Econometrica, Vol. 68, No. 3 (May, 2000), 605–641

REINFORCEMENT-BASED VS. BELIEF-BASED LEARNING MODELS IN EXPERIMENTAL ASYMMETRIC-INFORMATION GAMES

BY NICK FELTOVICH[1]

This paper examines the abilities of learning models to describe subject behavior in experiments. A new experiment involving multistage asymmetric-information games is conducted, and the experimental data are compared with the predictions of Nash equilibrium and two types of learning model: a reinforcement-based model similar to that used by Roth and Erev (1995), and belief-based models similar to the "cautious fictitious play" of Fudenberg and Levine (1995, 1998). These models make predictions that are qualitatively similar—cycling around the Nash equilibrium that is much more apparent than movement toward it. While subject behavior is not adequately described by Nash equilibrium, it is consistent with the qualitative predictions of the learning models. We examine several criteria for quantitatively comparing the predictions of alternative models. According to almost all of these criteria, both types of learning model outperform Nash equilibrium. According to some criteria, the reinforcement-based model performs better than any version of the belief-based model; according to others, there exist versions of the belief-based model that outperform the reinforcement-based model. The abilities of these models are further tested with respect to the results of other published experiments. The relative performance of the two learning models depends on the experiment, and varies according to which criterion of success is used. Again, both models perform better than equilibrium in most cases.

KEYWORDS: Equilibrium, asymmetric information, zero-sum game, learning, calibration, model comparison.

1. INTRODUCTION

LEARNING MODELS ARE BECOMING more and more widely used by experimental economists as alternatives to the equilibrium predictions of game theory. A number of researchers have presented results demonstrating that, in many cases, learning models are better able to describe and predict experimental results than static Nash equilibrium. The models that have been proposed vary widely in the way they suppose learning occurs, and they often differ to some extent in their predictions. Many fall into one of two broad classes: *belief-based* models and *reinforcement-based* models. Belief-based models have in common the assumption that players hold beliefs over the likely play of others, and they choose their own actions based in some way on their expected payoff given these beliefs. Reinforcement-based models, on the other hand, do not explicitly

[1] Part of this research was done while the author was supported by a University of Pittsburgh Andrew Mellon predoctoral fellowship. I thank David Cooper, John Duffy, Ido Erev, John Kagel, John Miller, Jack Ochs, Florenz Plassman, Bob Slonim, Nathaniel Wilcox, three anonymous referees, and especially an editor, Alvin Roth, and Shmuel Zamir for helpful discussions and comments. Any remaining errors are actually part of the equilibrium strategy of a more complex game.

require players to form beliefs about other players' likely actions (in fact, they may not even require players to realize that there are other players). Rather, their strategies receive reinforcement related to the payoffs they earn, and over time players adjust their play so that strategies leading to higher payoffs become more likely.

The objective of this paper is to examine the abilities of reinforcement-based and belief-based learning models to characterize the play of subjects in experiments. We accomplish this objective in two ways. First, we conduct a new experiment and compare the observed subject behavior to the predictions of these two classes of learning model, as well as Nash equilibrium, using several criteria of goodness of fit. Second, we use these same criteria to compare data from previous experiments by other researchers to the predictions of these models.

The new experiment uses members of a class of multistage two-player constant-sum games (described in Section 2), whose strategic structure involves the optimal use of private information. Besides being a severe test of the assumptions of Nash equilibrium, this environment seems a natural one for eliciting belief-based learning. Both types of learning model can capture some of the subtle strategic interplay essential to games of this type. In addition, the learning models give rise to dynamics which behave quite differently from equilibrium play, and to some extent, from each other. Play in the experiment is described poorly by stationary Nash equilibrium; in both games, play starts and remains far away from the equilibrium outcome, even after many repetitions. In contrast, given the behavior of subjects during early rounds of the experiment, both types of learning model describe well some gross qualitative features of the changes in subject behavior from early rounds to later rounds. We quantitatively compare the models using several criteria of goodness of fit. Which learning model best characterizes the experimental data depends on which criterion is used. The belief-based models (for some parameterizations) describe aggregate behavior better, while the reinforcement-based model makes more accurate predictions of behavior at the individual level. However, according to almost all of our criteria, both types of learning model perform substantially better than Nash equilibrium; the improvement achieved by moving from one learning model to another is small relative to that achieved by moving from equilibrium to either type of learning model.

This paper is not the first attempt to compare learning models. Boylan and El-Gamal (1993) compared the predictions of two belief-based models (Cournot play and fictitious play) applied to two earlier experiments using three 15×15 simultaneous-move games and six 3×3 simultaneous-move games. They found that for some games, play was more likely to be consistent with Cournot learning, while for other games, play was more likely to be consistent with fictitious play learning. However, when the entire set of experimental data was considered (using a Bayesian updating procedure), fictitious play was overwhelmingly more likely to be the correct model. Cheung and Friedman (1997) also compared different belief-based models; in addition to Cournot play and fictitious play, they considered a hybrid model that included these two models as special cases, and further generalized the model with two additional parameters. (The resulting model is very similar to the belief-based models we use in this paper.) Cheung and Friedman performed an experiment using four 2×2 games and several experimental procedures, and found that there is substantial heterogeneity in learning behavior across experimental subjects. The play of some subjects was consistent with Cournot, the play of others was consistent with fictitious play, and the play of yet others was consistent with neither, but with the more general model. Like Boylan and El-Gamal, Cheung and Friedman found that both the game and the experimental procedures used affected the type of learning that took place, in the sense that the distribution of model parameters varied across games (though one parameter changed very little from game to game). Sarin and Vahid (1999) used data from several experiments (the same experiments as used by Erev and Roth (1998), described below) to compare two reinforcement-based models: the “fixed reference point” model of Erev and Roth (1998), and their own “SV” model, which assumes that an agent chooses the strategy that maximizes payoff, given the agent’s beliefs about the likely payoff to each strategy, which are a weighted average of past payoffs to each strategy. Sarin and Vahid found that the SV model describes play at least as well as, and sometimes better than, the fixed reference point model.

Recently, some researchers have compared not just different learning models, but different *types* of learning model. Mookherjee and Sopher (1994) compared the predictions of a reinforcement-based model to those of a belief-based model (fictitious play). They used data from a Matching Pennies experiment, which they performed with two information treatments: one in which players were told only their own payoffs at the end of a round, and one in which they were told the actual payoff matrix and their opponent choices, as well as their own payoffs. The former treatment gives subjects enough information for reinforcement learning, but not enough for belief learning, and thus seems tailor-made for reinforcement learning; indeed, observed play was consistent with the reinforcement-based model. The latter treatment gives subjects enough information for belief learning, but neither the belief-based model nor the reinforcement-based model described play well. Mookherjee and Sopher (1997) also compared reinforcement-based and belief-based models, this time with regard to data from a new experiment which used two 4×4 and two 6×6 games. When these slightly more complex games were used, they found that the reinforcement-based model did a better job of describing play than the belief-based model, even though subjects were given enough information for belief learning.

A possible shortcoming of Mookherjee and Sopher’s work is that, though they compared many different learning models, they used only the data from their own experiments. Thus the generality of their results is open to question.[2] Two recent papers have used data from several experiments, performed by different

[2] The advantage of using data from many experiments has been pointed out by Erev and Roth (1998).

researchers under different sets of experimental procedures. Erev and Roth (1998) compared the ability of reinforcement-based models and belief-based models to characterize aggregate behavior in several one-stage simultaneous-move games with unique equilibria in completely mixed strategies. They found that the reinforcement-based models describe play better than the belief-based models, and that the better belief-based models are the ones that are more similar to reinforcement-based models. Camerer and Ho (1996, 1999) considered a general model that includes both belief-based and reinforcement-based models as special cases, and used several one-stage simultaneous-move games. By fitting model parameters via maximum likelihood estimation, they concluded that versions of their model that combine elements of belief-based and reinforcement-based learning perform much better than either pure belief-based models or pure reinforcement-based models.

These papers have shed light on the nature of learning in games, but they are not without their shortcomings. Erev and Roth considered only a handful of belief-based models chosen ex ante, so the fact that they found a reinforcement-based model to work well may merely reflect "bad luck" in their choice of belief-based model specification.[3] (They perform sensitivity analyses on the reinforcement-based model, so it is unlikely that their results are due to particularly "good luck" in their choice of reinforcement-based model specification.) On the other hand, Camerer and Ho's model is so general that there may exist a parameterization that is consistent with any experimental outcome; moreover, when they find that a particular combination of parameters best fits a set of experimental data, questions come to mind concerning how sensitive these best parameters are to small changes in the game, whether one could predict which parameter values are appropriate for which games, and so on. Even within their sample of games, they found that the best parameters vary greatly. Additionally, all of the work cited above considered only a narrow class of games —those in which players' strategies are single actions—so it is unclear how generalizable the conclusions are to more complex games. The asymmetric-information games used in our new experiment address this last limitation; they are more complex than the simultaneous-move games studied by Erev and Roth and by Camerer and Ho. Also, we look at a two-parameter family of belief-based models, which is part of the class proposed by Fudenberg and Levine (1995, 1998); this family of models is more general than that used by Erev and Roth, but not as general as that used by Camerer and Ho.

The results of the second part of this paper reconcile somewhat the seemingly different results of Erev-Roth and Camerer-Ho. We revisit some of the data sets they considered, and apply the same criteria we used to compare the learning models' predictions regarding the asymmetric-information game. Some of our criteria are similar in spirit to those used by Erev and Roth, and some are similar to those used by Camerer and Ho. No criterion is used by both, however,

[3] To be fair, it should be pointed out that they make no claim to have found the "correct" model of learning, merely that the reinforcement-based model does a good job of characterizing certain features of the data.

so even though they appeared to reach different conclusions (particularly concerning the utility of reinforcement-based models), it was unclear whether this was because they used different games or different criteria of goodness of fit. We find that both are important; the relative abilities of the learning models to characterize experimental data sets depend on both the game used and the criterion of goodness of fit. Thus neither type of model is always better. However, as was true when the asymmetric-information game data were used, both types of learning model usually represent a substantial improvement over static Nash equilibrium.

2. THE NEW EXPERIMENT

The new experiment uses a class of games from Aumann and Maschler (1995); see Figure 1. First, nature chooses one of two payoff matrices; the Left matrix is chosen with probability $p \in (0,1)$. The chosen matrix determines a constant-sum stage game, which is played twice.[4] We denote by *stage* one play of

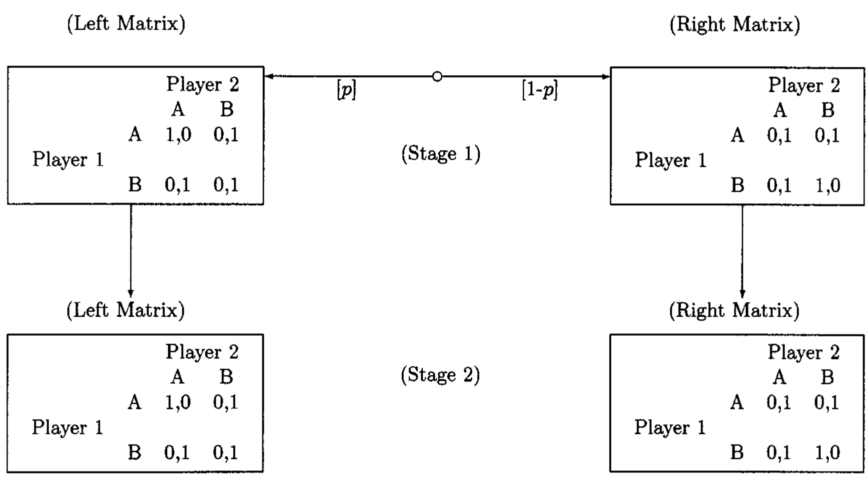

FIGURE 1.—The component game.

[4] A constant-sum game is used in an attempt to keep subjects' preferences as closely tied as possible to the payoffs of the game; we try to limit the influence of subjects' tastes for nonpecuniary aspects of outcomes, such as fairness or spite. (A further step in this direction is our repeated use of loaded words such as "opponent" in the instructions to subjects.) These steps are taken in order to maximize the likelihood that any lack of equilibrium play in the experiment is not due to subjects' unwillingness to play what we think to be their equilibrium strategies, but rather their inability to figure out what they are (or belief that their opponents are unable to figure out what they are, or belief that their opponents believe that they are unable to figure out what they are, etc.). This is important when studying learning, because some researchers have shown that static concepts that generalize Nash equilibrium to allow for players' tastes for such nonpecuniary aspects can often describe experimental subject behavior well. Some examples of such generalizations are Rabin's (1993) fairness equilibrium, Fehr and Schmidt's (1999) fairness theory, Bolton and Ockenfels's (1998) theory of equity, reciprocity, and competition, and Levine's (1998) theory of altruism.

the stage game, and by *round* one play of the entire (two-stage) game. Player 1 (the row player) is told at the beginning of the round which matrix was chosen; Player 2 (the column player) is not told until the end of the round. A player's payoff for the round is equal to the sum of the two stage-game payoffs, calculated according to the matrix that was chosen. Players' actions are announced at the end of each stage, but payoffs are not announced until the end of the round. We refer to members of this class of games as $G(p)$ for given p. We will deal with two particular games: $G(.50)$ and $G(.34)$.

2.1. *Strategies and Notation*

We will restrict our attention to two components of players' behavioral strategies: Player 1's first-stage move (conditional on nature's move) and Player 2's second-stage move (conditional on Player 1's first-stage move). It is in these components that the complexity of the game lies. To see why, recall that Player 1 has a piece of private information—the actual payoff matrix. It can be seen from Figure 1 that in either matrix, one of her actions (A in the Left matrix and B in the Right matrix) can possibly earn her a payoff of one for the stage (depending on Player 2's choice of action), while the other action gives her zero for the stage with certainty. We will use the term *stage-dominant action* (*sda*) to refer to the former, since these actions correspond to Player 1's weakly dominant strategy in the one-stage analogue to this game. Since the *sda* depends on which matrix was chosen by nature, Player 1's private information is potentially valuable in the sense that she may use it to earn a payoff higher than $2p(1-p)$, the maxmin expected payoff she would receive if she did not know the payoff matrix. However, the only way Player 1 can benefit from her private information is by *revealing* it, that is, by playing in the Left game differently (possibly in a stochastic way) from her play in the Right game. Player 1 *completely reveals* by playing the *sda* in the first stage with certainty, that is, by playing A if the Left game is chosen and B if the Right game is chosen. She *partially reveals* by playing A in the first stage with higher probability in the Left game than in the Right game. It may seem that completely revealing is the best course of action for Player 1, since it gives her the best chance of earning the point in the first stage. But since her first-stage actions are observable by Player 2 before he chooses his second-stage actions, Player 2 would then be able to infer the payoff matrix from Player 1's first-stage action, and could hold Player 1 to a payoff of zero in the second stage by playing B in the Left game and A in the Right game; this set of actions is the best response to Player 1's completely revealing strategy (*brcr*). Player 2's playing the *brcr* against a completely revealing Player 1, along with suitable play in the first stage, limits Player 1 to an expected payoff of $\text{Min}\{p, 1-p\}$, making her no better off, and for almost all p strictly worse off, than if she hadn't had the information at all. Even if Player 1 is only partially revealing, Player 2 can use Bayesian updating to revise his prior beliefs about which matrix was chosen and lower Player 1's expected payoff. Player 1's primary strategic problem is to choose her first-stage action to balance the

TABLE I

NOTATION

Symbol	Meaning: "Probability of ..."
$P(sda_1 \mid L)$:	stage-dominant action in the first stage of the Left matrix (Player 1)
$P(sda_1 \mid R)$:	stage-dominant action in the first stage of the Right matrix (Player 1)
$P(brcr \mid A)$:	best response to completely revealing following a Player 1 choice of A (Player 2)
$P(brcr \mid B)$:	best response to completely revealing following a Player 1 choice of B (Player 2)
$P(sda_1)$:	stage-dominant action in the first stage (Player 1)
$P(brcr)$:	best response to completely revealing (Player 2)

current (first-stage) gain from using her extra information against the future (second-stage) loss caused by revealing it. Player 2's main strategic problem is to choose his second-stage action to take into account his best inference of which matrix was chosen, given his observation of Player 1's behavior. These problems are closely related

Table I shows some of the notation used to describe players' behavioral strategies, and Table II summarizes the Nash equilibrium predictions for the two games. (A full description of the Nash equilibria of $G(p)$ for any $p \in (0,1)$ is given by Feltovich (1997).) The prediction for $G(.50)$ is not unique; there is a one-parameter family of Player 2 strategies, each of which is consistent with equilibrium play. Also, the equilibrium value of $P(brcr)$ depends on Player 1's choices of $P(sda_1 \mid L)$ and $P(sda_1 \mid R)$ and is generally not equal to β. In both games, $P(sda_1 \mid L)$ and $P(sda_1 \mid R)$ are strictly less than one, reflecting Player 1's strategic problem—how quickly to reveal her private information. (In the *second* stage, Player 1 optimally plays her *sda* with probability one, reflecting the fact that at that point, there is no harm to her in revealing all her information.) In $G(.50)$, Player 1's equilibrium strategy has her essentially ignoring the private information in the first stage, while in $G(.34)$, she optimally partially reveals her private information.

2.2. *Experimental Design*

In the experiment, one of these two-stage component games was chosen and played for 40 rounds, with players remaining in the same roles and playing against the same opponent, but with the payoff matrix (Left or Right) chosen

TABLE II

NASH EQUILIBRIUM PREDICTIONS

Game	$P(sda_1 \mid L)$	$P(sda_1 \mid R)$	$P(brcr \mid A)$	$P(brcr \mid B)$	$P(sda_1)$	$P(brcr)$	Expected payoff: Player 1	Expected payoff: Player 2
$G(.50)$	.500	.500	β	$1.500 - \beta$	.500	[.500, 1.000]	0.75	1.25
$G(.34)$	.971	.500	.500	1.000	.660	.670	0.67	1.33

Note: β may take on any value in [0.500, 1.000].

independently each round (and p held fixed). A total of 122 subjects, mostly undergraduates at the University of Pittsburgh, participated in the experiment. Nine experimental sessions were conducted, with between eight and twenty subjects in a session. In addition to a participation fee of \$10.00, subjects could earn a bonus of \$10.00 if they earned a point in a round and stage that were randomly chosen after the experiment.[5]

The payment scheme used here is equivalent to the "binary lottery" scheme of Roth and Malouf (1979); in particular, the utility functions of expected utility maximizing individuals would be linear in the number of points earned. Without loss of generality, we can therefore define payoffs in the forty-times-repeated game to be the sum of the payoffs in each individual component game. Since the games are constant-sum, the corresponding repeated games are also constant-sum. The unique equilibrium of the repeated $G(.34)$ game is the component game equilibrium repeated 40 times. Since $G(.50)$ has a continuum of Nash equilibria, the repeated game does not have a unique Nash equilibrium; any sequence of component-game Nash equilibria (including stationary play of a particular component-game equilibrium) yields a Nash equilibrium of the repeated game.

2.3. *The Reinforcement-Based Model*

The equilibrium prediction for $G(.50)$ and $G(.34)$ is play that is stationary from round to round (the only exception being that, in $G(.50)$, Player 2 can choose each round from a one-parameter family of equilibrium strategies). We now examine the alternative predictions of two types of learning model.

The reinforcement-based model is adapted from the models of Roth and Erev (1995). (See Roth and Erev (1995) and Erev and Roth (1998) for additional motivation behind this model and its generalizations.) Because of its close relation to their basic model, and its lack of free parameters, it will be referred to as the RE_0 model. Roth and Erev successfully use variations of this model for predicting subject behavior in many experiments, despite (or perhaps because of) the low level of rationality the model attributes to individuals. Rather than endowing players with the high degree of cognitive sophistication implicit in equilibrium predictions, this model posits that players merely learn, over time, to play better strategies (strategies leading to higher realized payoffs) more often and worse strategies less often.

Specifically, in round t, players have a nonnegative initial propensity $q^t(\alpha \mid \Psi)$ for playing action α ($\alpha = A$ or B) at information set Ψ. The *strength of propensities* ($Q^t(\Psi)$) in round t is the sum of the propensities for playing both actions at information set Ψ: $Q^t(\Psi) = q^t(A \mid \Psi) + q^t(B \mid \Psi)$. For any $t \geq 1$, propensities for round $t+1$ are found by adding the payoff earned in round t to

[5] The Appendix contains the instructions given to subjects. More details about the experimental procedures are available from the author.

the round-t propensities of the actions that were played in round t:

$$q^{t+1}(\alpha \mid \Psi) = \begin{cases} q^t(\alpha \mid \Psi) + \pi & \text{if } \Psi \text{ was reached, } \alpha \text{ was played, and} \\ & \quad \text{the payoff for the round was } \pi, \\ q^t(\alpha \mid \Psi) & \text{if } \Psi \text{ was not reached or } \alpha \text{ was not played.} \end{cases}$$

Thus in each round, each player will augment two propensities by the total payoff for the round.[6] Initial (round-1) propensities are exogenous. The probability of playing strategy α at information set Ψ in round t is the corresponding propensity, divided by the strength of propensities at information set Ψ in round t:

$$p^t(\alpha \mid \Psi) = \frac{q^t(\alpha \mid \Psi)}{Q^t(\Psi)}.$$

2.4. *The Belief-Based Models*

The RE_0 model assumes that players make decisions according only to past payoffs from actions. Taken literally, this means that decisions are made without regard to many features of the payoff matrix and the history of opponent's plays (though, of course, these affect the player's payoffs and thus indirectly influence later decisions). In the experiment, however, players do know the payoff matrix and are matched to the same opponent in every round (this latter fact should increase the relative usefulness in this experiment of the history of past play in forecasting future opponent actions, compared to many experimental designs). It seems reasonable to expect that an appropriate model of learning in this experiment should take this knowledge into account. Also, the very nature of Player 2's situation (trying to infer the payoff matrix from Player 1's move) suggests that he ought to form beliefs about the way his opponent behaves. Therefore, in addition to the RE_0 model, we consider a two-parameter family of belief-based models adapted from the model of Fudenberg and Levine (1995, 1998).[7]

According to these models, players hold beliefs (conjectures) concerning the likely play of their opponent(s), and they choose strategies based on their expected payoffs given these beliefs. Specifically, players' beliefs are characterized by nonnegative *belief weights* over opponents' actions at each information

[6] The alternative, augmenting each propensity by the payoff for the stage in which it was played, would result in *bluffing* (Player 1 actions that sacrifice any chance of positive payoff in the first stage in favor of a higher probability of positive payoff in the second stage) never being reinforced, even when it is successful.

[7] Similar models have been used for describing subject behavior in experiments by, for example, Cheung and Friedman (1997) and Mookherjee and Sopher (1997).

set. The weight on an opponent playing action α ($\alpha = A$ or B) at information set Ψ in round t is $\omega^t(\alpha \mid \Psi)$. The *strength of beliefs* at information set Ψ ($\Omega^t(\Psi)$) is the sum of the weights at Ψ: $\Omega^t(\Psi) = \omega^t(A \mid \Psi) + \omega^t(B \mid \Psi)$. For any $t \geq 1$, weights for round $t+1$ are found by increasing the weight of each action that was observed in round t:

$$\omega^{t+1}(\alpha \mid \Psi) = \begin{cases} (1-\delta)\,\omega^t(\alpha \mid \Psi) + 1 \\ \quad \text{if } \Psi \text{ was reached and } \alpha \text{ was played by the opponent,} \\ (1-\delta)\,\omega^t(\alpha \mid \Psi) \\ \quad \text{if } \Psi \text{ was not reached or } \alpha \text{ was not played by the opponent.} \end{cases}$$

The parameter δ determines the relative amount of bearing given to past outcomes relative to current outcomes in forming beliefs. If $\delta = 0$, outcomes in all rounds have equal import, while if $\delta = 1$, only the most recent outcome is considered. If $\delta \in (0,1)$, more recent outcomes are more important than previous outcomes, while if $\delta < 0$, the opposite is true. Initial belief weights are exogenous. The assessed probability of the opponent's playing action α at information set Ψ in round t is the corresponding belief weight, divided by the strength of beliefs at information set Ψ in round t:

$$\mu^t(\alpha \mid \Psi) = \frac{\omega^t(\alpha \mid \Psi)}{\Omega^t(\Psi)}.$$

Given these assessed probabilities, a player's perceived expected payoff $\Pi^e(s \mid \mu^t)$ to each available pure strategy s can be calculated. The player's chosen strategy in round t is determined from these expected payoffs; the probability of a player choosing the pure strategy s in round t given beliefs μ^t is

$$\text{(1)} \qquad \text{Prob}(s \text{ chosen in round } t) = \frac{\exp(\lambda \cdot \Pi^e(s \mid \mu^t))}{\sum_{s \in S} \exp(\lambda \cdot \Pi^e(s \mid \mu^t))},$$

where S is the set of all pure strategies. The parameter λ determines the extent to which the player responds optimally to her beliefs. If $\lambda = 0$, she chooses A and B with equal likelihood at each information set irrespective of expected payoff, while as λ gets large, her strategy approaches best-response play; we will call this limiting case the version of the model with $\lambda = \infty$. We will refer to members of this family of models as BA(λ, δ) for given λ and δ (where BA stands for Bayesian).[8]

[8] Fudenberg and Levine (1995, 1998) examine some of the theoretical properties of models such as this one, which they call "cautious fictitious play." In our notation, standard fictitious play would be called the BA(∞, 0) model. The relationship of cautious fictitious play to standard fictitious play is analogous to that between McKelvey and Palfrey's (1995, 1998) quantal (logistic) response equilibrium to Nash equilibrium. In particular, the (stochastic) fixed points of cautious fictitious play are logistic response equilibria.

2.5. *Implications of the Learning Models*

Since play according to either the reinforcement-based or the belief-based models is allowed to change over time, it is plausible that the predictions of these models could differ from those of static Nash equilibrium. In order to get some idea of what these differences are, we simulate experiments based on the two games. For each game, sets of 200 simulations were performed according to the RE_0 and $BA(\infty, 0)$ models. Each simulation consisted of one simulated Player 1 and one simulated Player 2 playing the same game for at least forty rounds. Initial propensities and beliefs were chosen randomly; strengths of initial propensities and beliefs were set to six at each information set.[9]

The simulation results are shown in Figure 2. In order to show the dynamic features more clearly, the unit $(P(sda_1), P(brcr))$ square is partitioned into four rectangular subregions determined by the Nash equilibrium points (in the case of $P(brcr)$ in $G(.50)$, the midpoint of the interval of equilibrium values). In each subregion is a circle and several connected line segments. The center of each circle represents the average starting point of all simulations with starting points in that region, the area of a circle is proportional to the number of simulations

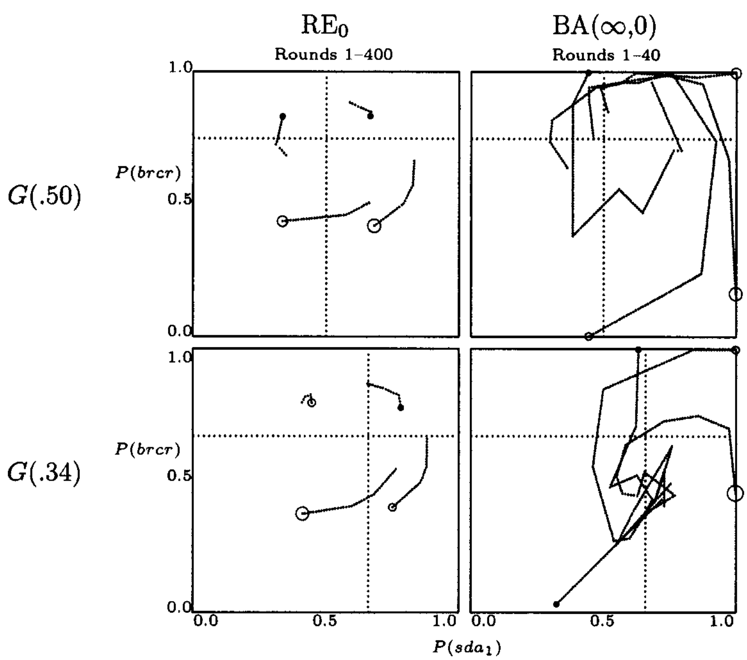

FIGURE 2.—Trajectories of learning model simulations (random initial propensities and beliefs).

[9] The qualitative aspects of the simulation results are robust to small changes in parameter values. For details, see Feltovich (1997).

with starting points in that region, and the line segments follow the average path of play of the simulations starting in that region. In the RE_0 simulations, segment endpoints show average play in rounds 1, 40, 100, and 400; in the $BA(\infty, 0)$ simulations, they show average play in rounds 1 and multiples of 5 up to 40.

In both games, and for both learning models, there is a rough counterclockwise movement of simulation trajectories around the equilibrium (the intersection of the dotted lines). In the RE_0 simulations, there is little if any discernible movement of trajectories toward equilibrium. In the $BA(\infty, 0)$ simulations, such movement can readily be seen; however, it is less pronounced than the counterclockwise movement. Thus, if these learning models accurately describe the way players change their behavior over time, we would expect the trajectories from the actual experiment to move counterclockwise about the equilibrium point (or center of the segment of equilibria) with much less, if any, movement toward this point. Importantly, this implication is quite robust to the choice of learning model. Here, we see that fictitious play and RE_0 have similar implications; Feltovich (1997) shows that simulation trajectories arising from other BA models are qualitatively similar to both.[10]

The implications of the learning models contrast sharply with the equilibrium prediction of stationary play. However, like the equilibria of the games, the learning models demonstrate the dilemma faced by Player 1 in deciding how much (if at all) to reveal in the first stage. If players play according to RE_0, Player 1's over-revealing makes Player 2's expected payoff to playing the *brcr* higher, so that it will be reinforced more on average and eventually played more often. Then, sda_1 will result in lower payoffs on average and will eventually be played less often. If players play according to one of the BA models, Player 1's revealing changes Player 2's beliefs so as to increase his assessed probability of Player 1 revealing, making *brcr* a better response and increasing the likelihood that he actually plays the *brcr*; such play eventually changes Player 1's beliefs, lowering the expected payoff to, and eventually the probability of choosing, the *sda*. By similar reasoning, according to either learning model, Player 1's "under-revealing" would eventually result in Player 2's playing the *brcr* less often, and the change in Player 2's play would eventually cause Player 1 to reveal more often. This argument also explains the counterclockwise direction of the simulation trajectories. At any point in the unit square except for the Nash equilibrium point itself, the best-response correspondence gives $(P(sda_1),$

[10] The similarity between belief-based and reinforcement-based models is not surprising, in view of the fact that they can be shown to be special cases of a more general model. Camerer and Ho's (1999) "experience-weighted attraction (EWA)" is a learning model in which, like RE_0, strategies that are played are reinforced by their payoff, but unlike RE_0, strategies that are *not* played are also reinforced by the payoff they would have earned had they been played, scaled by some constant (normally between zero and one). If the value of this constant is zero, EWA reduces to a model similar to RE_0, given suitable values for EWA's other parameters. Less obviously, if the value of the constant is one, Camerer and Ho show that EWA reduces to a model very much like our BA models (again, given suitable values for other parameters).

$P(brcr)$) pairs that lie counterclockwise from that point. According to either model, if play is initially out of equilibrium, it will (on average) move in the direction of the best response—counterclockwise. As an extreme example, note that Cournot best-response play (the BA(∞, 1) model) would result in individual play paths that jump from the top-right corner to the top-left corner, to the bottom-left corner, to the bottom-right corner, back to the top-right corner, and so on, each found moving to the next corner in the counterclockwise direction.

3. SUMMARY OF EXPERIMENTAL RESULTS

Table III gives a summary of aggregate behavior in the experiment.[11] Shown are the population relative frequencies corresponding to the probabilities $P(sda_1)$ and $P(brcr)$ for the entire forty rounds of each cell, and also for each of four ten-round "blocks" within each cell. (We will refer to the rounds 1–10 as "block 1," rounds 11–20 as "block 2," and so on.) Starred relative frequencies are those that are significantly different from Nash equilibrium behavior.[12]

It is easy to see that Nash equilibrium poorly describes aggregate behavior in this experiment. In both cells, the population relative frequency of sda_1 starts and remains well above the equilibrium prediction, though it decreases over time.[13] In both cells, the population relative frequency of *brcr* increases over time (the direction of the best response to the play of the Player 1 population).[14] These features of behavior are also visible in Figure 3, which plots the relative frequencies of sda_1 and *brcr* in the two cells for each block (the dots connected by lines) and the Nash equilibria. The equilibrium of the $G(.34)$ cell is the center of the '+' and the equilibria of the $G(.50)$ cell are the points of the vertical line segment (the horizontal segment marks the center of the vertical segment). The

[11] The raw data from this experiment are available on request; see Feltovich (1997) for further analysis of the data.

[12] Binomial test. For descriptions of the nonparametric statistical tests used in this paper, see Siegel and Castellan (1988).

[13] A referee pointed out the similarity between the tendency of Player 1s here to overplay the *sda* in the first stage, earning a higher first-stage payoff than in equilibrium but sacrificing the second-stage payoff, and the "melioration" theory of Herrnstein and Prelec (1992). The motivation behind this theory usually relies on examples of dynamic inconsistency in individual choice (for example, "self-control" problems such as overeating or overspending); however, the experimental results used as evidence for melioration generally rely on individuals' not understanding the relationship between current actions and the future decision-making environment (see, e.g., Herrnstein, Prelec, and Vaughan (1986)). A hypothesis explaining the high frequency of *sda* play in the first stage is that Player 1s notice that this action is better (in a myopic sense), but fail to consider the likely response of Player 2s in the second stage; this explanation is consistent with the latter aspect of melioration. It is difficult to imagine a realistic scenario in which high frequencies of *sda* play in the first stage are due to lack of self-control.

[14] As was mentioned in Section 2.1, the equilibrium probability of *brcr* in $G(.34)$ depends on Player 1 play in the first stage. Taking the observed values of $P(sda_1)$ into account, the equilibrium predictions are .800, .769, .770, .772, for blocks 1 through 4, respectively, and .778 for the entire $G(.34)$ cell.

TABLE III

EXPERIMENTAL RELATIVE FREQUENCIES

		Block 1	Block 2	Block 3	Block 4	All Rounds
$G(.50)$	$P(sda_1)$	.845**	.852**	.797**	.741**	.808**
	$P(brcr)$	.759	.886	.917	.914	.869
$G(.34)$	$P(sda_1)$	.872**	.819**	.800**	.781**	.818**
	$P(brcr)$	.684**	.794	.728	.179	.749*

*: significantly different from equilibrium at 5% level.
**: significantly different from equilibrium at 0.1% level.

directions of changes between blocks are consistent with the learning models, though more so for the $G(.50)$ cell than the $G(.34)$ cell.

In order to quantify this apparent counterclockwise movement and detect convergence, if any, of play to equilibrium, we disaggregate the data by pairs of players and change the coordinate system from standard rectangular coordinates to polar coordinates. The location of a point is represented by the ordered pair (r, θ), where r is the distance from that point to the origin (here, the Nash equilibrium point or center of the segment of equilibria), and $\theta \in [0, 2\pi)$ is the measure of the angle the point makes with the origin and some fixed reference ray (here, the ray pointing straight down from the origin). The values for r and θ are given by

$$r = \sqrt{\left(P(sda_1) - P_{equilibrium}(sda_1)\right)^2 + \left(P(brcr) - P_{equilibrium}(brcr)\right)^2};$$

$$\cos\theta = \frac{.75 - P(brcr)}{r}.$$

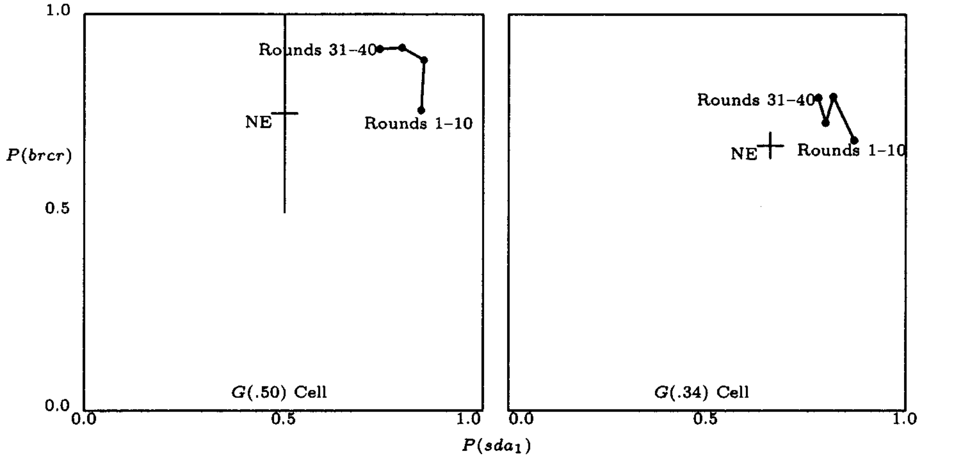

FIGURE 3.—Aggregate subject behavior in experiment (all players, 10-round blocks).

For each pair of players, *increases* in θ indicate movement in the counterclockwise direction around the equilibrium point and *decreases* in r indicate movement toward equilibrium. Wilcoxon signed ranks tests of changes in r and θ from the first to the last block in $G(.50)$ find a significant increase in the population distribution of θ ($p < .001$) but no significant change in the population distribution of r ($p > .10$). Page tests of changes in r and θ over all four blocks yield the same result in all cases. Wilcoxon tests of changes in r and θ from the first to the last block in $G(.34)$ find a significant *increase* in the population distribution of r ($p < .01$), as well as a significant increase in the population distribution of θ ($p < .001$). Page tests of changes in $G(.34)$ yield somewhat weaker results; the increase in r is not significant, while the increase in θ is only significant at the 10% level. We conclude that individual $(P(sda_1), P(brcr))$ pairs move counterclockwise over time in both games, but there is no evidence of movement toward equilibrium.[15]

4. COMPARISON BETWEEN EXPERIMENTAL RESULTS AND LEARNING MODEL PREDICTIONS

The results mentioned in the previous section suggest the experimental data possess features that are inconsistent with static Nash equilibrium, but consistent with the gross qualitative predictions of the learning models.[16] We now wish to compare the degree of success of these models. In the next section, we test the individual-level predictions of the models by comparing the decisions made by players in particular situations with the decisions predicted by the models in those situations. In the following section, we test the aggregate-level predictions of the models by running additional simulations in which initial behavior is similar to that in the experiment, and seeing how closely simulation trajectories track observed experimental trajectories.

4.1. *Use of the Learning Models for Characterizing Individual Decision-Making Behavior*

One way to look at learning models is as forecast rules that, given information from previous rounds, predict (possibly probabilistically) a subject's choices in the current round. In this section, we will examine the accuracy of the forecasts of the RE_0 and BA models, as well as stationary equilibrium, using as baselines

[15] There is no inconsistency in this difference between aggregate results and individual-level results. Changes in individual play paths may not show up when aggregated, particularly if there is a lot of heterogeneity in the individual play paths. As a simple example, consider two pairs of players. The first pair's $(P(sda_1), P(brcr))$ pair is $(.7, .7)$ in the first block and $(.5, .8)$ in the second, and the second pair's $(P(sda_1), P(brcr))$ pair is $(.3, .8)$ in the first block and $(.5, .7)$ in the second. The $(P(sda_1), P(brcr))$ pair for each pair of players moves counterclockwise about $(.50, .75)$ as well as toward it, but aggregating the pairs gives stationary play of $(.50, .75)$ "on average."

[16] Feltovich (1999) gives some evidence that the RE_0 model outperforms static Nash equilibrium in characterizing the experimental data.

three inertial models. The inertial models predict that players will behave in the current round exactly the same as in the previous round with probability p_{same}. For example, a Player 1 who chose the *sda* in the first stage in the previous round will choose it in the first stage of this round with probability p_{same}, irrespective of which payoff matrix was chosen by nature in either round. We denote this class of inertial models by IN(p_{same}); the members we examine are the IN(0.50), IN(0.75), and IN(1.00) models (IN(0.50) predicts that both actions are played with equal probability at each reached information set and could thus be considered "completely random" play).

Because predictions of early-round play according to the RE_0 and BA models depend heavily on unknown initial conditions (propensities or beliefs), we look only at the models' predictions of behavior in the last thirty rounds. In the case of the RE_0 model, we assume that the propensity for playing action α at information set Ψ in round t is exactly equal to the sum of payoffs received in rounds up to $t-1$ in which Ψ was reached and α was played; in the case of the BA models, we assume that the belief weight on an opponent's playing action α at information set Ψ is equal to the number of times Ψ was reached and α was played (for models with $\delta = 0$) or the weighted sum of times Ψ was reached and α was played (for models with $\delta \neq 0$). In other words, initial propensities or initial beliefs have been completely drowned out.[17] Given their propensities or beliefs, players' predicted probabilities are obtained as discussed in Section 2.

We first assess the accuracy of our models using three measures of closeness of predictions to actual choices: mean squared deviation (MSD), log likelihood ($\ln(L)$), and a proportion of inaccuracy (POI) score. All three criteria are derived by pairing the predicted probability of A being chosen (denoted $p_{pred}(A)$) according to the behavioral model being considered and the actual probability that A was chosen (denoted $p_{act}(A)$)—which is either zero or one—for each choice made by either type of player at any information set. Then

$$MSD = \left(\frac{1}{N} \sum \left[p_{pred}(A) - p_{act}(A) \right]^2 \right)^{1/2},$$

$$\ln(L) = \sum_{A \text{ chosen}} \ln\big(p_{pred}(A)\big) + \sum_{B \text{ chosen}} \ln\big(1 - p_{pred}(A)\big),$$

where N is the total number of observations. The MSD statistic used here is equivalent to the "quadratic scoring rule," whose desirable theoretical properties are examined by Selten (1998). It is also very similar to the "mean probability score" discussed by Yates (1982) and used later in this section. The POI score is meant to put models with deterministic predictions on the same

[17] If, contrary to this assumption, initial propensities or beliefs have not been drowned out at this time, there should be substantial differences between the analysis performed here and the analysis that would have been performed had we used the last twenty rounds or the last ten (in which case initial propensities or beliefs would have a noticeably smaller effect), rather than the last thirty. This turns out not to be the case; using the last ten or twenty rounds does not change the "ranking" of the various behavioral models. Some evidence of this will be given shortly.

TABLE IV

ABILITIES OF BEHAVIORAL MODELS TO PREDICT DISAGGREGATED DECISIONS
ROUNDS 11–40 (21–40)—$N = 1830$ (1220)

Model	*MSD*	*POI*	$\ln(L)$	Posterior Prob.
RE_0	.348 (.350)	.182 (.186)	-3161.3 (-2037.9)	> 0.999 (> 0.999)
BA(∞, 0)	.548 (.553)	.302 (.307)	$-\infty$ ($-\infty$)	0.000 (0.000)
BA(4.44, -0.105)	.424 (.423)	.264 (.266)	-3969.6 (-2627.7)	< 0.001 (< 0.001)
BA(0.19, -0.205)	.492 (.493)	.236 (.236)	-4963.9 (-3313.2)	< 0.001 (< 0.001)
equilibrium	.414 (.414)	.299 (.299)	$-\infty$ ($-\infty$)	0.000 (0.000)
IN(0.50)	.500 (.500)	.500 (.500)	-5073.8 (-3382.6)	< 0.001 (< 0.001)
IN(0.75)	.426 (.428)	.239 (.241)	-4025.1 (-2695.9)	< 0.001 (< 0.001)
IN(1.00)	.489 (.491)	.239 (.241)	$-\infty$ ($-\infty$)	0.000 (0.000)

footing as those with stochastic predictions; it treats the highest-probability choice according to a model as "the" prediction of the model and determines the proportion of wrong predictions. The *POI* score is found by calculating the mean of the following values over all choices: when action α (either A or B) is chosen, the value 0 if $p_{pred}(\alpha) > 1/2$, the value $1/2$ if $p_{pred}(\alpha) = 1/2$, or the value 1 if $p_{pred}(\alpha) < 1/2$. For a deterministic model such as fictitious play or the IN(1.00) model, the *POI* score is exactly equal to the square of the model's *MSD* score.

Table IV summarizes the predictive abilities of the models. In addition to RE_0, Nash equilibrium, fictitious play (BA(∞, 0)), and the three inertial models, we show the *MSD*, *POI*, and $\ln(L)$ of the "best" BA models in terms of each of these criteria.[18] Keeping in mind that better predictive power is implied by lower *MSD* and *POI* and by higher $\ln(L)$, we can see that RE_0 is best according to all three criteria. Fictitious play fares particularly badly here—worse than fifty-fifty randomization according to the *MSD* criterion. Even the best of the BA models performs better than the IN(0.75) model according to only two of the three criteria, and is far worse than RE_0. We also see that while Nash equilibrium performed poorly in describing many features of the data, it actually does reasonably well here. In fact, according to the *MSD* criterion, RE_0 is the only model that performs better than stationary Nash equilibrium play. Shown in parentheses are corresponding statistics that cover only the last twenty rounds, rather than the last thirty. Using this smaller set of rounds does not change any of our conclusions.

[18] The best model, according to a given criterion, was found by a grid search over values of λ and δ to three significant digits. In order to give the BA model the best possible chance, different parameter values were allowed each time the criterion of goodness of fit was changed. The best model according to the $\ln(L)$ criterion was almost the same as the best one according to *MSD*, not only in the parameter space but also in their values for the three criteria, so the best model according to $\ln(L)$ is reported here as their common optimizer. A sensitivity analysis suggests that at the optima, all three statistics are robust to small changes in both λ and δ, so adding significant digits will not lead to substantially better values.

Because the models used here are not nested, we cannot use a straightforward likelihood-ratio test to compare them. However, we can use the very similar "minimal prior information" posterior odds criterion that was developed by Klein and Brown (1984) and used by Harless and Camerer (1994) to compare models of decision making. The posterior odds criterion for some Model 1 versus some other Model 2 is

$$[n^{-(K_1-K_2)/2}][\text{Maximized Likelihood under Model 1}/\text{Maximized Likelihood under Model 2}],$$

where n is the sample size and K_1 and K_2 are the number of free parameters in Model 1 and Model 2, respectively. (The BA model has two free parameters; the other models have none.) Given these pairwise odds, we can calculate the posterior probability of each model being correct, given that one of these models is correct; these are shown in Table IV. Not surprisingly, given the values of the $\ln(L)$ statistic, RE_0 is the "odds-on" favorite. Of course, there is always the possibility that none of the given models is the "correct" one, so we cannot conclude that RE_0 is the correct model, only that the other models are (with extremely high probability) incorrect.

In order to illustrate the predictive ability of the models, we also show reliability diagrams for each model.[19] A reliability diagram is a graphical representation of the predictive ability (sometimes called *external correspondence*) of a model, and depicts its decomposition into two distinct components, calibration and resolution. *Calibration* is a measure of the "accuracy" of a model's forecasts; in a well calibrated model, predicted probabilities conform closely on average to actual relative frequencies. (For example, if one looks at the situations in which a well calibrated model predicted that the probability of choosing A is 0.2, A should actually have been chosen about 20% of the time.) *Resolution* is a measure of the "informativeness" of a model's forecasts; a well resolved model partitions the set of predicted/actual choices into subsets in which A is actually chosen either almost never or almost always. Calibration and resolution are both desirable properties for a model to have, but it is possible for a model to be well calibrated but poorly resolved, well resolved but poorly calibrated, good at both, or poor at both (hypothetical examples are shown in Figure 4). For example, since there are only two possible predictions at any information set, a model that *always* makes the wrong prediction is very poorly calibrated but very well resolved. The forecasts from an ideal behavioral model would be both perfectly calibrated and perfectly resolved; the model would predict A (with probability one) every time A is actually chosen, and B every time B is chosen. For probabilistic models such as RE_0 and the BA models with finite λ, as well as for mixed-strategy Nash equilibrium, this is generally not possible. However, we will examine the extent to which the models approach this ideal.

[19] A thorough review of many of the concepts used here, particularly the notions of calibration and resolution, is given by Yates (1982).

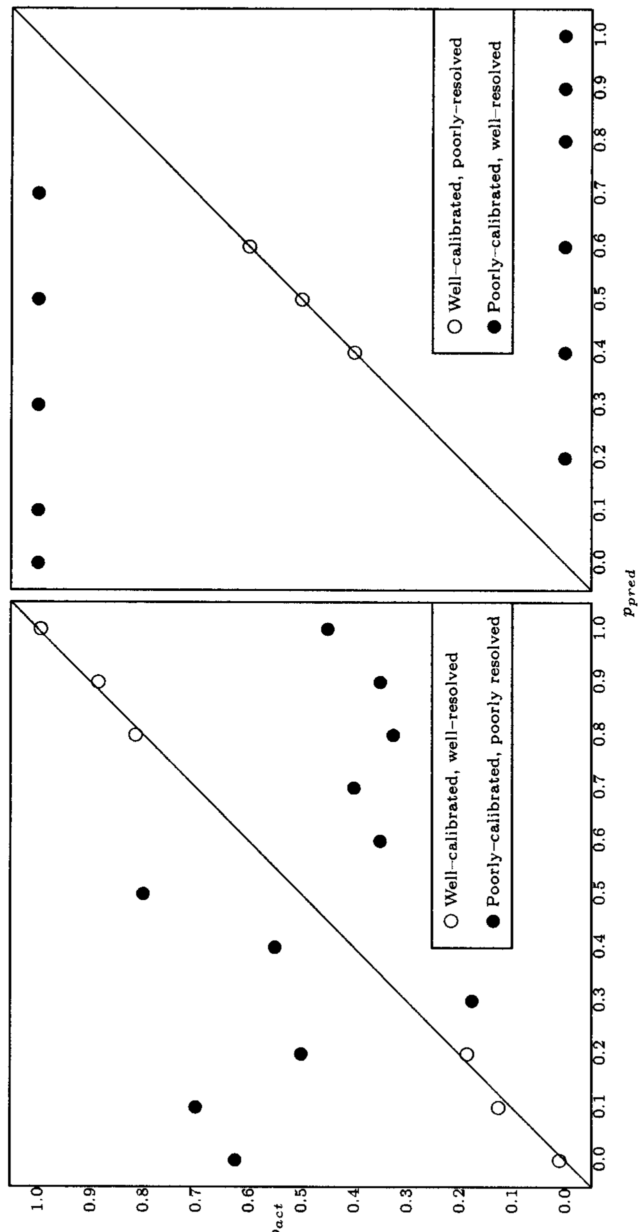

FIGURE 4.—Examples of calibration graphs.

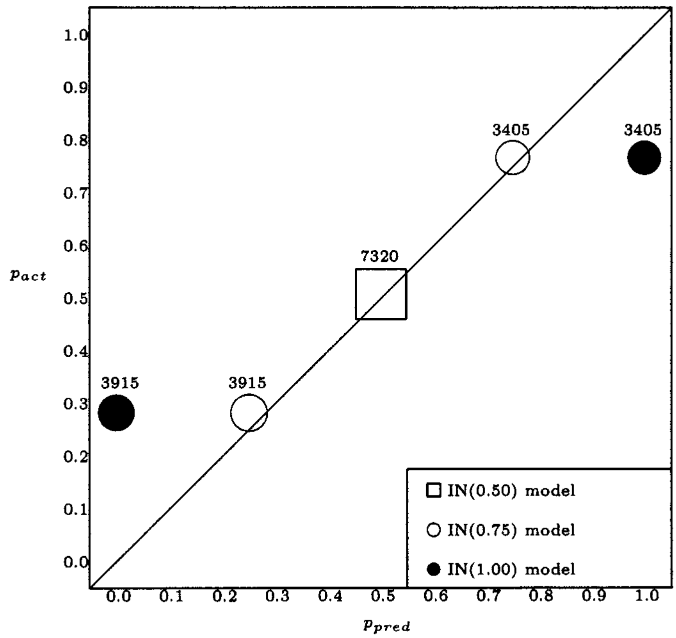

FIGURE 5.—Calibration graphs of inertial models.

The reliability diagrams are constructed as follows: first, actual and predicted decisions over rounds 11–40 are paired; second, each decision pair is classified according to its predicted probability of choosing A into one of eleven intervals ($[0, 0.05], (0.05, 0.15], (0.15, 0.25], \ldots, (0.95, 1]$); third, the mean predicted probability of an A choice (denoted $\bar{p}_{pred}(A)$) and the actual relative frequency of A choices (denoted $\bar{p}_{act}(A)$) are calculated for the pairs in each interval, yielding up to eleven ordered pairs; fourth, these ordered pairs are graphed in $(P(sda_1), P(brcr))$ space, along with the number of observations represented by each ordered pair. A perfectly calibrated model would yield ordered pairs on the 45° line. A perfectly resolved model would yield ordered pairs on the $\bar{p}_{act} = 0$ and $\bar{p}_{act} = 1$ lines (see Figure 4).

The models' reliability diagrams are shown in Figures 5 and 6. We can see the high degree of calibration of the IN(0.50) and IN(0.75) models, though both are very poorly resolved. The IN(1.00) model is both poorly calibrated and poorly resolved. The RE_0 and BA models, as well as equilibrium, are well calibrated in the very weak sense that $\bar{p}_{pred}$ and $\bar{p}_{act}$ are positively correlated. However, RE_0 and the "best" BA model are clearly better calibrated than fictitious play or equilibrium play (though it should be noted that, as poorly as equilibrium characterized other aspects of subject behavior, it performs rather well here).

It is difficult to tell from the reliability diagrams whether RE_0 or the best BA model is better calibrated, but based on the diagrams, we can quantitatively examine the success of each behavioral model using the following three criteria:

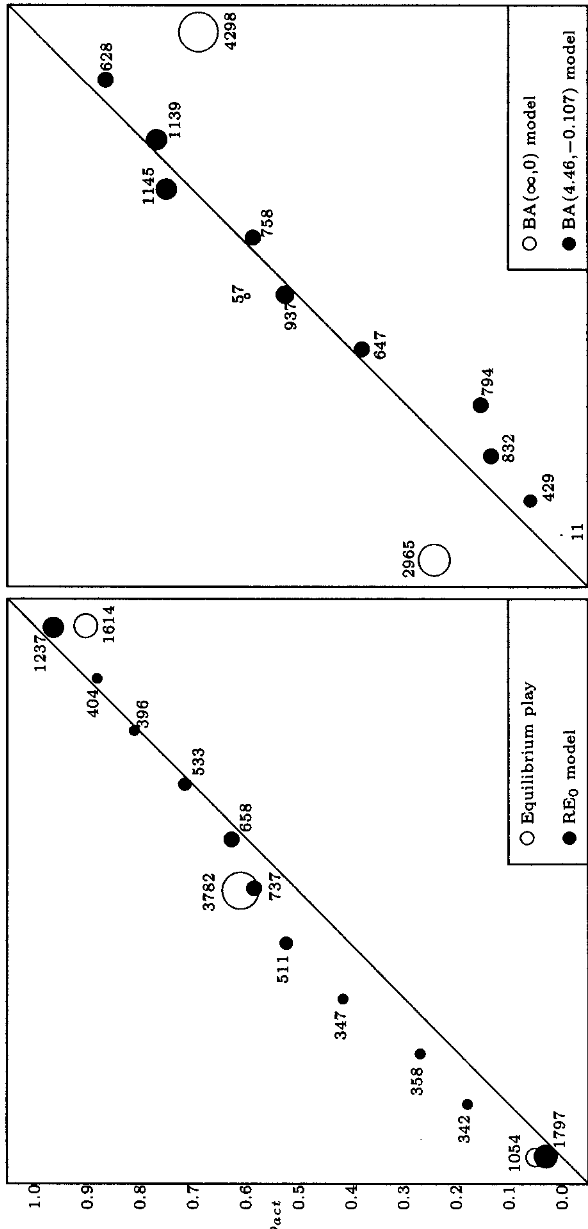

FIGURE 6.—Calibration graphs of equilibrium and learning models.

Sanders calibration (C_s), Sanders resolution (R_s), and mean probability score ($\overline{PS}$), measures of calibration, resolution, and overall predictive ability, respectively. These measures are defined as follows:

$$C_s = \left(\frac{1}{N}\right) \sum_{j=1}^{11} N_j \left(\bar{p}_{pred}(A) - \bar{p}_{act}(A) \right)^2,$$

$$R_s = \left(\frac{1}{N}\right) \sum_{j=1}^{11} N_j \bar{p}_{act}(A)(1 - \bar{p}_{act}(A)),$$

$$\overline{PS} = C_s + R_s,$$

where N_j is the number of pairs whose predicted probability lies in the jth interval, and N is the total number of pairs.[20] Table V reports the performance of the models according to these criteria. According to calibration scores alone, the best models are BA(0.19, −0.205) and fifty-fifty randomization; however, they are the worst resolved. The difficulty in ascertaining from the diagrams whether RE_0 or the best BA model is better calibrated is apparent here; RE_0 is slightly better than BA(4.46, −0.107), but the difference between them is negligible. The RE_0 model has by far the best resolution, and while the best BA model has better calibration than equilibrium, its resolution is worse.

4.2. *Use of Initialized Models for Tracking Aggregate Play*

It was shown in the previous section that the RE_0 model predicts individual decisions, given histories of play up to the current round, better than the other models we considered. While this provides evidence that individuals' decision-making processes (in this game) might be usefully approximated by a reinforce-

TABLE V

CALIBRATION AND RESOLUTION OF MODELS
ROUNDS 11–40—N = 1830

Model	C_s	R_s	$\overline{PS}$
RE_0	.0034	.1329	.1362
BA(∞, 0)	.0799	.2015	.2815
BA(4.44, −0.105)	.0037	.1750	.1788
BA(0.19, −0.205)	.0000	.2499	.2500
BA(4.46, −0.107)	.0035	.1750	.1786
Equilibrium	.0050	.1726	.1775
IN(0.50)	.0001	.2499	.2500
IN(0.75)	.0007	.1917	.1923
IN(1.00)	.0629	.1917	.2546

[20] It can be shown that $\overline{PS}$ is approximately the square of *MSD*, so we already know that RE_0 will have the best $\overline{PS}$ score. However, it is still instructive to see if it is best calibrated, best resolved, or both.

ment-based model, such predictions are not the only use of learning models. Also useful are predictions of aggregate behavior; a model that performs poorly in predicting individual decisions may work well on average. Furthermore, we have been looking at models' predictions of play in a given round *given what has happened from the first round up to that round*. Since a researcher will not have this information until after the experiment is performed, it would be useful for a model to make accurate predictions of behavior in *all* rounds, given only some initial conditions. We will now analyze the success of the models in making these types of predictions.

In order to accomplish this, we have run additional sets of simulations similar to those in Section 2.5, but different in two notable ways. First, the output of the new simulations consists of observed actions averaged over blocks, instead of behavioral strategies from particular rounds, so that the simulation results are more directly comparable to the experimental data reported in Section 3. Second, instead of using randomly chosen first-round propensities or beliefs, we initialize propensities or beliefs so that first-round play in the simulations matches first-round play in the experiment as closely as possible. For the RE_0 simulations, we use the actual first-round relative frequencies from the experiment as estimates of initial probabilities. For the BA simulations, we set initial belief probabilities to obtain play as close as possible to the actual first-round relative frequencies as the outcome of expected-payoff maximization for the simulations with $\lambda = \infty$ or via Equation 1 in Section 2.4 for the simulations with finite λ. As before, strengths of initial propensities and beliefs were set to six. For each model, we ran 100 sets of 29 simulations of the $G(.50)$ game and 100 sets of 32 simulations of the $G(.34)$ game, 100 times the actual number of observations in the experiment.

Figures 7 and 8 plot, for the RE_0 and BA(6.67, −0.030) simulations, respectively, mean relative frequencies of sda_1 and *brcr* over each ten-round block. (The BA(6.67, −0.030) model is the best-fitting according to the criterion that will be used in this section.) Each small circle represents the $(P(sda_1), P(brcr))$ pair corresponding to the average play of a set of pairs of *simulated* players; each large circle represents the $(P(sda_1), P(brcr))$ pair corresponding to the average play of all pairs of *actual* players over the same block. (In these figures, the radii of the circles are not meant to indicate any aspect of the data; they are meant only to make the one experimental point stand out among the 100 simulation points.) Large pluses show the Nash equilibria, and in the figures corresponding to the BA model, small pluses show the logistic response equilibrium.

None of the simulations exactly match the experimental data. The RE_0 simulation means and those of the BA(6.67, −0.030) model are clustered near the corresponding experimental means in each block, though the former seems to be somewhat closer to the $G(.50)$ data and the latter to the $G(.34)$ data. This closeness is quantified in Table VI, which reports the squared difference between simulation and experimental means for both games, both players, and each of the four blocks. In the bottom row of the table, the sixteen squared

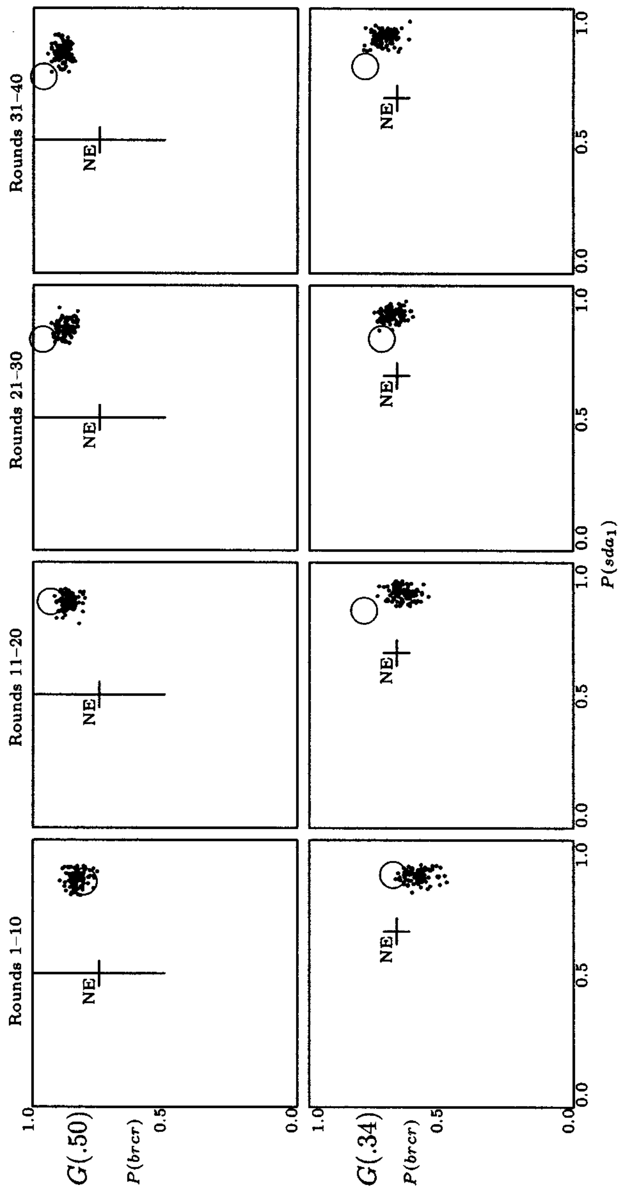

FIGURE 7.—Experimental data (circles) and initialized RE_0 simulations (dots).

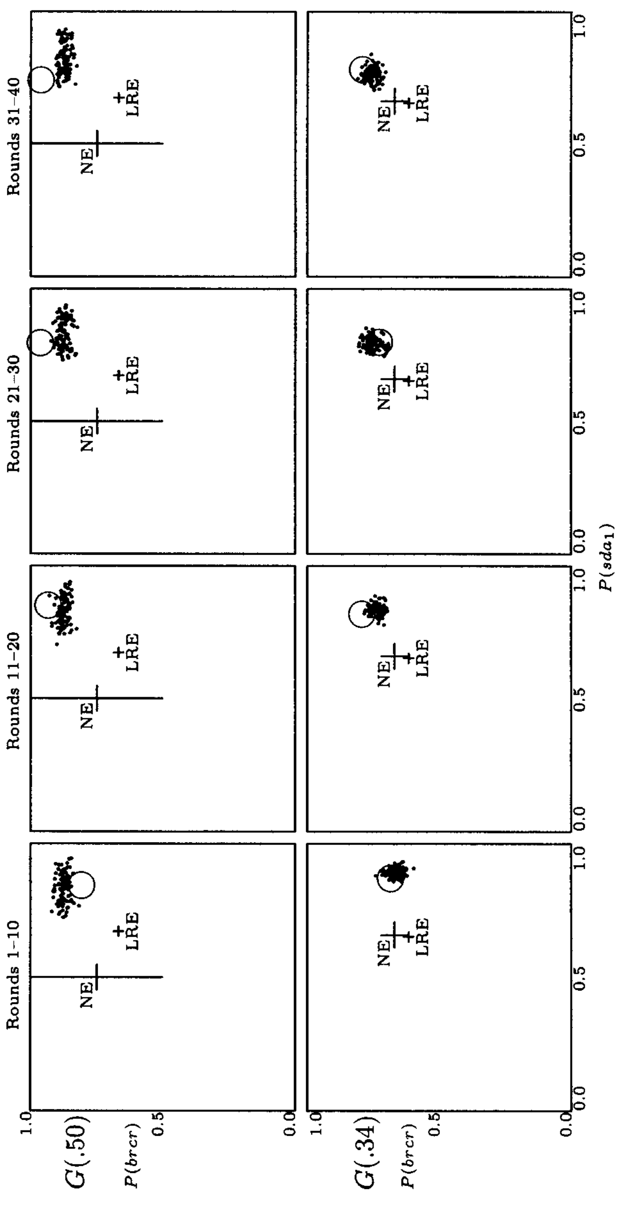

FIGURE 8.—Experimental data (circles) and initialized BA(6.67, −0.030) simulations (dots).

TABLE VI

FIT OF MODEL SIMULATIONS TO EXPERIMENTAL MEANS (SQUARED DEVIATIONS)

Game	Variable	Block	RE_0	BA models $(\infty, 0)$	BA models $(4.46, -0.107)$	BA models $(6.67, -0.030)$	IN(0.50)	Equil. i.i.d.
$G(.50)$	$P(sda_1)$	1	.00006	.00002	.00020	.00032	.00639	.11902
		2	.00001	.02826	.00067	.00001	.12390	.12390
		3	.00179	.05471	.00035	.00002	.08821	.08821
		4	.00781	.05724	.00425	.00015	.05808	.05808
	$P(brcr)$	1	.00602	.00407	.00063	.01156	.06708	.00000
		2	.00038	.00062	.01069	.00000	.14900	.00000
		3	.00108	.00000	.01831	.00128	.17389	.00000
		4	.00028	.00049	.01831	.00156	.17140	.00000
$G(.34)$	$P(sda_1)$	1	.00001	.00043	.00432	.00036	.13838	.04494
		2	.00462	.00571	.00038	.00040	.10176	.02528
		3	.00845	.03047	.00007	.00000	.09000	.01960
		4	.01249	.03532	.00004	.00032	.07896	.01464
	$P(brcr)$	1	.00837	.00001	.00763	.00026	.03386	.00020
		2	.02043	.00722	.02565	.00317	.08644	.01538
		3	.00163	.01506	.00871	.00073	.05198	.00336
		4	.00601	.09992	.02289	.00142	.08468	.01464
Sum of Squared Deviations			0.07945	0.33955	0.12310	0.02762	1.6166	0.52726

differences are summed to produce a single measure of a model's closeness to actual play. For comparison, we also report squared differences between the experimental means and the means implied by stationary equilibrium play, fictitious play, the BA$(4.46, -0.107)$ model (the best BA model according to $\ln(L)$), and the IN(0.50) model.[21]

Keeping in mind that lower numbers imply simulated play closer to actual play, we see that the worst models by far were Nash equilibrium play and the IN(0.50) model. Even fictitious play performs better than these models, though it is also far from actual play. The RE_0 model and the BA model that was best at describing individual decisions (according to $\ln(L)$) are much closer, though the former is somewhat better than the latter. However, the BA$(6.67, -0.030)$ model is much closer to the actual data than even the RE_0 model.

Since there exist values of λ and δ that make the BA model better able to track the experimental data than RE_0, one may be tempted to conclude that the BA model is better than the RE_0 model. Two caveats apply, though. First, which model is "better" depends on what criteria of "goodness" are being used; as was shown earlier, RE_0 is better than *all* of our BA models when the criteria are those used in the Section 4.1 (the BA$(6.67, -0030)$ model has a *MSD* of .434, a

[21] Here, we are only comparing simulation and experimental means, with no regard paid to variances. This is intentional; our interest here is in determining the accuracy of predictions of experimental means using simulation means, not in determining which model would have been "most likely" to produce the observed data (in which case a measure of dispersion would have been important).

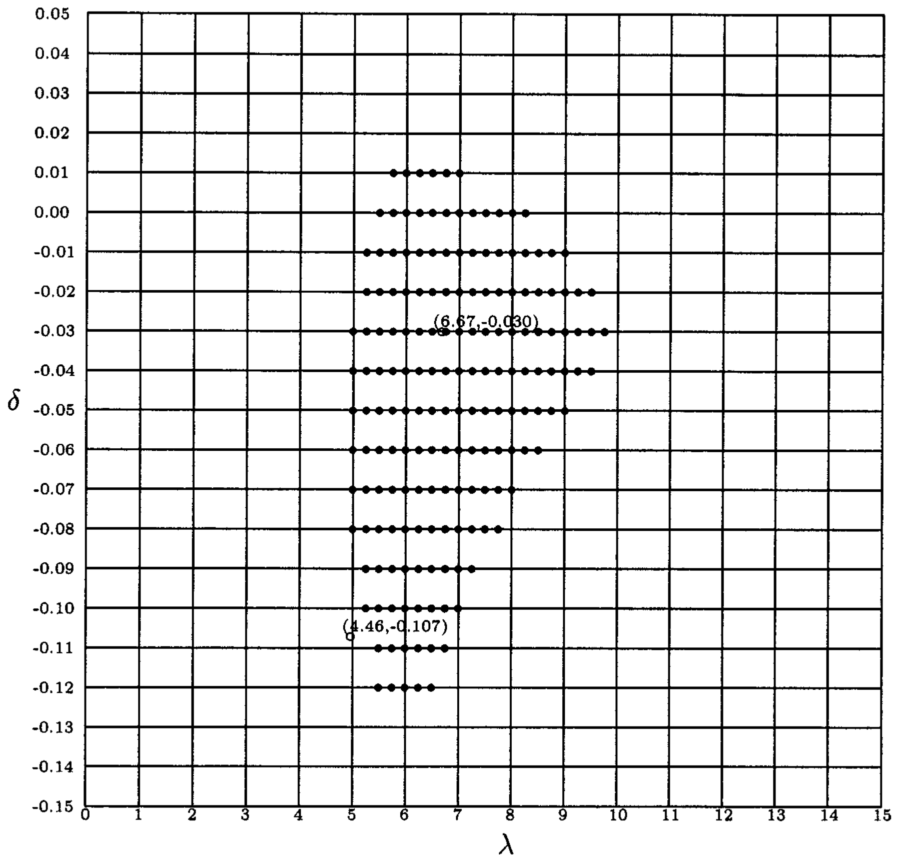

FIGURE 9.—Parameter values for which the BA model outperforms the RE_0 model.

POI of .280, a $\ln(L)$ of -4162.4, a C_s of .0061, a R_s of .1768, and a $\overline{PS}$ of .1830, all worse than those of RE_0). Second, even if the criterion used is the one from this section, there are also values of λ and δ that make the BA model *less* able to track the experimental data than RE_0. Before running the experiment, there was no way of knowing which parameter values should have been chosen.

It is therefore of some value to know how likely we would have been to have chosen parameter values that gave predictions better than those of RE_0. Figure 9 shows the (λ, δ) pairs that make the BA model better according to the criterion used in this section. We can see that while it would certainly have been possible to luckily guess good parameter values, it would by no means have been a sure bet. In particular, if one supposed a positive value for δ—in forming beliefs, players put more weight on recent opponent actions than on earlier ones—then it would have been unlikely indeed that a better model than RE_0 would have been selected, and the improvement over RE_0 would have been negligible (the best model with positive δ, the BA(6.50, 0.01) model, yields a sum of squared deviations of .07252, only slightly better than RE_0).[22]

This is not to say that the BA model is not appropriate as a model of learning in this experiment, only that at this time, RE_0 is more useful for generating quantitative predictions ex ante (as opposed to fitting play ex post). Before the experiment, we had no idea which BA model was appropriate, and the results of this, and the previous, section suggest that an "uneducated guess" of parameter values would have produced predictions worse than those of RE_0. There is currently no intuition available regarding how to make an "educated guess." Further research may accomplish this; eventually it may be possible to know, before the experiment is performed, which BA model to use to generate predictions. The simplest way this could occur would be if one small subset of the BA parameter space always produced suitable models (the results of the next section imply that this is unlikely). More generally, it might be the case that there exists a mapping from "types" of game (and possibly experimental procedures) to parameter combinations, such that given a strategic environment, one could choose appropriate parameters. If it is eventually determined that such a correspondence exists, the BA model will turn out to be very useful indeed. But until this is determined, a simple model like RE_0 will continue to be functional not only as a model of individual choice, but also as a model of aggregate behavior.

5. SOME EVIDENCE FROM OTHER EXPERIMENTS

We now look at the abilities of the RE_0 and BA models to characterize data from other experiments. Using data from other experiments allows us to determine how generalizable our results are; that is, whether RE_0 is best for a large class of games, or whether the game used here happened to be a lucky choice. As mentioned in the introduction, Erev and Roth (1998) and Camerer and Ho (1999) use data from previous experiments to test learning models. Erev and Roth examine the data from six previous experiments and one new experiment. They assess the ability of several models to characterize aggregate features of data and find that in the games they examine, a generalization of the RE_0 model works best. Camerer and Ho examine the data from five previous experiments (all different from those examined by Erev and Roth). They develop a very general learning model that includes as special cases all of the models in this paper, and use likelihood ratios to test the general model versus "restricted" versions of the model such as reinforcement learning and fictitious

[22] Furthermore, we should remember that we have only used one particular reinforcement-based model throughout this paper. As mentioned previously, Roth and Erev (1995) and Erev and Roth (1998) consider several variations of the RE_0 model. We use only one here because it does such a good job that it seems that the small improvement in fit that comes from adding free parameters would not be worth the reduced generality. However, if it is claimed that one should try all versions of BA models in order to give it the best possible chance, an argument could be made that one should do this for the reinforcement-based model, too.

play. In almost all cases, they find that the best (maximum-likelihood) version of the model is quite different from the restricted versions, and that the former is significantly better than any of the latter.

We will look at four data sets, each of which comes from an experiment involving a simultaneous-move game. Two were used by Erev and Roth: Erev and Roth's (1998) "Matching Pennies" game and Ochs's (1995) 2×2 games. The other two were used by Camerer and Ho: Mookherjee and Sopher's (1997) 6×6 constant-sum games, and Van Huyck, Battalio, and Beil's (1991) median-action game. As we did in Section 4.1 with the asymmetric-information game data, we look at the models' abilities to predict individual decisions. As before, we attempt to limit the reliance of the models on unobservable initial propensities or beliefs by not comparing predictions of behavior with actual play in early rounds. For each data set, we compare model predictions with play from round 11 through the last round of play, with the exception of the data from van Huyck et al., whose experiment only lasted ten rounds; we consider rounds 3 through 10 of this data set. The data we examine range from 75% to 96% of each experiment. For comparison, and in order to examine the plausibility of our assumption that initial propensities or beliefs are drowned out by the time we reach the rounds at which we look, we also consider the data from a smaller subset of rounds of each experiment (roughly the last 50% of each experiment).

As a basis for comparison, in addition to the RE_0 and BA models, we consider stationary equilibrium (if the equilibrium is unique), the IN(0.75) model in which players repeat their previous action with probability 0.75 and play each other action with equal share of the remaining 0.25 probability, and the $IN(1/m)$ model in which players play each of the m possible actions with equal probability. For each of these models, we report in Table VII the *MSD*, *POI*, and $\ln(L)$ scores, as well as the posterior probability of each model (again, given that one of these models is the correct one) and the rounds considered.

According to both *MSD* and $\ln(L)$ (and thus posterior probability), RE_0 is the best for characterizing Erev and Roth's data, but it is only slightly better than the IN(0.75) model according to *MSD*, and worse according to *POI*. Even the best BA model fares worse than RE_0 according to all three criteria, but it nonetheless improves upon the stationary equilibrium prediction.[23] None of these results changes when we look at rounds 251–500 rather than 11–500.

Ochs's data present a somewhat weaker case for the RE_0 model; according to *MSD*, it outperforms the rest of the models, and the best BA model is even worse than the IN(0.75) model. On the other hand, according to *POI*, RE_0 is worse than BA and only slightly better than IN(0.75). According to $\ln(L)$, RE_0 is worse than BA and both inertial models; as with *POI*, the BA model outper-

[23] The best BA model is determined separately for each data set, time frame, and criterion, again using a grid search. The optimal values of λ and δ vary greatly between experiments and between criteria; and vary somewhat between time frames. For example, using the longer time frame and the $\ln(L)$ criterion, the best BA model is $BA(0.791, -0.133)$ for the Erev and Roth data set, $BA(0.612, 0.236)$ for the Ochs data set, $BA(31.0, -27.6)$ for the van Huyck et al. data set, and $BA(0.556, 0.153)$ for the Mookherjee and Sopher data set.

TABLE VII

ABILITIES OF MODELS TO PREDICT DISAGGREGATED DECISIONS (PREVIOUS EXPERIMENTS)

Experiment	Model	*MSD*	*POI*	ln(*L*)	Posterior Prob.
ER: Matching pennies	RE_0	.474 (.459)	.361 (.333)	−6370.3 (−3043.7)	> 0.999 (> 0.999)
	BA	.496 (.495)	.441 (.428)	−6715.9 (−3411.0)	< 0.001 (< 0.001)
	equilibrium	.500 (.500)	.500 (.500)	−6792.8 (−3465.7)	< 0.001 (< 0.001)
Rounds 11(251)−500	IN(0.50)	.500 (.500)	.500 (.500)	−6792.8 (−3465.7)	< 0.001 (< 0.001)
$N = 4900(2500)$	IN(0.75)	.482 (.473)	.340 (.322)	−6474.4 (−3206.1)	< 0.001 (< 0.001)
Ochs: three 2 × 2 games	RE_0	.335 (.333)	.271 (.261)	−7490.4 (−4045.7)	< 0.001 (< 0.001)
	BA	.370 (.380)	.260 (.265)	−4344.4 (−2508.9)	> 0.999 (> 0.999)
	equilibrium	.379 (.380)	.393 (.393)	−10441.8 (−5628.3)	< 0.001 (< 0.001)
Rounds 11(33)−56 or 64	IN(0.50)	.390 (.397)	.500 (.500)	−7674.4 (−4219.2)	< 0.001 (< 0.001)
$N = 2464(1408)$	IN(0.75)	.350 (.348)	.273 (.260)	−7472.7 (−4113.6)	< 0.001 (< 0.001)
vHBB: three 7 × 7 games	RE_0	.188 (.145)	.134 (.070)	−598.8 (−190.5)	< 0.001 (< 0.001)
	BA	.154 (.112)	.090 (.046)	−526.0 (−130.9)	> 0.999 (> 0.999)
	equilibrium	—	—	—	—
Rounds 3(6)−10	IN(1/7)	.350 (.350)	.979 (.991)	−1681.3 (−1050.8)	< 0.001 (< 0.001)
$N = 864(540)$	IN(0.75)	.190 (.147)	.126 (.056)	−563.6 (−242.1)	< 0.001 (< 0.001)
MS: two 6 × 6 games	RE_0	.369 (.372)	6.70 (.691)	−2896.1 (−1919.8)	< 0.001 (< 0.001)
	BA	.372 (.372)	.771 (.762)	−2144.3 (−1429.3)	0.269 (0.079)
	equilibrium	.355 (.357)	.622 (.634)	−∞ (−∞)	0.000 (0.000)
Rounds 11(21)−40	IN(1/6)	.373 (.373)	.833 (.833)	−2150.1 (−1433.4)	0.731 (0.921)
$N = 1200(800)$	IN(0.75)	.422 (.429)	.708 (.734)	−2647.1 (−1819.8)	< 0.001 (< 0.001)

Note: Statistics outside (inside) parentheses correspond to round numbers outside (in) parentheses.

forms all other models. Looking at only rounds 33 and afterward does not change the *MSD* or ln(L) results, but it does eliminate the BA model's advantage in *POI* (the IN(0.75) model is now slightly better).

The data of Van Huyck, Battalio, and Beil present even stronger evidence against RE_0; according to all three criteria, the best BA model is better than RE_0 (and every other model), and in all cases the difference is substantial. The RE_0 model performs even more poorly than IN(0.75) (though slightly better according to *MSD*). The BA model is still the best when we look at rounds 6−10 rather than 3−10, but RE_0 is better than IN(0.75) according to ln(L).

The conclusions based on the Mookherjee and Sopher data are ambiguous. According to the ln(L) criterion and hence the posterior probabilities, the best BA model and the two inertial models are far better than RE_0, which is in turn far better than stationary equilibrium; the best model according to ln(L) is the BA model, though the difference between BA and the IN(1/6) model is small enough that the posterior odds actually favor the IN(1/6) model by about 3 to 1. On the other hand, according to the *MSD* and *POI* criteria, stationary equilibrium is best, followed by RE_0, then by the other three. (This game has a dominated strategy which is played occasionally by subjects. Thus, though equilibrium does rather well in terms of *MSD* and *POI*, its ln(L) score is $-\infty$.)

TABLE VIII

CALIBRATION AND RESOLUTION OF MODELS (PREVIOUS EXPERIMENTS)

Experiment	Model	C_s	R_s	$\overline{PS}$
ER: Matching pennies	RE_0	.0017	.0558	.0575
	BA(0.21, −0.19)	.0041	.0583	.0624
	equilibrium	.0042	.0583	.0625
Rounds 11–500	IN(0.50)	.0042	.0583	.0625
$N = 4900$	IN(0.75)	.0051	.0551	.0602
Ochs: three 2 × 2 games	RE_0	.0020	.2082	.2102
	BA(0.60, 0.22)	.0119	.2246	.2365
	equilibrium	.0066	.2349	.2415
Rounds 11–64 or 11(21)–56	IN(0.50)	.0026	.2474	.2500
$N = 2464$	IN(0.75)	.0021	.2198	.2219
vHBB: three 7 × 7 games	RE_0	.0021	.0338	.0359
	BA(31.0, −27.6)	.0001	.0239	.0240
	equilibrium	—	—	—
Rounds 3–10	IN(1/7)	.0000	.1224	.1224
$N = 864$	IN(0.75)	.0026	.0334	.0359
MS: two 6 × 6 games	RE_0	.0090	.1274	.1364
	BA(−0.560, −0.120)	.0004	.1382	.1387
	equilibrium	.0002	.1258	.1259
Rounds 11–40	IN(1/6)	.0000	.1389	.1389
$N = 1200$	IN(0.75)	.0420	.1358	.1778

The *POI* and $\ln(L)$ results are not affected by looking only at rounds 21–40, and the *MSD* results are changed only slightly (RE_0 is now tied with BA), but the IN(1/6) model becomes by far the best according to posterior probability.

The calibration, resolution, and mean probability scores of the models are presented in Table VIII. (Here, only the larger subset of rounds is considered.) These scores motivate conclusions similar to those obtained from the *MSD* scores; according to the Erev and Roth data and the Ochs data, RE_0 is best; according to the van Huyck et al. data, BA is best; and according to the Mookherjee and Sopher data, static equilibrium is best.

We are thus led to the inconvenient conclusion that neither the RE_0 nor the BA model is always the best. Though both the RE_0 and the BA model are usually better able to characterize experimental data than static equilibrium, which of these two models is better depends not only on which set of data is used, but also on which criterion of "goodness" is used. Looking at only the five data sets we consider, there seems to be a weak relationship between the length of the experiment and the ability of the RE_0 model to characterize its data; the longer the experiment, the better RE_0 performs. This is consistent with the results of Erev and Roth (1998), who look exclusively at games of 100 rounds or longer and find that reinforcement-based models do very well (though they use few belief-based models for comparison).[24] This result is also consistent with those of Camerer and Ho (1999), who look at games played no more than 40 times and find that reinforcement models do poorly, and that the only game in which reinforcement models do reasonably well is one in which the game lasted exactly 40 rounds (the Matching Pennies game of Mookherjee and Sopher (1994)). If it is true that the length of an experiment influences the ability of particular learning models to predict behavior, then using only very long experiments (in order to gather a lot of data relatively cheaply) or very short experiments (in order to make parameter estimation problems tractable) to test theories of learning may bias the likely results.

One possible explanation for the better performance of a model like RE_0 in longer games is that individuals have a "toolbox" of learning models available, and choose one by balancing its benefits against its costs. The benefit of choosing a belief-based model over a reinforcement-based model is that the former is likely to yield a higher payoff (by taking into account more information and making strategy choice more sensitive to expected payoff); the cost is the disutility the individual incurs by "thinking hard." In games lasting many rounds, the amount of expected payoff at stake (per round) is relatively small, so that the benefit obtained by choosing the best strategy is also small. Thus, subjects may be more willing to adopt a less cognitively demanding model like RE_0 (which is simple, even compared with the other reinforcement-based models of Erev and Roth). An alternative explanation is that in longer games, the number of rounds is usually much larger than the number of available strategies, so it is less costly for subjects to gather experience in earlier rounds by "experimenting" with various strategies and finding which lead to higher payoffs, rather than using more introspective methods.

The relative abilities of the models also depend on which criterion of goodness is used. According to the *MSD* criterion, RE_0 is better than the BA model in four of the five games, and according to *POI*, it is better in three games and roughly the same in one. On the other hand, according to $\ln(L)$ and posterior probability, the BA model is better in three of the games. When one model is better according to one criterion and worse according to another, which should win out? The answer depends partly on the researcher's opinion of how learning models should be used. The $\ln(L)$ statistic is a measure of how probable a model would be to produce the exact pattern of observed events. The *MSD* statistic is a measure of how closely the probabilistic predictions of a learning model conform to observed events. The *POI* statistic is a measure of how closely the deterministic predictions of a learning model conform to observed events. If it is believed that one of the available learning models is the

[24] Additionally, Roth, Erev, and Slonim (1999) look at a sample of 2×2 constant-sum games, played 500 times, and find that in later rounds, even (stage-game) Nash equilibrium describes play well. Their result is also consistent with reinforcement-based models working well in long games, as both RE_0 and its analogue in their paper (a one-free-parameter model which they call RE1) converge to equilibrium in that class of games.

"correct" model, and the problem is just to pick out the right one, then the $\ln(L)$ statistic (or the closely related posterior probability criterion) is the natural choice. If it is believed that *all* of the models are misspecified, then arguing about which model is the correct one is unjustified; instead, looking at the utility of the models as predictive tools (*MSD* or *POI*) is more productive.[25] In the case of the models we use here, it is extremely unlikely that any of them accurately describes the way individuals actually learn; at best, they are good approximations. (As evidence, consider how much the choice of "best" BA model depended on which data set was used, as mentioned in footnote 23.) Thus, an argument could be made that the *MSD* or *POI* comparisons should be more relevant than $\ln(L)$.

6. CONCLUSION

In order to compare the predictive abilities of several behavioral models, we have run a new experiment using a two-stage constant-sum game of asymmetric information. This game is more complex than the games usually used for learning-model comparisons. We have examined the predictions of stationary Nash equilibrium and of two types of learning model: a class of belief-based models and a reinforcement-based model. While quite different in rationale and in mathematical specification, the two models yield qualitatively similar patterns of behavior. These patterns, implied by the nature of the best-response correspondence, can be intuitively explained in terms of the strategic problems faced by the players in this game, and show that learning models can capture some of the subtleties underlying strategic situations (in this game, having to do with the gradual revelation of private information) that theorists initially learned about by studying equilibrium. The learning models are successful in other ways, too. The qualitative predictions of the learning models pertaining to the evolution of play over time are borne out in the experimental data. In contrast, Nash equilibrium fails convincingly as a predictor of the behavior of experimental subjects playing this game, especially in its aggregate predictions; players start far from equilibrium and do not get closer over time.

The models are not equally successful, however; quantitative aspects of play depend on which model is used. If we view learning models as predictors, given an individual's history of past play, of the probability of his or her choosing a particular action at a particular information set of this game, we must conclude that the RE_0 model is the best of the models we have considered. On the other hand, if we view learning models as predictors, given initial aggregate behavior, of trajectories of aggregate behavior over time, there is no longer an unambiguous choice. For some parameterizations, the BA model works better than the RE_0 model, although the parameter combinations for which this is true make up a small portion of the universe of possible parameter combinations. The comparison between models becomes still more equivocal when a larger set of

[25] This point has been made by Roth, Erev, and Slonim (1999).

games is considered. Even looking at only the predictions of individual behavior, we find that in some cases, RE_0 outperforms the BA model, while in other cases, the reverse is true. Which model is better depends not only on the experimental data set used, but also on the criterion of success.

It is important to remember, however, that equally important to the comparison between the RE_0 and BA models is the comparison between learning models and stationary Nash equilibrium play. Although most of the discussion in this paper has been devoted to the former comparison, this is to some extent only true because the difference between the two learning models is much smaller than the difference between ether of them and equilibrium play. For almost all of the games considered, and according to almost any of the criteria we have used, both RE_0 and the BA model fare better than equilibrium; often the improvement is substantial. Because the difference between the learning models and equilibrium is so much greater than that between the two learning models, it is the former difference that should be expected to carry over to other strategic situations. It may well be that there is no single model that is able to describe behavior well in all situations; much more experimental work is necessary before such a question can be answered. Fortunately, the "best" learning model is not required in order to improve on the predictions of equilibrium play.

Dept. of Economics, University of Houston, Houston, TX 77204-5882, U.S.A.

Manuscript received December, 1997; final revision received May, 1999.

APPENDIX: INSTRUCTIONS USED IN THE EXPERIMENT

Below is an example of the written instructions given to subjects participating in the experiment. The actual instructions were written in WordPerfect 5.1; so they look somewhat different from these due to the way LATEX formats text. Copies of the actual instructions are available from the author upon request.

INSTRUCTIONS

This is a decision-making experiment. You will be playing a game in which the object is to score as many points as possible. If you have any questions at any time, feel free to ask the experimenter.

1. THE PLAYERS

There are two types of players, informed players and uninformed players. Your type is chosen randomly at the beginning of the session and will stay the same throughout the session. You will be randomly assigned an opponent, who is of the opposite type.

You will have the same opponent throughout the session. You will not be told the identity of your opponent—not even at the end of the session.

2. SEQUENCE OF PLAY IN EACH ROUND

The session is made up of many rounds, each divided into 2 stages.

A. The round begins. One of the two payoff tables on page 3 is chosen by the computer, and the informed player is told which table was chosen. There is a 50-50 chance of either table being chosen.

B. Stage 1 begins. You and your opponent each choose an action (either A or B).

C. Stage 1 ends. Each player is told what action his/her opponent chose. The informed player is also told her payoff for the stage.

D. Stage 2 begins. You and your opponent each choose an action (either A or B).

E. Stage 2 ends. Each player is told what action his/her opponent chose. The informed player is also told her payoff for the stage.

F. The round ends. The uninformed player is told which table had been chosen, and both players are told how many points they have earned in the round.

3. THE PAYOFF TABLES

There are two payoff tables, the Left table and the Right table. In each round you and your opponent will be assigned one of these tables. The informed player is told at the beginning of the round which table was chosen; the uninformed player is not told until the end of the round. The table you use may change from round to round, but it will never change during a round. Each player has two possible moves, A and B. In each stage, either you or your opponent will get a point. The point is awarded based on the table you have been assigned, your move, and your opponent's move.

Example 1: Suppose that the Left table was chosen, and that both players choose B. Then the informed player knows that she gets 0 points in this stage. The uninformed player will not know until the end of the round what his payoff is, but he does know that if the Left table was chosen, he gets 1 point, and if the Right table was chosen, he gets 0 points.

Example 2: Suppose instead that the Right table was chosen. Then the informed player knows that she gets 1 point. The uninformed player will not know until the end of the round what his payoff is, but he does know that if the Left table was chosen, he gets 1 point, and if the Right table was chosen, he gets 0 points.

Example 3: Suppose that the Left table was chosen, the informed player chooses A, and the uninformed player chooses B. Then the informed player knows that she gets 0 points. The uninformed player knows that he gets 1 point—in this case his payoff is the same no matter which payoff table had been chosen.

4. PLAYING THE GAME

Once the game begins, your computer screen will be divided into 3 parts. The top part will show the round and stage numbers, and will have a space for messages. The middle part will show "personalized versions" of the payoff tables. The bottom part of the screen will show the results of the most recent 10 rounds you have played. By looking at these results, you may get some insight into how your opponent plays the game.

PAYOFF TABLES

LEFT TABLE

Informed player move	A	A	B	B
Uninformed player move	A	B	A	B
Informed player payoff	1	0	0	0
Uninformed player payoff	0	1	1	1

RIGHT TABLE

Informed player move	A	A	B	B
Uninformed player move	A	B	A	B
Informed player payoff	0	0	0	1
Uninformed player payoff	1	1	1	0

In each stage, the computer will ask you to type in a move (either **A** or **B**). After you have done this, the computer will ask you to verify your move. Type **Y** to verify or **N** to change your move. After you and your opponent have verified, each of you will be able to see the move of the other. After observing these results, press **G** to go on to the next stage.

At the end of each round, uninformed players are told which table had been chosen, and all players are told their total payoff for the round. After observing these results, press **G** to go on to the next round.

5. GETTING PAID

All players will receive $10 for their participation. In addition, at the end of the session, one round and stage will be chosen at random from those that have been played, and all players who earned a point in that round and stage will receive an additional $10. In each stage, either you or your opponent earns a point, so either you or your opponent will get the $10 bonus. Each round and stage is equally likely to be chosen, so you should try to earn as many points as possible in order to give yourself the best chance at winning the extra $10. All earnings will be paid in cash at the end of the session.

REFERENCES

AUMANN, R., AND M. MASCHLER (1995): *Repeated Games of Incomplete Information*. Cambridge, MA and London: MIT Press.

BOLTON, G. E., AND A. OCKENFELS (1998): "ERC: A Theory of Equity, Reciprocity and Competition," *American Economic Review*, forthcoming.

BOYLAN, R., AND M. EL-GAMAL (1993): "Fictitious Play: A Statistical Study of Multiple Economic Experiments," *Games and Economic Behavior*, 5, 205–222.

CAMERER, C., AND T.-H. HO (1996): "Experience-Weighted Attraction Learning in Games: Estimates from Weak-Link Games," mimeo.

——— (1999): "Experience-Weighted Attraction Learning in Normal Form Games," *Econometrica*, 67, 827–874.

CHEUNG, Y.-W., AND D. FRIEDMAN (1997): "Individual Learning in Normal Form Games: Some Laboratory Results," *Games and Economic Behavior*, 19, 46–76.

EREV, I., AND A. E. ROTH (1998): "Predicting How People Play Games: Reinforcement Learning in Experimental Games With Unique, Mixed Strategy Equilibria," *American Economic Review*, 88, 848–881.

FEHR, E., AND K. M. SCHMIDT (1999): "A Theory of Fairness, Competition and Cooperation," *Quarterly Journal of Economics*, 114, 817–868.

FELTOVICH, N. (1997): "Learning and Equilibrium in Asymmetric-Information Games: An Experimental Study," dissertation, University of Pittsburgh.

——— (1999): "Equilibrium and Reinforcement Learning in Private-Information Games: An Experimental Study," *Journal of Economic Dynamics and Control*, 23, 1605–1632.

FUDENBERG, D., AND D. K. LEVINE (1995): "Consistency and Cautious Fictitious Play," *Journal of Economic Dynamics and Control*, 19, 1065–1089.

——— (1998): *The Theory of Learning in Games (Economics Learning and Social Evolution)*. Cambridge: MIT Press.

HARLESS, D. W., AND C. F. CAMERER (1994): "The Predictive Utility of Generalized Expected Utility Theories," *Econometrica*, 62, 1251–1289.

HERRNSTEIN, R. J., AND D. PRELEC (1992): "Melioration," in *Choice Over Time*, ed. by G. Loewenstein and J. Elster. New York: Russell Sage Foundation.

HERRNSTEIN, R. J., D. PRELEC, AND W. VAUGHAN, JR. (1986): "An Intra-personal Prisoners' Dilemma," paper presented at the IX Symposium on Quantitative Analysis of Behavior: Behavioral Economics, Harvard University.

KLEIN, R. W., AND S. J. BROWN (1984): "Model Selection When There Is 'Minimal' Prior Information," *Econometrica*, 52, 1291–1312.

LEVINE, D. K. (1998): "Modeling Altruism and Spitefulness in Experiments," *Review of Economic Dynamics*, 1, 593–622.

MCKELVEY, R. D., AND T. R. PALFREY (1995): "Quantal Response Equilibria for Normal Form Games," *Games and Economic Behavior*, 10, 6–38.

——— (1998): "Quantal Response Equilibria for Extensive Form Games," *Experimental Economics*, 1, 9–41.

MOOKHERJEE, D., AND B. SOPHER (1994): "Learning Behavior in an Experimental Matching Pennies Game," *Games and Economic Behavior*, 7, 62–91.

——— (1997): "Learning and Decision Costs in Experimental Constant Sum Games," *Games and Economic Behavior*, 19, 97–132.

OCHS, J. (1995): "Simple Games With Unique Mixed-Strategy Equilibrium: An Experimental Study," *Games and Economic Behavior*, 10, 202–217.

RABIN, M. (1993): "Incorporating Fairness into Game Theory and Economics," *American Economic Review*, 83, 1281–1302.

ROTH, A. E., AND I. EREV (1995): "Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term," *Games and Economic Behavior*, Special Issue: Nobel Symposium, 8, 164–212.

ROTH, A. E., I. EREV, AND R. SLONIM (1999): "Learning and Equilibrium as Useful Approximations: Accuracy of Prediction on Randomly Selected Constant Sum Games," mimeo.

ROTH, A. E., AND M. W. K. MALOUF (1979): "Game Theoretic Models and the Role of Bargaining," *Psychological Review*, 86, 574–594.

SELTEN, R. (1998): "Axiomatic Characterization of the Quadratic Scoring Rule," *Experimental Economics*, 1, 43–62.

SARIN, R., AND F. VAHID (1999): "Predicting How People Play Games: A Simple Dynamic Model of Choice," mimeo.

SIEGEL, S., AND N. J. CASTELLAN, JR. (1988): *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill.

VAN HUYCK, J., R. BATTALIO, AND R. BEIL (1991): "Strategic Uncertainty, Equilibrium Selection, and Coordination Failure in Average Opinion Games," *Quarterly Journal of Economics*, 106, 885–909.

YATES, J. F. (1982): "External Correspondence: Decompositions of the Mean Probability Score," *Organizational Behavior and Human Performance*, 30, 132–156.