models. For example, in Table 9 the entry in the M1-M2 cell indicates the results of the pair-wise test of the hypothesis that there was no difference between the goodness of fit of the Stated Belief (M1) and Fictitious play (M2) models. Test statistic values between -1.96 and +1.96 would indicate failure to reject at the 5% level while values greater than 1.96 would indicate that model M1 fits the data better than M2. A value less than -1.96 indicates just the opposite, M2 provides a better fit than M1.

Table 9 confirms, on the aggregate level, exactly the conclusion reached above based on our individual calculations. In every experiment the Stated belief model, Model 1, outperforms all other models and does so significantly at at least the 5% level. In addition, except for the comparison between the Cournot Model (M3) and the γ^*- empirical model (M4) in Experiment 3 (where the Cournot model provides a better fit to the data), none of the other models distinguish themselves from each other in a statistically significant manner. This result once again affirms our claim

Empirical vs. γ*- empirical: U=168, z = -0.865
Empirical vs. γ- empirical: U=211, z = -0.297
Gamma- Empirical vs. γ*- empirical: U=178, z = -0.595

[18]We run these tests on the aggregate regressions since we need to make binary comparisons in these tests and this would not be feasible on individual regressions since there are 76 of them in total.

that if belief learning is to provide a good guide to the behavior of laboratory subjects, one is going to have to be careful and get them to reveal their true beliefs. Using empirical proxies can lead one astray.

To reconfirm this finding using a more disaggregated approach which uses the information provided by the likelihoods estimated on an individual by individual basis, we compared the individual maximized likelihoods derived from the individual regressions run across models within any experiment to see whether one model fits the data better on an individual-by-individual basis.

More precisely, consider the sample of maximized likelihoods generated by our stated belief model, Model 1 in say Experiment 2, as compared to the sample of such maximized likelihoods generated by our empirical belief model, in that experiment. There are 28 such likelihoods, one for each subject in the experiment. In addition, each model generates one such sample, so within the low information experiment there are five such samples. The question we ask is whether within an experiment these samples, when compared pair-wise, came from populations with the same mean.

A Wilcoxon Matched-Pairs Signed-Ranks test performed to test this hypothesis rejects that hypothesis at the 5% level of significance for all comparisons between stated and $\hat{\gamma}$- empirical, fictitious play and Cournot beliefs in both Experiment 2 and 3. The exception is the comparison between stated and $\hat{\gamma}$ - empirical beliefs in Experiment 2 where the null can not be rejected.[19] In other words, on an individual-by-individual basis, using the sample of individual maximized likelihoods, our logit models fit the data best when stated beliefs are used as arguments in the logit function instead of any other types of beliefs. If people are basing behavior in our experiments on beliefs, then they are basing on stated beliefs.

Finally, it is interesting to note that our γ^* - empirical model outperforms the $\hat{\gamma}$ - empirical model using these matched-pairs tests. This is important because it indicates that not only are the estimates we get from our γ^* calculation different

[19] **Low Information Experiment (Experiment 2).**
T is the test statistic for the Wilcoxon test while z is a transformation of T with a standard normal distribution.
Stated vs γ empirical: T = 151, z=-1.184
Stated vs empirical: T=101, z=-2.322
Stated vs Cournot: T= 110, z= -2.117
Stated vs $\gamma*$ -Empirical: T = 0, z = 4.622
Empirical vs γ - Empirical , T = 66, z = -3.119
Empirical vs. $\gamma*$ - Empirical, T = 0, z = -4.6226
γ Empirical vs. $\gamma*$ - Empirical, T = 0, z = -4.622
High Information Experiment (Experiment 3)
Stated vs γ empirical: T = 58, z =-1.754
Stated vs empirical: T=41, z =-2.389
Stated vs Cournot: T= 39, z = -2.463
Stated vs $\gamma*$ -Empirical: T = 0, z = -3.919
Empirical vs γ - Empirical , T = 63, z = -1.567
Empirical vs. $\gamma*$ - Empirical, T = 0, z = -3.919
γ - Empirical vs. $\gamma*$ - Empirical, T = 0, z = -3.919

from the parameter estimates we get for γ in Model 3, but on an individual-by-individual basis, the logit model using γ^* - empirical beliefs actually outperforms the model which is based on inferring γ using only observable outcome data.

5 Conclusions

This paper has investigated belief learning. Unlike other investigators who have been forced to use observable proxies to approximate unobserved beliefs, we have, using a belief elicitation procedure (proper scoring rule), elicited subject beliefs directly. As a result we were able to perform a more direct test of the proposition that people behave in a manner consistent with belief learning. What we find is interesting.

First to the extent that subjects tend to "belief learn" the beliefs they use are the stated beliefs we elicit from them and not the "empirical beliefs" posited by fictitious play or Cournot models. Hence, while we present data that lends support to the notion that people behave in a manner consistent with belief learning, we must be careful to specify the type of beliefs that must be used as inputs to these models.

Second, we present evidence that the stated beliefs of our subjects differ dramatically, both quantitatively and qualitatively, from the type of empirical or historical beliefs usually used as proxies for them. While empirical beliefs, i.e. those beliefs formed by counting the frequency with which subjects have used their various strategies in the past, tend to generate a fairly stable time series, stated beliefs vary wildly from period to period and exhibit no tendency to settle down as the experiment progresses. Still, such differences would be inconsequential if they had no impact on behavior, i.e., if despite their apparent difference both stated and empirical beliefs prescribed the same behavior. As we have seen, such is not the case.

Third, our belief elicitation procedures allow us to examine how far we can be led astray when we are forced to infer the value of parameters using observable proxies for variables previously thought to be unobservable. By transforming a heretofore unobservable into an observable we can see directly how parameter estimates change when this new information is introduced. Again, we demonstrate that such differences can be dramatic.

Finally, in the future we hope to use our elicited beliefs to investigate whether subjects, when playing this game, are capable of achieving an "equilibrium in beliefs". In such an equilibrium each subject would believe that the other would use his or her strategies with the frequencies prescribed by the Nash (mini-max) equilibrium of the game. While in such an equilibrium beliefs should remain stable over time, the actions of the agents need not resemble those prescribed by the static equilibrium of the game. In other words, while others have investigated the predictive ability of the Nash theory by looking at the history of the observable actions, we feel that we can verify the theory independently using stated beliefs. Such an attempt might prove to discover equilibria where they previously had been thought not to exist.

References

[1] Allen, F. "Discovering Personal Probabilities When Utility Functions are Unknown", **Management Science,** vol. 33, 1987, pp. 542-544.

[2] Arthur, B., "Designing Economic Agents That Act Like Human Agents: A Behavioral Approach to Bounded Rationality", **AER Papers and Proceedings,** Vol. 81, May 1991, pp. 353-359.

[3] Boylan, R. and El-Gamal, M., "Fictitious Play: A Statistical Study of Multiple Economic Experiments", **Games and Economic Behavior,** vol. 5, 1993, pp. 205-222.

[4] Bush, R and Mosteller, F., **Stochastic Models of Learning,** New York, John Wiley and Sons**, 1955.**

[5] Camerer, C. "Individual Decision Making", in **The Handbook of Experimental Economics,** A. Roth and J. Kagel, eds., Princeton New Jersey, Princeton University Press, 1995.

[6] Camerer, C. and H. T. Ho, "Experience -weighted Attraction Learning in Games: A Unifying Approach", mimeo, California Institute of Technology, 1996.

[7] Camerer, C. and Ho, T. H., "Experience -weighted Attraction Learning in Games: A Unifying Approach", mimeo, California Institute of Technology, 1996.

[8] Cheung, Y. W., Friedman, D., "Individual Learning in Normal Form Games: Some Laboratory Results", **Games and Economic Behavior**, vol., 19, 1997, pp. 46-76.

[9] Fudenberg, D. and Levine, D., Theory of Learning in Games, Cambridge MA. MIT Press. 1998.

[10] Luce, R. D., **Individual Choice Behavior: A Theoretical Analysis,** New York, John Wiley & Sons, 1959.

[11] McKelvey, R. and Palfry, T., "Quantal Response Equilibrium for Normal Form Games", **Games and Economic Behavior,** Vol. 10, no.1, July 1995, pp. 6-38.

[12] McKelvey, R. and Page, T., "Public and Private Information :An Experimental Study of Information Pooling", **Econometrica,** vol. 58, 1990, pp. 1321-1339.

[13] Noussair, C. and Faith, T., "A Laboratory Study of Mixed Strategy Play", mimeo, Krannert School of Management, Purdue University, 1997.

[14] Offerman, T., **Beliefs and Decision Rules in Public Goods : Theory and Experiments,** Kluwer Academic Publishers, The Netherlands, 1997

[15] Offerman, T., Sonnemans, J., and Schram, A., "Value Orientations, Expectations and Voluntary Contributions in Public Goods", **Economic Journal,** July, vol. 106, 1996, pp. 817-845.

[16] Shachat, J., "Mixed Strategy Play and the Minimax Hypothesis", UCSD Economics Discussion Paper 96-37, University of California at San Diego, November 1996.

[17] Mookerherjee, D. and Sopher, B., "Learning Behavior in Experimental Matching Pennies", **Games and Economic Behavior**, vol., 7, 1994, pp. 62-91.

[18] Mookerherjee, D. and Sopher, B., "Learning and Decision Costs in Experimental Constant- Sum Games",**Games and Economic Behavior**, vol., 19, 1997, pp.97-132.

[19] Thorndike, E.L., "Animal Intelligence: An Experimental Study of the Associative Processes in Animals", Psychological Monographs 2, 1898.

[20] VanHuyck, J. Battalio, J., and Rankin, F., "Strategic Similarity and Emergent Conventions: Evidence From Scrambled Payoff Perturbed Stag-Hunt Games", Mimeo, Department of Economics, Texas A & M University, July 1997.

Table 2: Absolute Difference in Beliefs

Experiment 1: Pure Strategy						
	Stated vs. Fictitious Play Beliefs			Stated vs. γ^*- Empirical Beliefs		
Rounds	Mean	Median	Inter-Quantile Range	Mean	Median	Inter-Quantile Range
1-20	0.250	0.244	0.2141-0.2851	0.243	0.234	0.2024-0.2797
20-40	0.254	0.246	0.1554-0.3513	0.236	0.240	0.1515-0.3166
40-60	0.242	0.237	0.1604-0.3013	0.223	0.228	0.1490-0.2692
Experiment 2: Low Information						
	Stated vs. Fictitious Play Beliefs			Stated vs. γ^*- Empirical Beliefs		
Rounds	Mean	Median	Inter-Quantile Range	Mean	Median	Inter-Quantile Range
1-20	0.240	0.230	0.2131-0.2849	0.227	0.221	0.1952-0.2628
20-40	0.220	0.195	0.1560-0.2846	0.195	0.167	0.1281-0.2709
40-60	0.225	0.209	0.1458-0.2771	0.209	0.198	0.1383-0.2657
Experiment 3: High Information						
	Stated vs. Fictitious Play Beliefs			Stated vs. γ^*-Empirical Beliefs		
Rounds	Mean	Median	Inter-Quantile Range	Mean	Median	Inter-Quantile Range
1-20	0.2382	0.2566	0.1893-0.2747	0.2242	0.2335	0.1912-0.2730
20-40	0.2286	0.2331	0.2106-0.2656	0.2374	0.2110	0.2223-0.2800
40-60	0.2522	0.2458	0.2062-0.2944	0.2425	0.2327	0.2046-0.2802

Table 3: Calculated γ^*

Experiment 1 (Pure Strategy)			Experiment 2 (Low Information)			Experiment 3 (High Information)		
Player	γ^*	Min SSQ	Player	γ^*	Min SSQ	Player	γ^*	Min SSQ
1	0.751	3.676	1	1.168	6.392	1	0.961	7.617
2	1.034	1.925	2	0.925	3.748	2	1.531	2.898
3	0.873	11.066	3	0.919	4.704	3	0.942	6.262
4	0.972	4.447	4	1.825	0.552	4	0.948	3.496
5	0.948	9.992	5	0.768	2.411	5	0.774	2.737
6	1.926	8.376	6	0.935	1.674	6	1.740	5.465
7	1.238	4.071	7	1.356	3.438	7	0.927	4.879
8	1.066	3.919	8	1.337	1.906	8	0.893	4.065
9	0.994	3.286	9	1.012	5.176	9	1.037	3.875
10	2.754	4.181	10	1.487	3.054	10	1.037	3.875
11	1.124	11.738	11	1.100	1.997	11	1.057	5.179
12	1.430	4.485	12	1.661	2.245	12	1.064	2.806
13	1.009	1.976	13	1.327	0.444	13	1.039	3.514
14	1.085	1.856	14	1.019	7.889	14	2.493	4.372
15	1.946	5.125	15	1.044	4.809	15	0.960	6.749
16	1.152	6.553	16	1.040	4.395	16	1.139	4.060
17	0.556	11.450	17	1.637	12.617	17	1.932	1.411
18	1.012	4.201	18	0.924	2.387	18	0.873	4.690
19	1.012	0.653	19	1.109	4.532	19	1.860	5.908
20	1.029	2.342	20	0.882	1.246	20	1.031	7.290
21	1.367	0.884	21	1.097	3.056			
22	1.400	6.434	22	0.951	3.718			
23	1.114	1.295	23	0.936	9.931			
24	0.933	3.656	24	0.863	12.486			
25	0.981	4.319	25	0.844	3.546			
26	0.854	3.987	26	1.657	4.699			
27	1.402	10.978	27	2.086	6.685			
28	1.085	8.793	28	0.958	5.066			

Table 4: Correspondence Between Actions and Best Response Prescriptions

	Cournot	Fictitious Play	Stated	Total	None	All
Experiment 1: Pure Strategy						
Cournot	92	132	238	462	117	472
Fictitious Play	132	67	260	459		
Stated	238	260	302	800		
Experiment 2: Low Information						
	Cournot	Fictitious Play	Stated	Total	None	All
Cournot	23	38	47	108	1134	121
Fictitious Play	38	33	68	139		
Stated	47	68	216	331		
Experiment 3: High Information						
	Cournot	Fictitious Play	Stated	Total	None	All
Cournot	54	19	32	105	829	99
Fictitious Play	19	48	31	98		
Stated	32	31	88	151		

Table 5: Regression Results

Model	β_0	β_1	Std. Error (Prob): β_0	Std. Error (Prob): β_1	$\hat{\gamma}$	Std. Error (Prob)	Obs	Mean Log Likelihood
Pure Strategy Experiment (Experiment 1)								
Model 1	0.0753	0.5672	0.0610	0.0388	NA	NA	1680	−0.6154
			(0.1084)	(0.0000)				
Model 2	0.1049	0.3000	0.0522	0.0605	NA	NA	1680	−0.6841
			(0.0222)	(0.0000)				
Model 3	0.0892	0.2017	0.0498	0.0516	0.6098	0.1547	1680	−0.6831
			(0.0367)	(0.0000)		(0.0000)		
Model 4	0.0943	0.0912	0.0520	0.0199	NA	NA	1680	−0.6854
			(0.0348)	(0.0000)				
Model 5	0.0967	0.2686	0.0492	0.0544	NA	NA	1680	−0.6844
			(0.0326)	(0.0000)				
Low Information Experiment (Experiment 2)								
Model	β_0	β_1	Std. Error (Prob):β_0	Std. Error (Prob):β_1	$\hat{\gamma}$	Std. Error (Prob)	Obs	Mean Log Likelihood
Model 1	0.0494	0.4412	0.0435	0.0403	NA	NA	1680	-0.6514
			(0.1280)	(0.0000)				
Model 2	0.0469	0.2897	0.0606	0.0594	NA	NA	1680	-0.6832
			(0.2196)	(0.0000)				
Model 3	0.0682	0.2628	0.0514	0.0576	0.6672	0.1242	1680	-0.6775
			(0.0919)	(0.0000)		(0.0000)		
Model 4	0.1110	0.1056	0.0471	0.0201	NA	NA	1680	-0.6823
			(0.0093)	(0.0000)				
Model 5	0.0429	0.3328	0.0732	0.0613	NA	NA	1680	-0.6812
			(0.2790)	(0.0000)				
High Information Experiment (Experiment 3)								
Model	β_0	β_1	Std. Error (Prob): β_0	Std. Error (Prob): β_1	$\hat{\gamma}$	Std. Error (Prob)	Obs	Mean Log Likelihood
Model 1	-0.0431	0.2362	0.0596	0.0411	NA	NA	1200	-0.0678
			(0.2345)	(0.0000)				
Model 2	-0.0426	0.1571	0.0523	0.0594	NA	NA	1200	-0.6902
			(0.2074)	(0.0041)				
Model 3	-0.0348	0.0834	0.0484	0.0233	0.0207	0.0896	1200	-0.6877
			(0.2361)	(0.0002)		(0.0000)		
Model 4	-0.0275	0.0536	0.0599	0.0336	NA	NA	1200	-0.6919
			(0.3231)	(0.0557)				
Model 5	-0.0388	0.1413	0.0498	0.0647	NA	NA	1200	-0.6911
			(0.2177)	(0.0145)				

Model 1: Stated Beliefs
Model 2: Fictitious Play Beliefs
Model 3: $\hat{\gamma}$-Empirical Beliefs
Model 4: Cournot Beliefs
Model 5: γ^*-Empirical Beliefs

Table 6: Calculated γ^* and Estimated $\hat{\gamma}$

Experiment 1 (Pure Strategy)			Experiment 2 (Low Information)			Experiment 3 (High Information)		
Player	γ^*	$\hat{\gamma}$	Player	γ^*	$\hat{\gamma}$	Player	γ^*	$\hat{\gamma}$
1	0.751	0.847	1	1.168	0.469	1	0.961	0.608
2	1.034	-1.063	2	0.925	0.559	2	1.531	-0.220
3	0.873	-0.479	3	0.919	0.115	3	0.942	-0.587
4	0.972	-0.171	4	1.825	-0.801	4	0.948	-0.581
5	0.948	-0.902	5	0.768	-0.235	5	0.774	0.494
6	1.926	0.358	6	0.935	0.405	6	1.740	-0.219
7	1.238	0.959	7	1.356	0.724	7	0.927	0.193
8	1.066	-0.546	8	1.337	-0.884	8	0.893	0.476
9	0.994	0.404	9	1.012	0.145	9	1.037	-0.128
10	2.754	-0.519	10	1.487	0.170	10	1.037	-0.128
11	1.124	-0.803	11	1.100	0.395	11	1.057	-0.117
12	1.430	-0.846	12	1.661	-0.422	12	1.064	0.902
13	1.009	0.019	13	1.327	0.153	13	1.039	-0.197
14	1.085	0.068	14	1.019	0.269	14	2.493	0.946
15	1.946	-0.282	15	1.044	208.698	15	0.960	-0.494
16	1.152	-0.140	16	1.040	0.350	16	1.139	-0.163
17	0.556	-0.274	17	1.637	-0.388	17	1.932	-0.980
18	1.012	0.998	18	0.924	0.803	18	0.873	0.814
19	1.012	0.400	19	1.109	0.289	19	1.860	0.982
20	1.029	-0.365	20	0.882	-0.124	20	1.031	-0.175
21	1.367	1.468	21	1.097	0.442			
22	1.400	0.232	22	0.951	0.378			
23	1.114	0.679	23	0.936	-0.054			
24	0.933	0.040	24	0.863	0.582			
25	0.981	0.392	25	0.844	0.168			
26	0.854	0.445	26	1.657	-0.668			
27	1.402	-0.530	27	2.086	0.884			
28	1.085	-0.686	28	0.958	-0.212			

Table 7: Individual Goodness of Fit: Low Information

Player	Min SSQ	Stated Beliefs	$\hat{\gamma}$ Empirical Beliefs	Fictitious Play Beliefs	Cournot Beliefs	γ^* Empirical Beliefs
1	0.07116	0.29417	0.29823	0.29810	0.29495	0.29843
	0.10620	0.22695	0.11629	0.06626	0.10642	0.06692
2	0.11653	0.54876	0.40170	0.37630	0.36382	0.38208
	0.13084	0.21318	0.17141	0.13953	0.16548	0.14123
3	0.15691	0.17302	0.27201	0.28187	0.27269	0.28123
	0.14701	0.13792	0.15550	0.16299	0.15434	0.16019
4	0.06281	0.12886	0.17134	0.12477	0.12838	0.14314
	0.04364	0.06175	0.11339	0.06266	0.06956	0.08455
5	0.19451	0.20033	0.24046	0.24659	0.24123	0.23468
	0.16100	0.16838	0.15506	0.15802	0.15585	0.15243
6	0.11721	0.11881	0.13682	0.18480	0.15912	0.17529
	0.11528	0.12553	0.13278	0.11563	0.13098	0.12730
7	0.10386	0.11724	0.12672	0.11535	0.12034	0.10787
	0.12240	0.11306	0.11999	0.10639	0.11049	0.10076
8	0.09243	0.09178	0.14173	0.15289	0.13102	0.15423
	0.07901	0.08040	0.12552	0.09544	0.08165	0.08925
9	0.26997	0.27937	0.27657	0.28014	0.27653	0.28031
	0.14323	0.16339	0.15280	0.16185	0.15246	0.16106
10	0.15228	0.15441	0.20029	0.18203	0.20282	0.16230
	0.11246	0.11213	0.13556	0.15054	0.13522	0.14096
11	0.17266	0.18381	0.17357	0.17417	0.17336	0.17397
	0.12360	0.16563	0.12957	0.13177	0.13007	0.13084
12	0.14008	0.14013	0.14687	0.14804	0.14567	0.15217
	0.10907	0.10972	0.11444	0.11289	0.11086	0.11523
13	0.07663	0.10449	0.10128	0.11738	0.10137	0.09370
	0.11499	0.10736	0.10822	0.12239	0.10724	0.12221
14	0.40055	0.39882	0.41575	0.42288	0.41633	0.42410
	0.18753	0.19136	0.17281	0.16846	0.17210	0.16654
15	0.10141	0.13851	0.28232	0.28348	0.27958	0.28502
	0.13036	0.14521	0.10409	0.09407	0.08130	0.09627

Note: Standard Deviations below

Player	Min SSQ	Stated Beliefs	$\hat{\gamma}$ Empirical Beliefs	Fictitious Play Beliefs	Cournot Beliefs	γ^* Empirical Beliefs
16	0.21444	0.21792	0.22327	0.23187	0.20658	0.23327
	0.10995	0.10647	0.15995	0.13684	0.16109	0.13691
17	0.11920	0.13639	0.38342	0.44890	0.39805	0.45146
	0.22356	0.21538	0.21312	0.15323	0.20285	0.14970
18	0.11013	0.12823	0.41098	0.43593	0.45342	0.41652
	0.22610	0.21723	0.17838	0.14007	0.13022	0.16939
19	0.15395	0.15455	0.17220	0.15104	0.16546	0.15621
	0.09742	0.10506	0.11108	0.10329	0.10911	0.10323
20	0.09627	0.10454	0.20692	0.26919	0.20039	0.24116
	0.09046	0.11477	0.16139	0.10920	0.15861	0.10526
21	0.37736	0.37324	0.31878	0.37992	0.34667	0.38366
	0.17039	0.18615	0.17763	0.15814	0.17419	0.16354
22	0.10880	0.10788	0.12945	0.10860	0.14129	0.11358
	0.12940	0.13811	0.12771	0.13126	0.13172	0.13489
23	0.22600	0.23227	0.38038	0.42271	0.37976	0.41049
	0.22543	0.21940	0.18446	0.15695	0.18354	0.17066
24	0.23075	0.23314	0.23696	0.23056	0.23709	0.23365
	0.14906	0.16066	0.15897	0.15023	0.15797	0.15390
25	0.28478	0.28662	0.26454	0.29871	0.26216	0.27751
	0.17454	0.17437	0.15765	0.17599	0.16121	0.16650
26	0.29993	0.29300	0.27654	0.29770	0.27084	0.29087
	0.14362	0.15935	0.18076	0.16449	0.16468	0.16898
27	0.28514	0.28551	0.33550	0.34631	0.30738	0.31446
	0.18560	0.18971	0.19147	0.18403	0.17711	0.18915
28	0.06891	0.08371	0.40304	0.35874	0.40537	0.36156
	0.14599	0.14547	0.18482	0.22778	0.18224	0.22681
Average	0.17160	0.20391	0.25456	0.26318	0.25292	0.25832
	0.09364	0.10954	0.09850	0.10709	0.10195	0.10617

Note: Standard Deviations below

Table 8: Individual Goodness of Fit: High Information

Player	Min SSQ	Stated Beliefs	$\hat{\gamma}$Empirical Beliefs	Fictitious Play Beliefs	Cournot Beliefs	γ^* Empirical Beliefs
1	0.31841	0.31827	0.33591	0.36671	0.36028	0.35671
	0.16434	0.15733	0.16185	0.15285	0.16143	0.15387
2	0.12269	0.13583	0.29931	0.29300	0.29350	0.29073
	0.18949	0.19583	0.19509	0.19428	0.19967	0.18930
3	0.24280	0.23294	0.24580	0.23005	0.23130	0.25093
	0.12476	0.12395	0.16647	0.19013	0.16686	0.18374
4	0.36815	0.35378	0.38578	0.34699	0.34925	0.29891
	0.15767	0.16632	0.16010	0.16889	0.16999	0.16010
5	0.35517	0.33808	0.34368	0.33989	0.34488	0.23172
	0.26637	0.24726	0.26131	0.24522	0.24628	0.16264
6	0.26498	0.25734	0.25625	0.25215	0.24816	0.25124
	0.16876	0.16255	0.16744	0.17180	0.16683	0.14584
7	0.27255	0.26928	0.25313	0.26220	0.29984	0.21101
	0.19490	0.19189	0.18569	0.18245	0.19612	0.17723
8	0.38006	0.41193	0.39991	0.40180	0.40199	0.24746
	0.15793	0.16386	0.15355	0.15391	0.16225	0.16423
9	0.19872	0.19932	0.22185	0.22194	0.22299	0.22267
	0.20674	0.22118	0.19695	0.18981	0.18899	0.17762
10	0.26178	0.28566	0.28301	0.28895	0.29781	0.28659
	0.14021	0.14757	0.13417	0.13521	0.13207	0.13249
11	0.01458	0.11288	0.25587	0.23720	0.14417	0.19926
	0.15866	0.14973	0.15103	0.17150	0.18881	0.16431
12	0.33110	0.33072	0.32646	0.34645	0.34658	0.34508
	0.01731	0.01688	0.15480	0.16929	0.16130	0.16684
13	0.26666	0.27152	0.25938	0.27975	0.28365	0.31432
	0.14725	0.16383	0.16018	0.14392	0.14464	0.14322
14	0.03369	0.05535	0.28136	0.28861	0.28406	0.28836
	0.19080	0.19471	0.17933	0.18128	0.19348	0.18080
15	0.27709	0.27689	0.26231	0.27180	0.28282	0.26994
	0.01896	0.04523	0.16552	0.15936	0.15792	0.15886

Note: Standard Deviations below

Player	Min SSQ	Stated Beliefs	$\hat{\gamma}$Empirical Beliefs	Fictitious Play Beliefs	Cournot Beliefs	γ^* Empirical Beliefs
16	0.17749	0.18705	0.21048	0.20026	0.21786	0.19926
	0.12185	0.15777	0.13086	0.16866	0.13887	0.13850
17	0.27481	0.29470	0.33949	0.34206	0.34997	0.33714
	0.17768	0.17806	0.17812	0.18963	0.19452	0.18228
18	0.19589	0.18865	0.30969	0.31461	0.29943	0.33598
	0.18306	0.16874	0.14654	0.14287	0.14032	0.14605
19	0.03575	0.07466	0.28397	0.28519	0.29432	0.28254
	0.15410	0.16719	0.18618	0.18246	0.18415	0.17813
20	0.06406	0.08121	0.59625	0.57260	0.61017	0.56933
	0.03423	0.03955	0.18535	0.18855	0.18765	0.18920
Average	0.22282	0.23380	0.30749	0.30711	0.30815	0.28946
	0.11486	0.10133	0.08479	0.08156	0.09270	0.08158

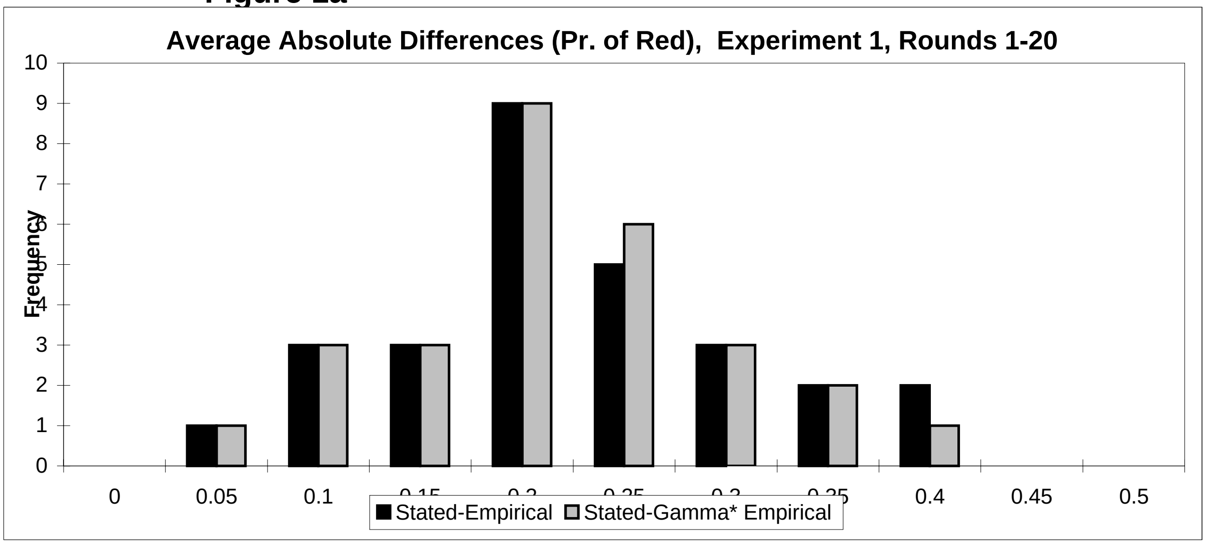
Note: Standard Deviations Below

Table 9: Model Selection Tests

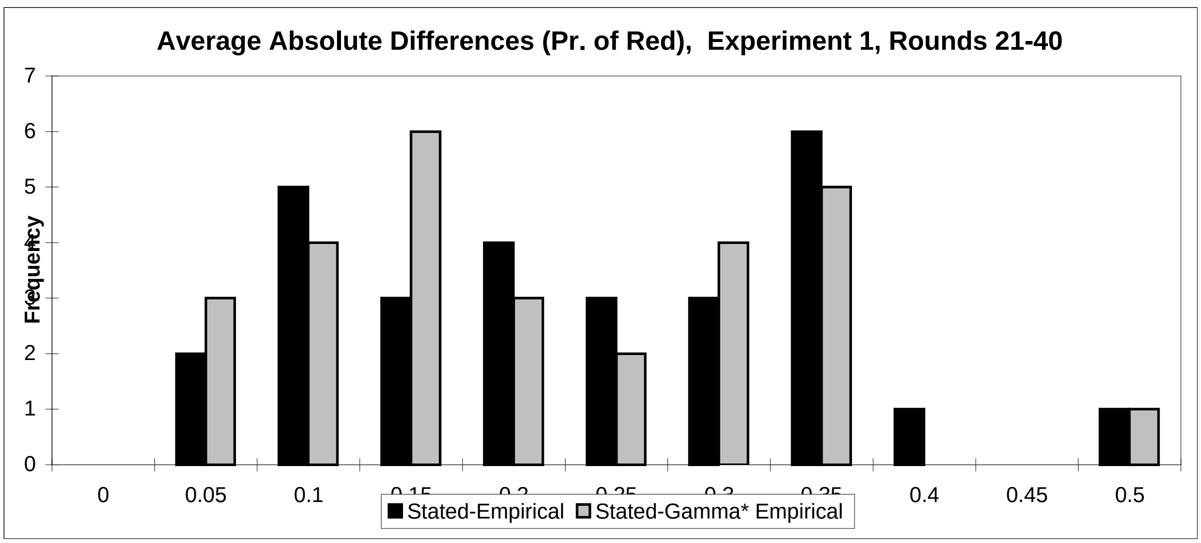
	Model 1	Model 2	Model 3	Model 4	Model 5
Low Information					
	Model 1	Model 2	Model 3	Model 4	Model 5
Model 1		4.704	3.909	4.200	4.368
Model 2			-1.004	-0.279	-1.108
Model 3				0.682	0.243
Model 4					-0.269
High Information					
	Model 1	Model 2	Model 3	Model 4	Model 5
Model 1		2.400	2.370	2.880	2.640
Model 2			0.325	0.960	0.600
Model 3				0.340	0.015
Model 4					-5.520
Pure Strategy					
	Model 1	Model 2	Model 3	Model 4	Model 5
Model 1		7.258	7.253	7.056	7.560
Model 2			0.422	0.269	0.185
Model 3				-0.107	-0.282
Model 4					-0.178

Model 1: Model Using Stated Beliefs
Model 2: Model Using Fictitious Play Beliefs
Model 3: Model Using $\hat{\gamma}$-Empirical Beliefs
Model 4: Model Using Cournot Beliefs
Model 5: Model Using γ^*-Empirical Beliefs

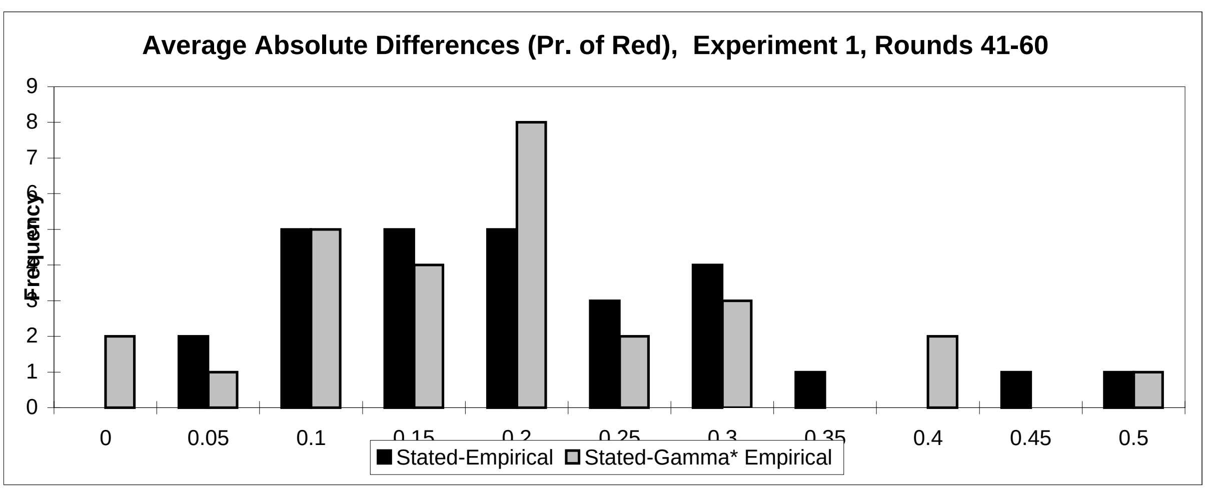
Figure 1a



Average Absolute Differences (Pr. of Red), Experiment 1, Rounds 1-20
Frequency
Stated-Empirical
Stated-Gamma* Empirical

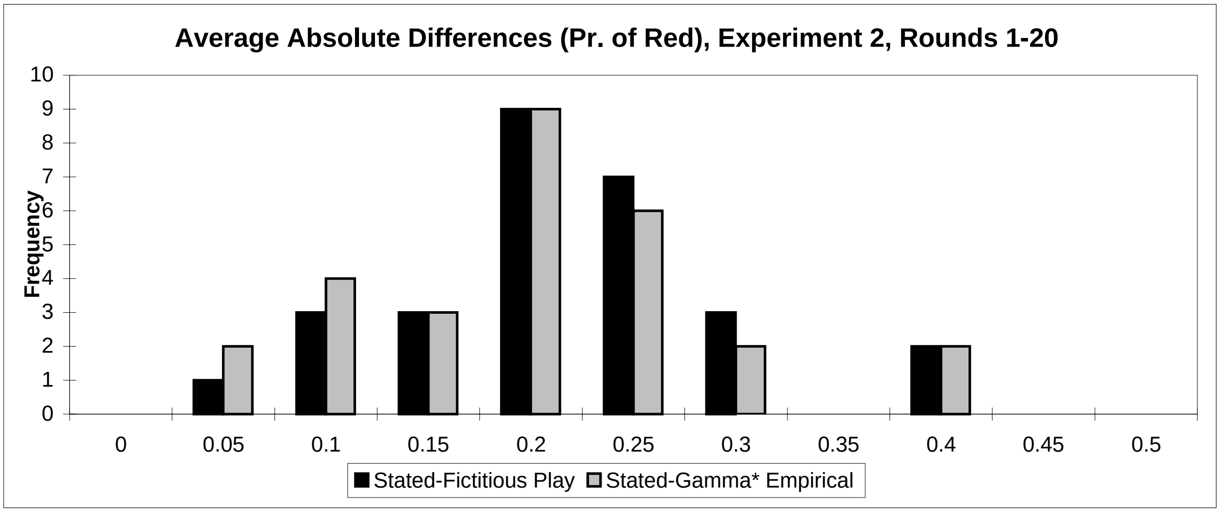


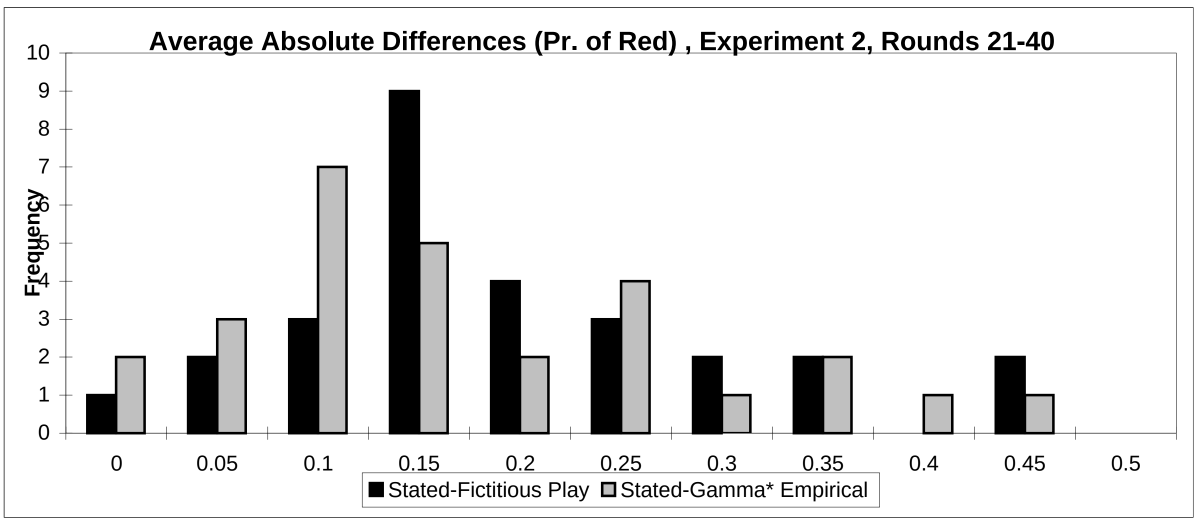
Average Absolute Differences (Pr. of Red), Experiment 1, Rounds 21-40
Frequency
Stated-Empirical
Stated-Gamma* Empirical



Average Absolute Differences (Pr. of Red), Experiment 1, Rounds 41-60
Frequency
Stated-Empirical
Stated-Gamma* Empirical

Figure 1b


Average Absolute Differences (Pr. of Red), Experiment 2, Rounds 1-20
Frequency
0
1
2
3
4
5
6
7
8
9
10
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
Stated-Fictitious Play
Stated-Gamma* Empirical


Average Absolute Differences (Pr. of Red) , Experiment 2, Rounds 21-40
Frequency
0
1
2
3
4
5
6
7
8
9
10
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
Stated-Fictitious Play
Stated-Gamma* Empirical

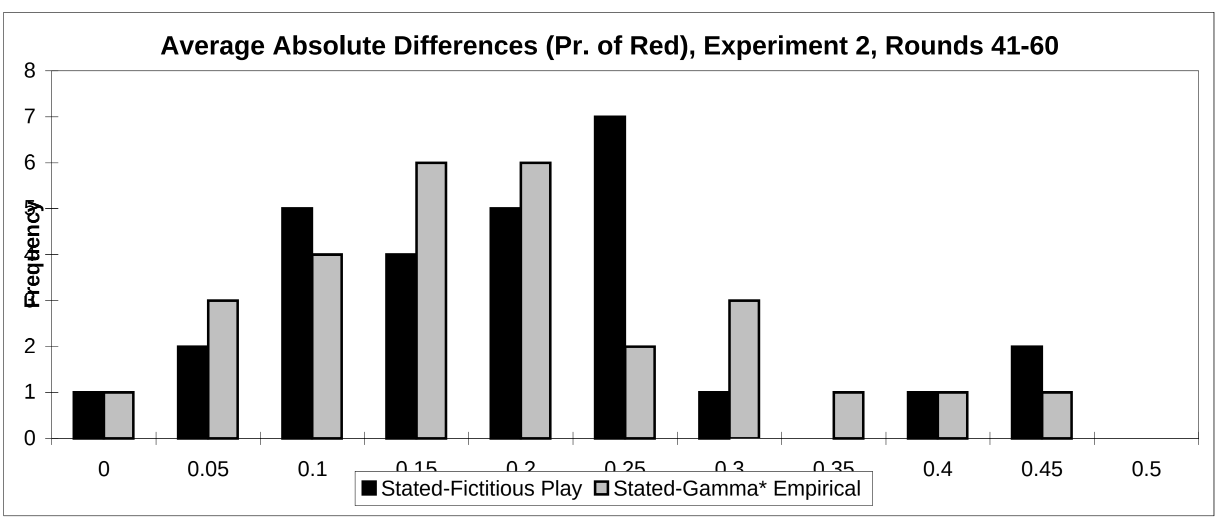
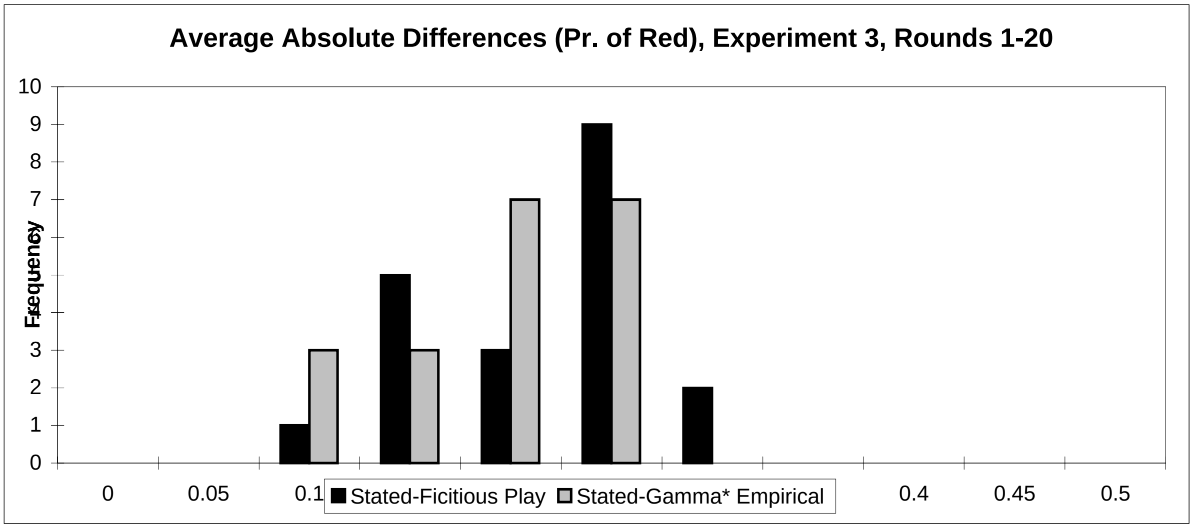
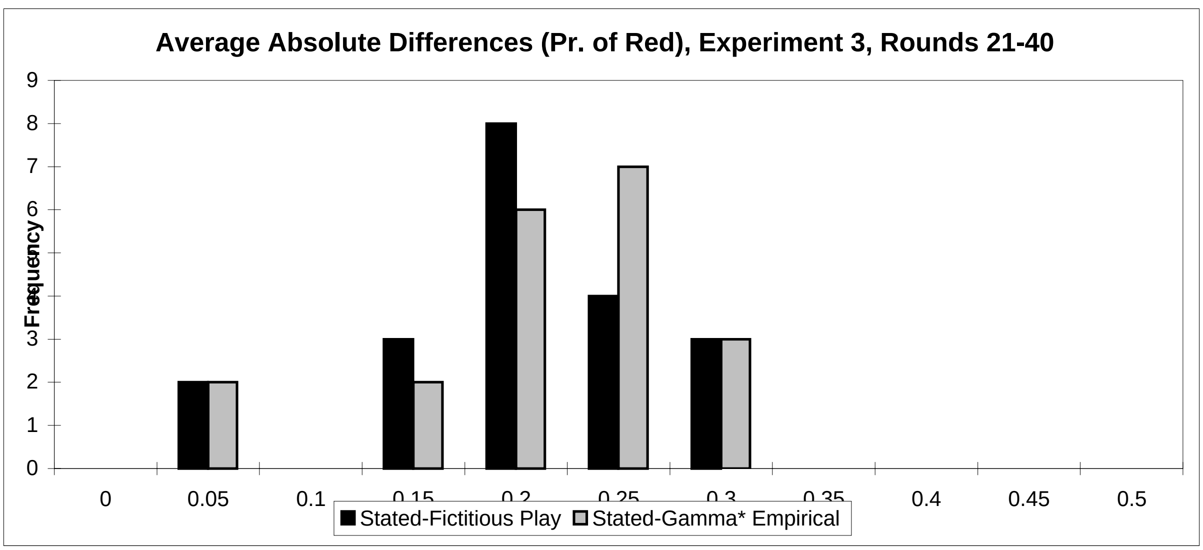
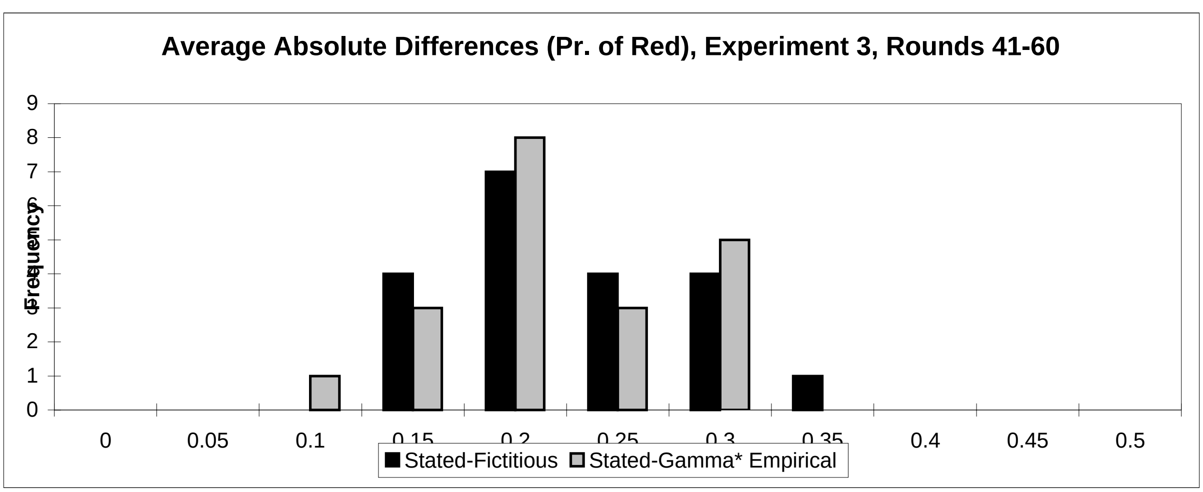

Average Absolute Differences (Pr. of Red), Experiment 2, Rounds 41-60
Frequency
0
1
2
3
4
5
6
7
8
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
Stated-Fictitious Play
Stated-Gamma* Empirical

Figure 1c


Average Absolute Differences (Pr. of Red), Experiment 3, Rounds 1-20
Frequency
0
1
2
3
7
8
9
10
0
0.05
0.1
0.4
0.45
0.5
Stated-Ficitious Play
Stated-Gamma* Empirical


Average Absolute Differences (Pr. of Red), Experiment 3, Rounds 21-40
Frequency
0
1
2
3
6
7
8
9
0
0.05
0.1
0.4
0.45
0.5
Stated-Fictitious Play
Stated-Gamma* Empirical


Average Absolute Differences (Pr. of Red), Experiment 3, Rounds 41-60
Frequency
0
1
2
7
8
9
0
0.05
0.1
0.4
0.45
0.5
Stated-Fictitious
Stated-Gamma* Empirical

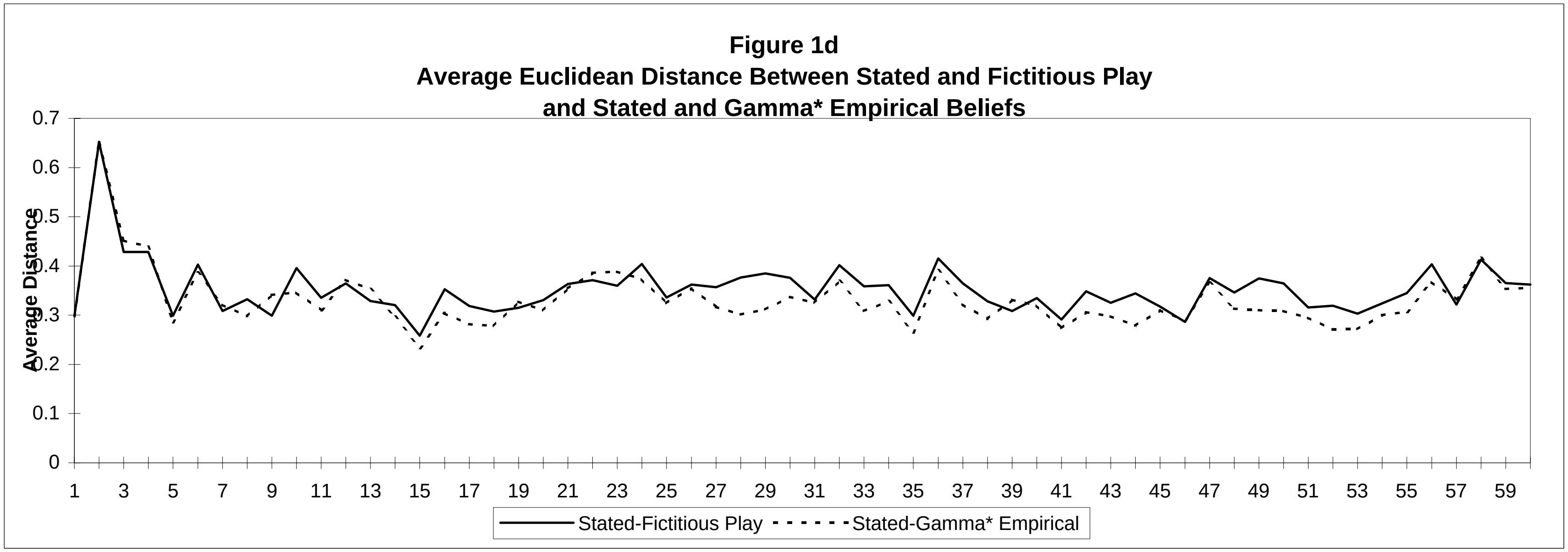


Figure 1d
Average Euclidean Distance Between Stated and Fictitious Play
and Stated and Gamma* Empirical Beliefs
Average Distance
Stated-Fictitious Play
Stated-Gamma* Empirical

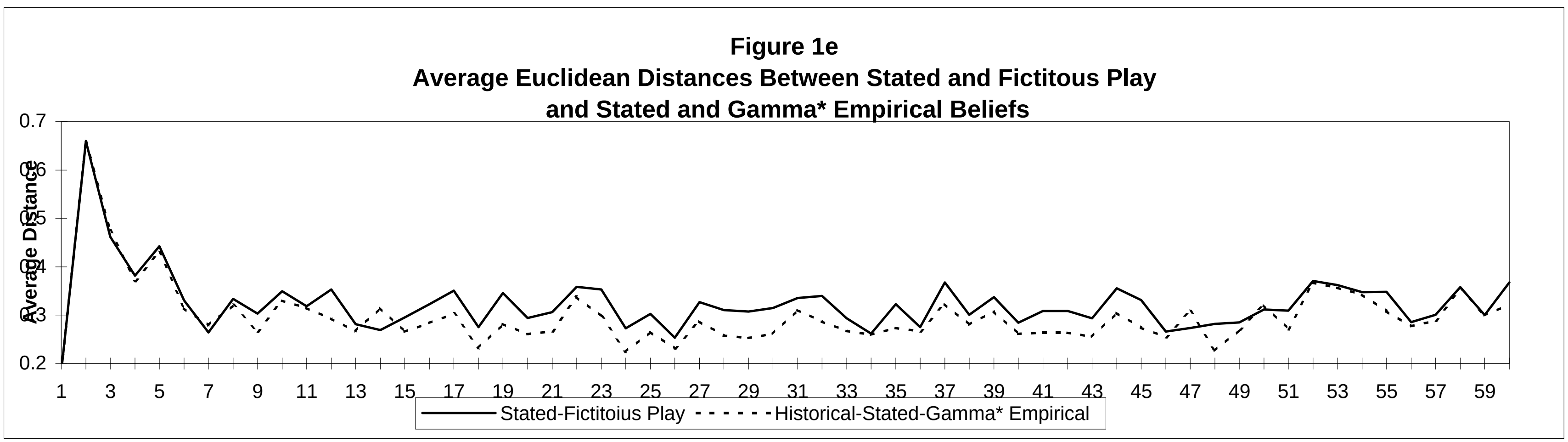


Figure 1e
Average Euclidean Distances Between Stated and Fictitous Play
and Stated and Gamma* Empirical Beliefs
Average Distance
Stated-Fictitoius Play
Historical-Stated-Gamma* Empirical

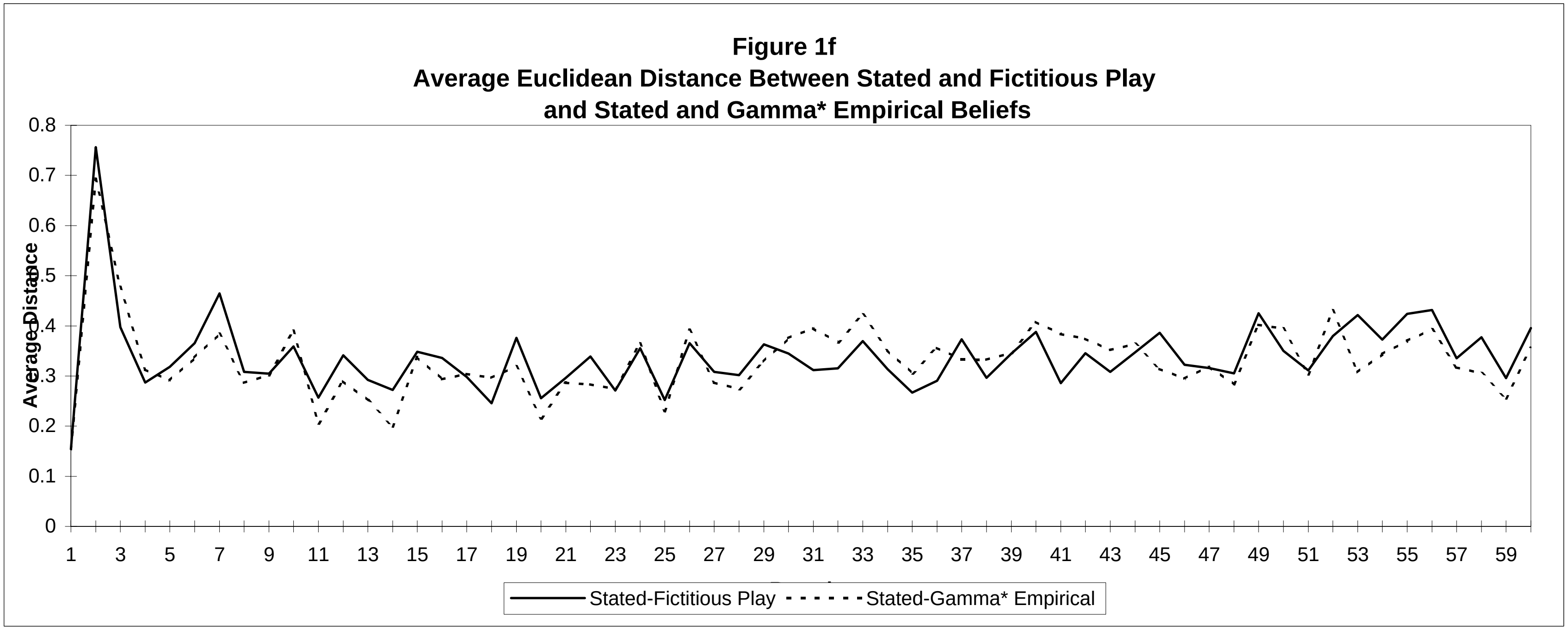


Figure 1f
Average Euclidean Distance Between Stated and Fictitious Play
and Stated and Gamma* Empirical Beliefs
Average Distance
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
3
5
7
9
11
13
15
17
19
21
23
25
27
29
31
33
35
37
39
41
43
45
47
49
51
53
55
57
59
Stated-Fictitious Play
Stated-Gamma* Empirical

Figure 2a: Stated vs Fictitious Play Beliefs Player 2 Low Information Experiment (Exp. 2))

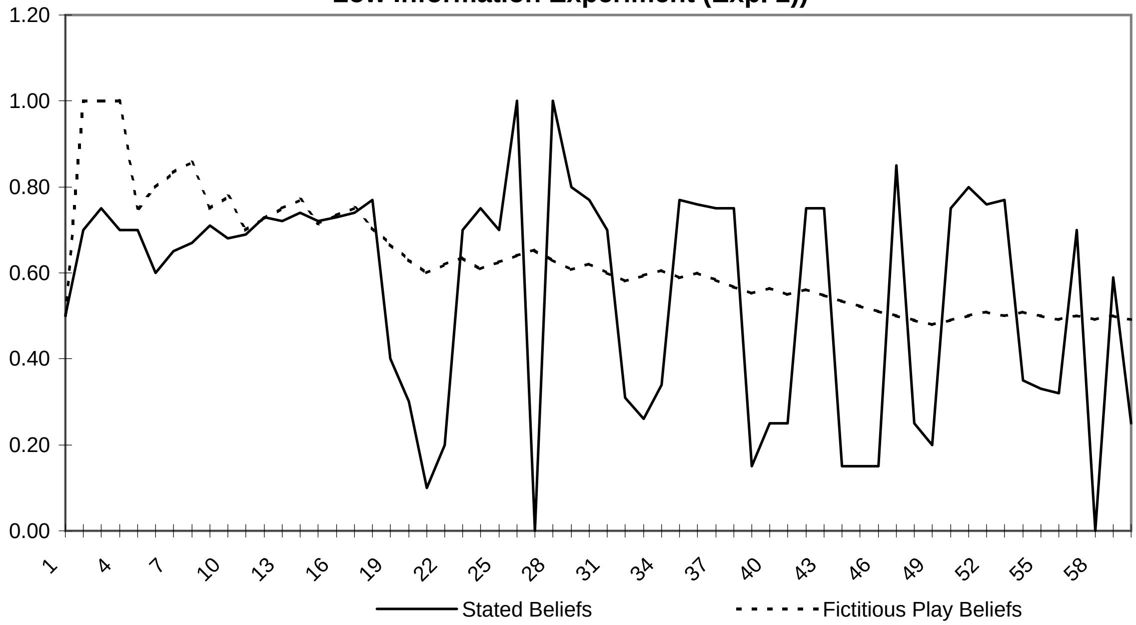


Figure 2b:Stated vs Fictitious Play Beliefs Player 7 Low Information Experiment (Exp. 2)

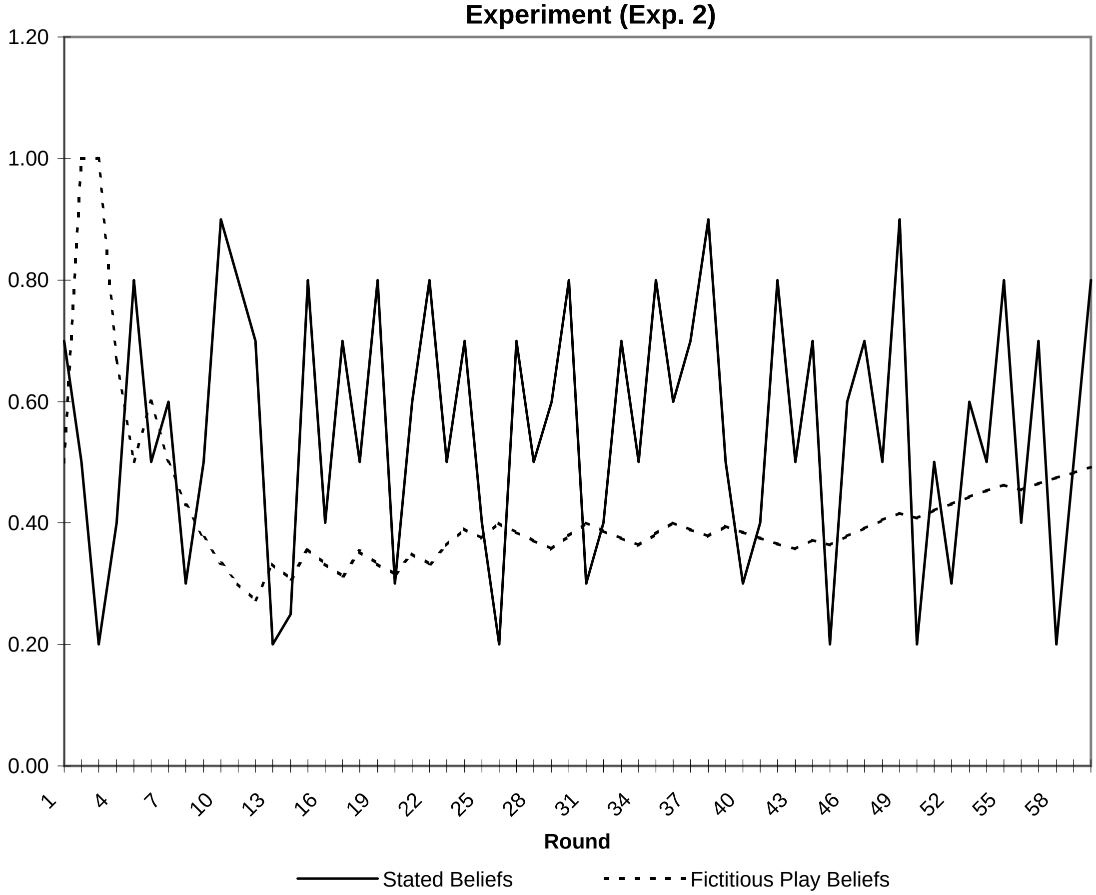


Figure 2c: Stated vs Fictitious Play Beliefs Player 5
Low Information Experiment (Exp. 2)

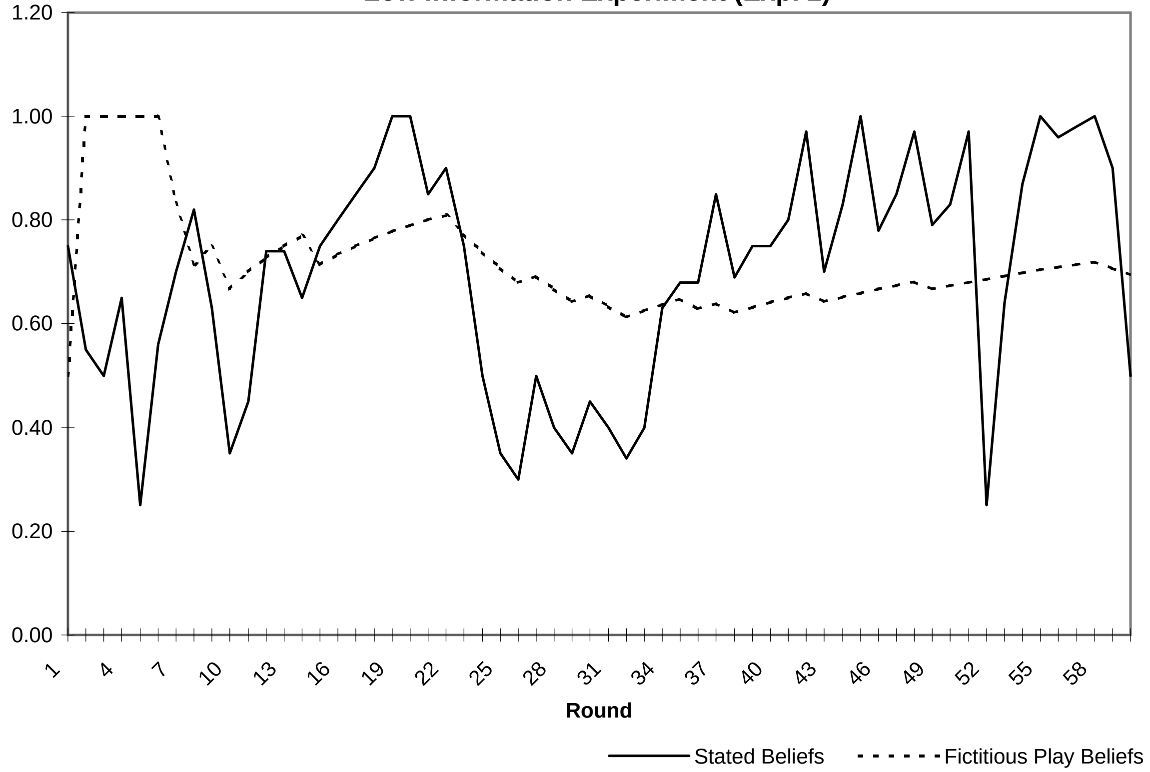
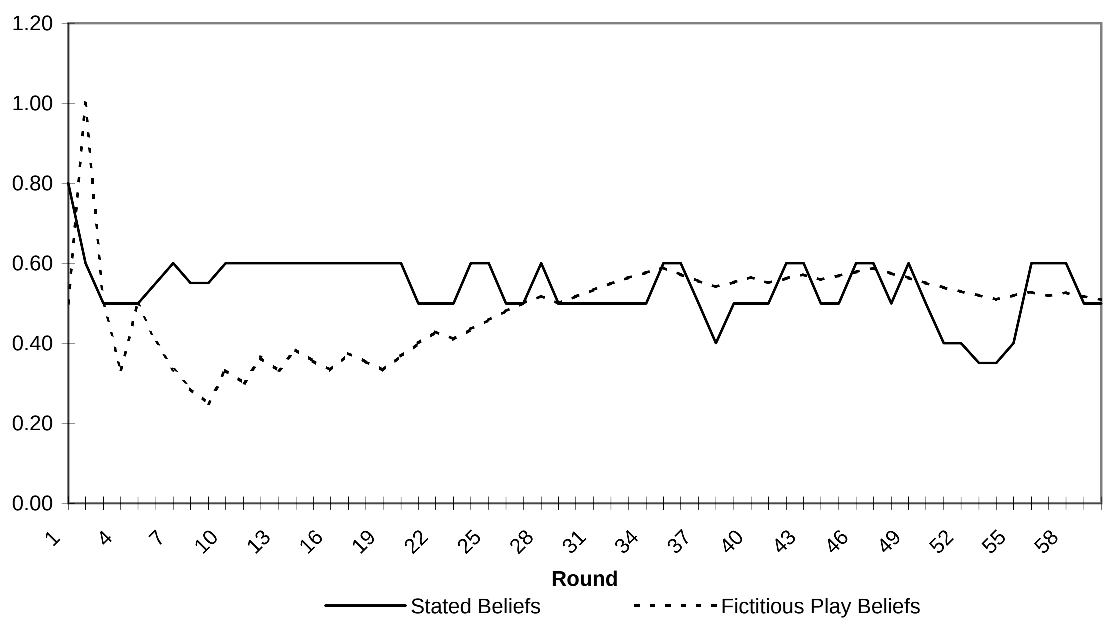


Figure 2d: Stated vs Fictitious Play Beliefs Player 4
Low Information Experiment (Exp. 2)



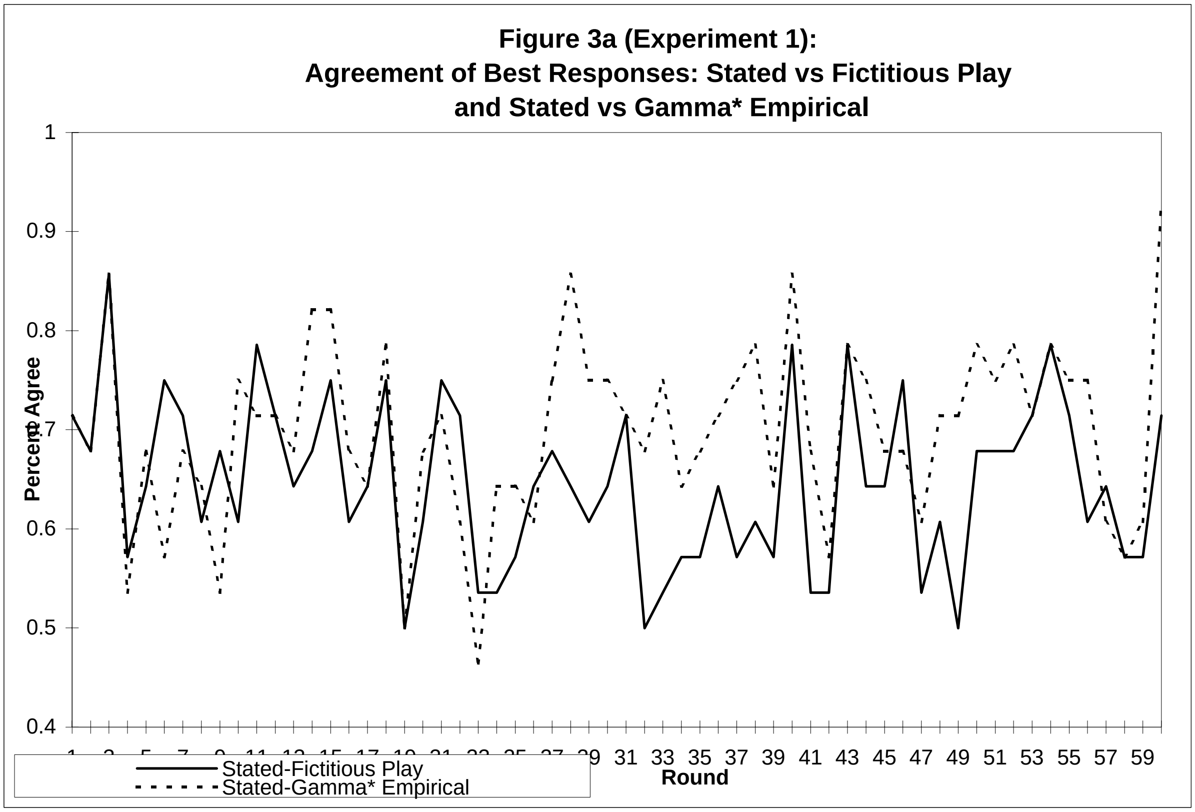


Figure 3a (Experiment 1):
Agreement of Best Responses: Stated vs Fictitious Play
and Stated vs Gamma* Empirical
Percent Agree
Round
Stated-Fictitious Play
Stated-Gamma* Empirical

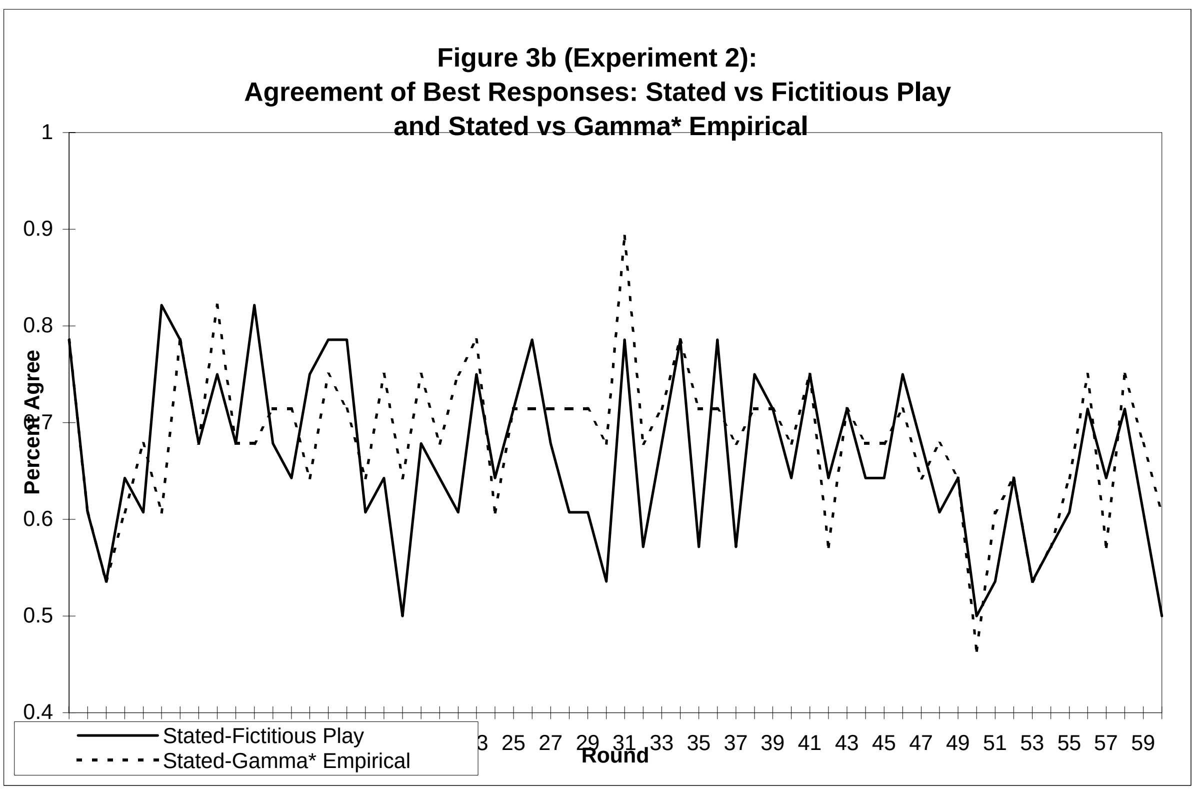


Figure 3b (Experiment 2):
Agreement of Best Responses: Stated vs Fictitious Play
and Stated vs Gamma* Empirical
Percent Agree
Round
Stated-Fictitious Play
Stated-Gamma* Empirical

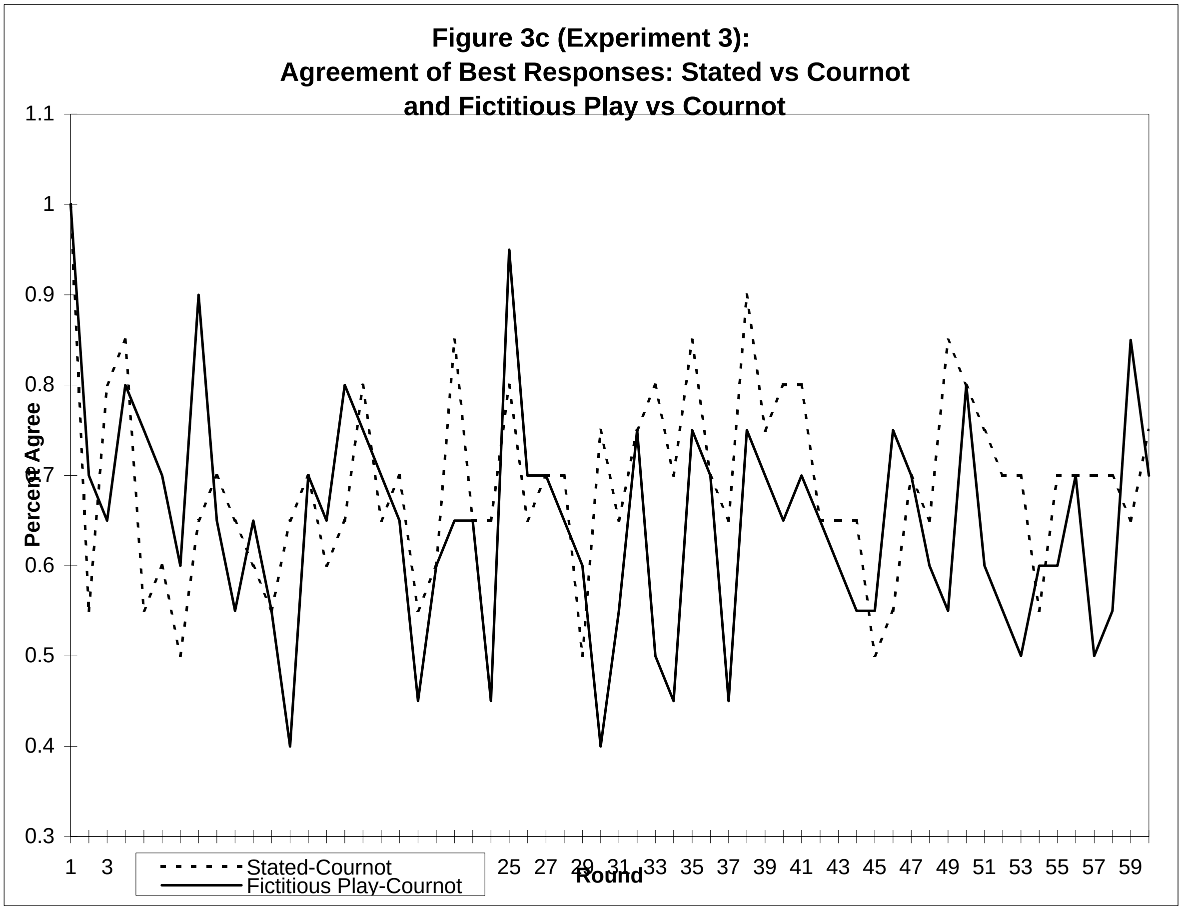


Figure 3c (Experiment 3):
Agreement of Best Responses: Stated vs Cournot
and Fictitious Play vs Cournot
Percent Agree
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
1 3 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59
Round
Stated-Cournot
Fictitious Play-Cournot

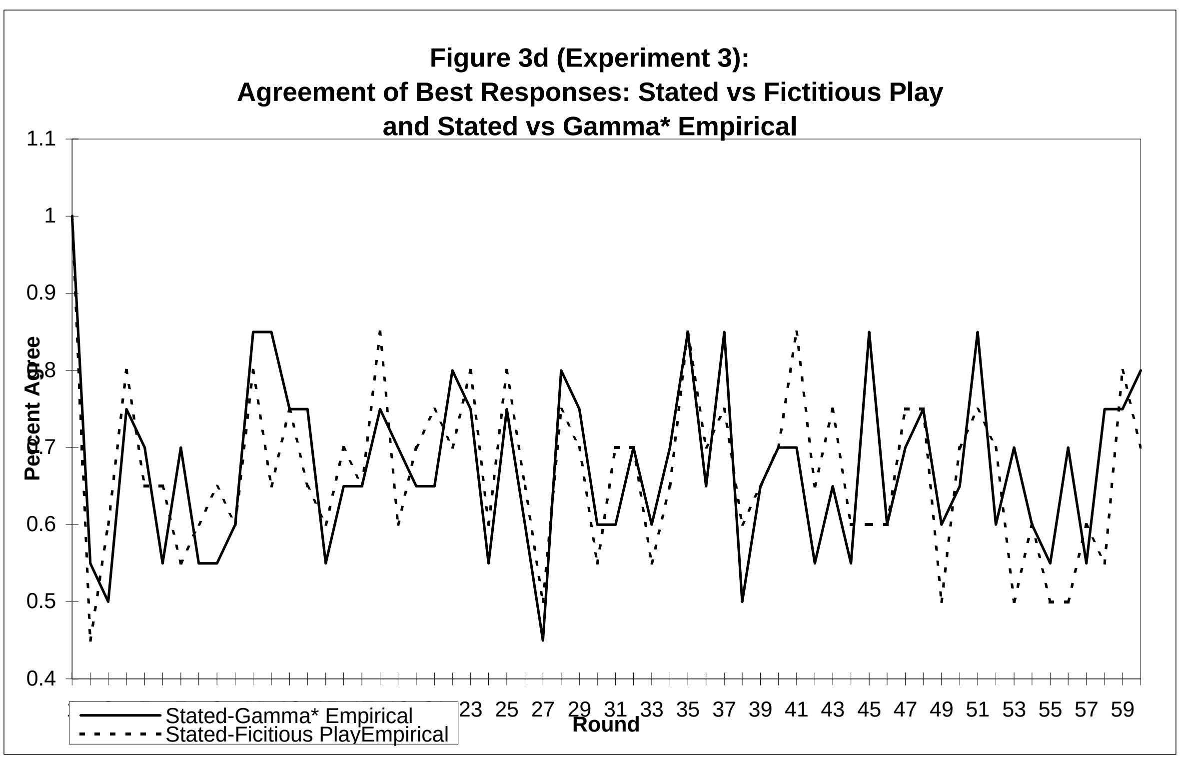


Figure 3d (Experiment 3):
Agreement of Best Responses: Stated vs Fictitious Play
and Stated vs Gamma* Empirical
Percent Agree
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59
Round
Stated-Gamma* Empirical
Stated-Ficitious PlayEmpirical

Figure 4a: Gamma* Empirical and Estimated Gammas:Exp. 1

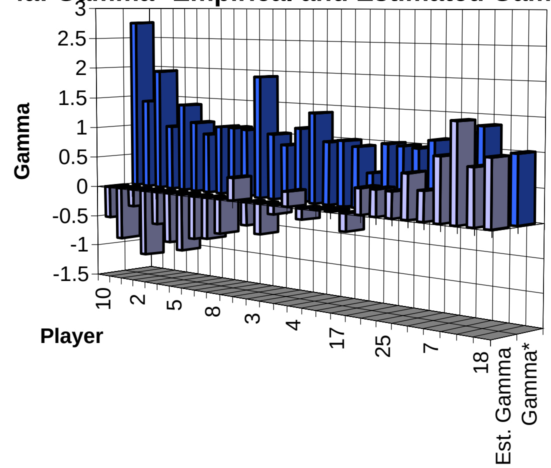


Figure 4b: Gamma* Empirical and Estimated Gammas: Exp. 2

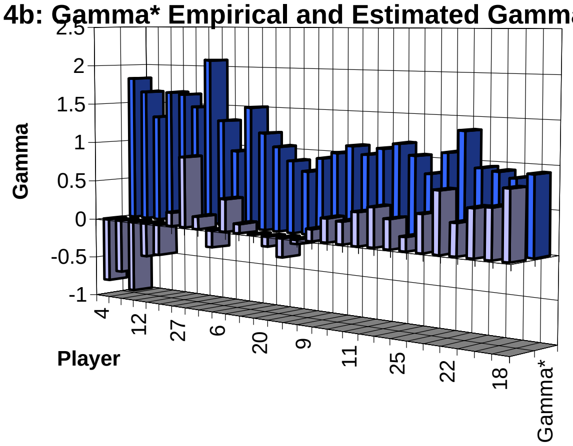
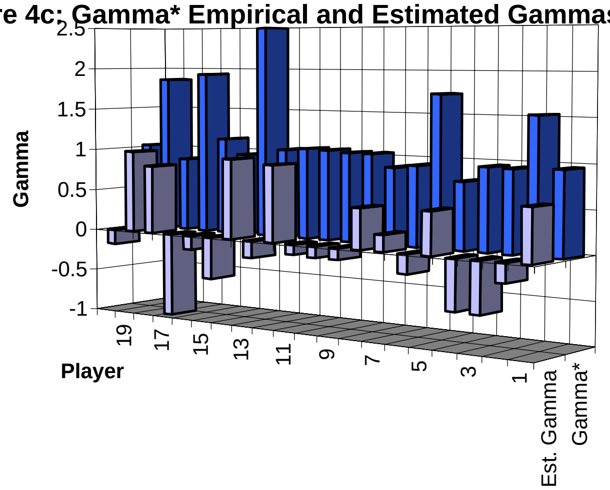


Figure 4c: Gamma* Empirical and Estimated Gammas: Exp. 3



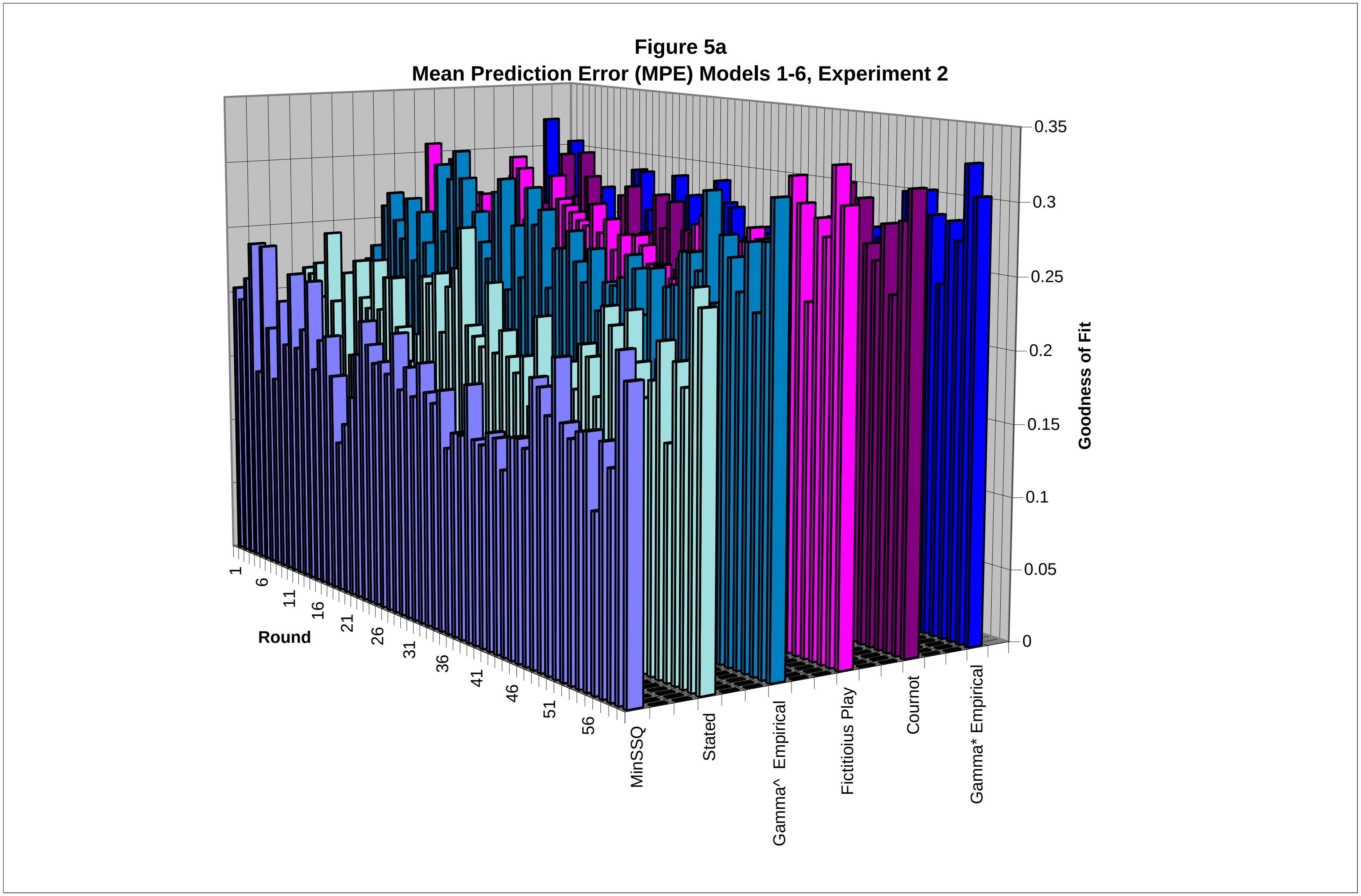


Figure 5a
Mean Prediction Error (MPE) Models 1-6, Experiment 2
Goodness of Fit
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Round
1
6
11
16
21
26
31
36
41
46
51
56
MinSSQ
Stated
Gamma^ Empirical
Fictitioius Play
Cournot
Gamma* Empirical

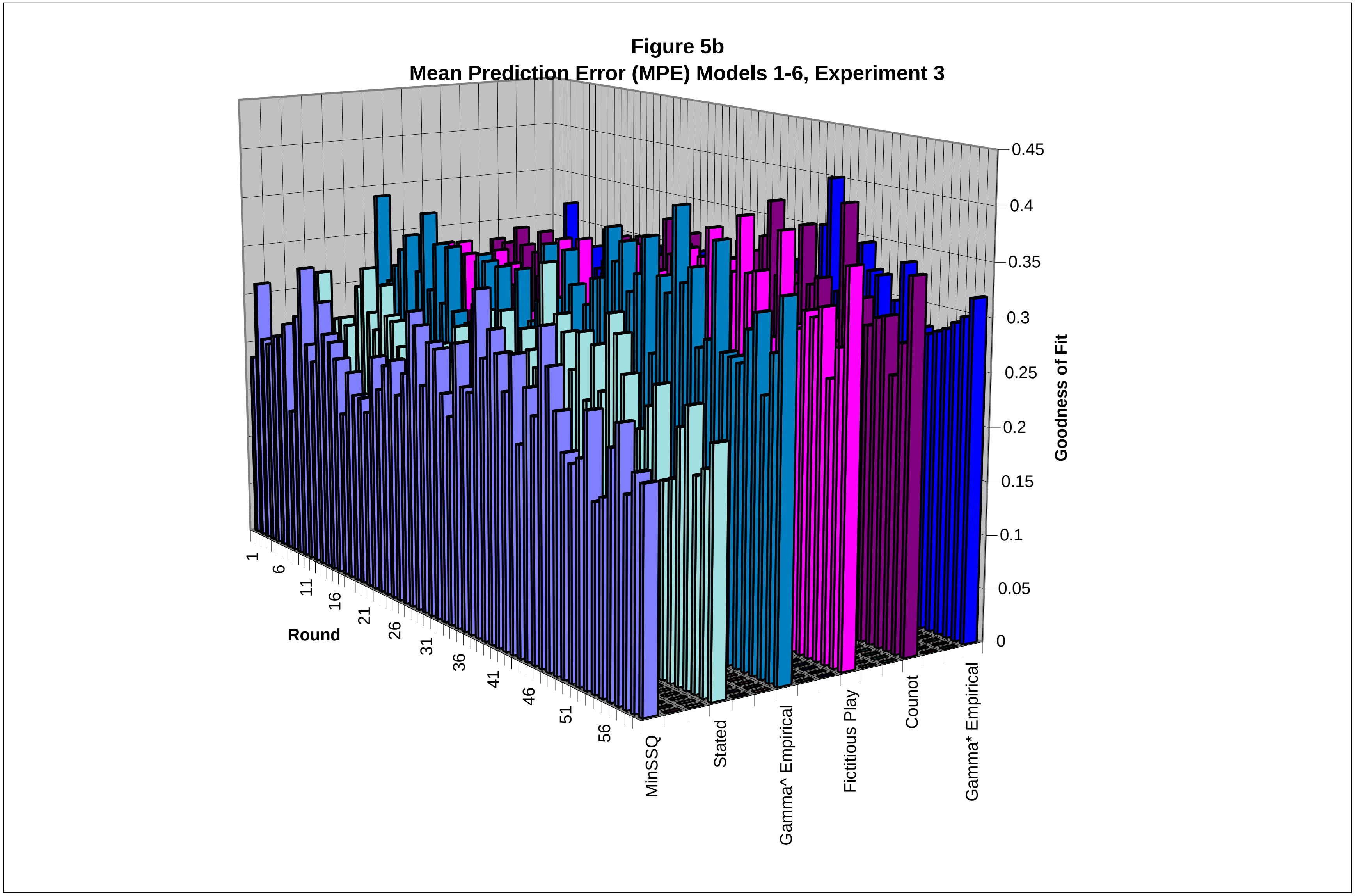


Figure 5b
Mean Prediction Error (MPE) Models 1-6, Experiment 3
Round
1
6
11
16
21
26
31
36
41
46
51
56
MinSSQ
Stated
Gamma^ Empirical
Fictitious Play
Counot
Gamma* Empirical
Goodness of Fit
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45