Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Economic Theory 114 (2004) 280–309
JOURNAL OF Economic Theory
http://www.elsevier.com/locate/jet

Unobserved heterogeneity and equilibrium: an experimental study of Bayesian and adaptive learning in normal form games ☆

Jason Shachat[a] and Mark Walker[b,*]

[a] IBM T.J. Watson Research Center, P.O. Box 218, Yorktown Heights, NY 10598, USA
[b] Department of Economics, University of Arizona, Tucson, AZ 85721-0108, USA

Received 13 June 2000; final version received 19 November 2002

Abstract

We describe an experiment based on a simple two-person game designed so that different learning models make different predictions. Econometric analysis of the experimental data reveals clear heterogeneity in the subjects' learning behavior. But the subjects follow only a few decision rules for basing their play on their information, and these rules have simple cognitive interpretations. There is a unique equilibrium in pure strategies, and many equilibria in mixed strategies. We find that the only equilibrium consistent with the data is one of the mixed strategy equilibria. This equilibrium is shown, surprisingly, to be consistent with Jordan's Bayesian model.

JEL classification: C72; C92; D82

Keywords: Bayesian learning; Experiments; Heterogeneity

Theoretical research on out-of-equilibrium behavior, or "learning," in games typically assumes that individuals all behave according to one single learning model.

☆ Prior research done in collaboration with James Cox influenced our thinking about experiments on learning in games. Timothy O'Neill developed the software for the experiment. John Joganic provided programming assistance for the maximum likelihood estimation. For helpful comments and suggestions we are grateful to Vincent Crawford, Mahmoud El Gamal, David Grether, James Jordan, Micheal Keane, Kathleen McGarry, Richard McKelvey, James Ratliff, Robert Rosenthal, Neil Wallace, John Wooders, and William Zame, and to many conference and seminar audiences.
* Corresponding author.
E-mail address: mwalker@arizona.edu (M. Walker).

0022-0531/03/$ - see front matter © 2003 Published by Elsevier Science (USA).
doi:10.1016/S0022-0531(03)00125-X

Following this lead, experimental research has largely been directed toward identifying the single model, or learning rule, that will explain everyone's action choices in a game. In response to experimental evidence that neither reinforcement models nor belief-based models can account for observed behavior by themselves, the ‘‘single model’’ approach has led to hybrid theories that combine elements of more than one model into a single model.

In this paper we take a different approach, explicitly allowing for heterogeneity across subjects. We present an experiment designed so that each one of several behavioral learning rules makes a different prediction about behavior. Moreover, in contrast with previous research, our experiment is designed so that the learning rules' predictions do not rely upon the values of unobservable parameters. This approach allows us to estimate which distinct behavioral rule each of our experimental subjects used.

We find clear evidence of heterogeneity in the ways that our subjects conditioned their behavior in the second play of an incomplete-information game on the information they had received in the first play. Moreover, this heterogeneity in learning behavior was concentrated almost exclusively on cognitively simple behavioral rules. The heterogeneity was systematic, and in principle predictable: the proportions of subjects who can be said to have used various simple behavioral rules coincide with the proportions one would observe if two players were playing a sophisticated mixed strategy Bayesian equilibrium of the repeated game.[1] The equilibrium's path of play is shown to coincide with the more myopic sequence of Bayesian equilibria described by Jordan [7].

Previous experimental research on learning in games has placed subjects in simple games which were repeated many times. The typical approach to identifying a single model that will explain the experimental data first uses the data to estimate the values of unobservable parameters of two or more candidate models, and then uses the estimated parameter values in a forecasting or hypothesis testing exercise that compares the performance of the models. In one strand of research,[2] the estimated models are evaluated against one another according to various statistical criteria in order to determine the ‘‘best’’ model for describing the subjects' behavior. In another strand,[3] a composite of various learning rules is specified, and the composite model's parameters are estimated from the data to obtain the ‘‘best’’ model.

Our approach avoids the estimation of parameter values by using a much shorter learning environment: a once-repeated two-person game of incomplete information.[4]

[1] We will use the term ‘‘repeated game’’ to refer to this twice-played game. As described below, this repeated game was itself played many times between opponents randomly rematched after each two-play sequence.

[2] Examples are Boylan and El Gamal [1], Erev and Roth [6], Mookherjee and Sopher [9], and Roth and Erev [10].

[3] Examples are Camerer and Ho [2], Cheung and Friedman [3], and Stahl [11].

[4] In an earlier attempt to evaluate learning models without estimating parameters (Cox et al. [4]), we studied Jordan's Bayesian learning model by itself, using an experiment in which subjects' ‘‘types’’ (their payoff tables) were fixed for sequences of 15 plays and redrawn at the beginning of each new sequence. Several aspects of that paper played a central role in influencing the research reported here: (1) Most if not

In this very short learning environment, alternative learning rules make precise and distinct predictions about subjects' choices, and the predictions are not conditional on the values of unobservable parameters. For example, two theories that assume significantly different levels of sophistication, best response and Jordan's Bayesian learning model [7], make unique and completely opposite predictions about play in our experiment.

Subjects participated in many repetitions of our short learning environment, with random rematching of opponents between repetitions. This yields a rich panel data set, which permits a maximum likelihood estimation of the individual subjects' behavior, using the approach pioneered by El Gamal and Grether (EG). This analysis of the data, in Sections 5 and 6, indicates that no single theory is consistent with the observed individual behavior: heterogeneity in the subjects' behavior is shown to be a central feature of the data.

We find that most of the subjects' behavior is best described by a few cognitively simple behavioral rules, often even simpler than best response. Nevertheless, the cognitively sophisticated Bayesian model provides a powerful organizing principle for the data by giving us an equilibrium explanation of the heterogeneity: while the Bayesian model is not consistent with most of the individual subjects' behavior, the *proportions* in which the subjects played simple decision rules constitute a mixed strategy equilibrium that is as if two "representative players" were following a sophisticated Bayesian equilibrium learning path.

We establish this equilibrium explanation of the data in Sections 7 and 8. In Section 7, we model the experiment's repeated game as an extensive form game and identify the game's equilibria. There is an enormous number of equilibria, but there is only one equilibrium in pure strategies. Surprisingly, the experimental data are not consistent with the pure strategy equilibrium, but are consistent instead with one of the mixed strategy equilibria, as we demonstrate in Section 8. The path of play in this equilibrium corresponds to a mixed strategy path prescribed by Jordan's Bayesian learning model. Thus, Jordan's model rationalizes the aggregate observed behavior as an equilibrium—an equilibrium in which heterogeneity is an essential feature.

We begin, in Section 1, with a description of the repeated game of incomplete information on which the experiment is based. Section 2 presents the experiment, Section 3 provides a brief description of the aggregate data obtained from the experiment, and Section 4 develops an analytically tractable representation of individual behavior, which forms the basis for the remainder of the paper.

(*footnote continued*)
all learning models make the same predictions about out-of-equilibrium play in the CSW experiment. (2) In most of the experimental sessions subjects were randomly rematched between every stage of each 15-play sequence, essentially eliminating the common knowledge of prior play required in Bayesian learning. (3) All Jordan-model learning takes place in the first two plays of the CSW experiment's sequences. For all these reasons, estimation of subjects' decision rules could not be performed, and the analysis in CSW focused only on the question whether play converged to the Jordan equilibrium.

1. A simple learning environment

Suppose two people are playing a normal form game, such as any one of the four 2×2 games depicted in Table 1, and suppose that we want to predict how they will play. We could use equilibrium theory to make our prediction, but equilibrium theory does not tell us what actions the players will choose if their expectations about one another are incorrect. Theories about learning in games, on the other hand, do not depend upon players having correct expectations, but instead make assumptions about the players' behavior when they are not in equilibrium. A particular theory of learning might or might not predict the players' actions in a single play of the game, but the primary objective of learning theories is to predict how play will proceed when the game is played repeatedly.

Now suppose that one of the games in Table 1 is indeed played repeatedly by the same two players, and that after each play of the game each of the players observes the action his opponent has just taken. It is useful to consider how some alternative learning theories proceed.

Consider, for example, two simple adaptive models of learning: best response and fictitious play. Neither model has anything to say about what the players will do the first time they play the game. But then, at every subsequent play, each model assumes that a player will choose an action which is a best response to a "forecast" of the action his opponent will take. The forecast is based upon the opponent's observed actions in past plays and is formed differently in the two models.

Table 1
Possible games and their Nash equilibria

Game 00

	L	R
T	0,0	2,1
B	1,2	0,0

Pure Nash equilibrium: (B,L) and (T,R)
Mixed strategy N.E.: (2/3, 1/3)

Game 01

	L	R
T	0,2	2,0
B	1,0	0,1

Pure Nash equilibrium: none
Mixed strategy N.E.: (2/3, 1/3)

Game 10

	L	R
T	1,0	0,1
B	0,2	2,0

Pure Nash equilibrium: none
Mixed strategy N.E.: (1/3, 1/3)

Game 11

	L	R
T	1,2	0,0
B	0,0	2,1

Pure Nash equilibrium: (T,L) and (B,R)
Mixed strategy N.E.: (1/3, 1/3)

Note: Mixed strategy profiles are expressed as (Row's prob. of B, Col's prob. of R).

Conditional on the game they are playing, and on the players' actions in the first play, each model predicts the ensuing path of play.

Bayesian theories of learning (examples are Jordan [7] and Kalai and Lehrer [8]) assume a much greater degree of rationality and cognitive sophistication. Each player is assumed to have a prior belief about his opponent's characteristics, and he is assumed to use his belief to determine the best action to take. The learning that takes place in these models consists of each player using his observations of his opponent's past actions to update his belief via Bayes' Rule. Conditional on the players' characteristics and on their beliefs prior to their first play, the model makes either a deterministic or a stochastic prediction about the path of play.

We wish to design an experiment in which various learning models, which differ in the degree of rationality they assume, will each make precise and distinct predictions about play, without the need to estimate the values of behavioral parameters. This will require that we add enough structure so that one or more Bayesian learning models will be well specified. We accomplish this by embedding the four 2×2 games in Table 1 into a single incomplete-information structure.

Assume, then, that two players are going to repeatedly play one of the four 2×2 games in Table 1. Assume that neither player knows which game they are playing: each knows his own payoff function, but does not know his opponent's payoff function. Note that in the two games in the top half of Table 1 (the games labelled 00 and 01), the Row player has the same payoff function; he also has the same payoff function in both of the games in the bottom half of Table 1 (the games labelled 10 and 11). Similarly, the Column player's payoff function is the same in the two games in the left half of Table 1 (games 00 and 10) and is also the same in the two games in the right half (games 01 and 11).

Our move to an incomplete information environment has no effect on the predictions made by adaptive models, such as best response and fictitious play: these models continue to make no prediction about first-period play, and then, conditional on play at the first period, they predict the subsequent path of play. Turning to Bayesian models of learning, perhaps the simplest model is the one introduced by Jordan. Jordan's model requires us to specify the players' prior beliefs about which game they are playing. In other words, the incomplete information environment we have just described must be formalized as a *game* of incomplete information.

Thus, we assume that each player can be one of only two types (called Type 0 and Type 1), where a player's type is simply his payoff function. The Row player is Type 0 if his payoff function is the one that appears in the top half of Table 1 for the Row player; he is Type 1 if his payoff function is the one in the bottom half of Table 1. Similarly, the Column player is Type 0 when he has the payoff function that appears in the left half of Table 1, and he is Type 1 when he has the payoff function in the right half of Table 1. Table 2 displays each player's two types, or payoff functions. We assume that each player believes the other player's types to be equally likely, and that this is common knowledge.

Jordan introduces the useful notion of the *true game* the players are playing, i.e., the complete-information game defined by the players' actual (realized) payoff functions. In the incomplete-information game we have just defined there are four

Table 2
Player types

Row

Type 0

T	0	2
B	1	0

Type 1

T	1	0
B	0	2

Column

Type 0

L	R
0	1
2	0

Type 1

L	R
2	0
0	1

possible true games, the games depicted in Table 1, and they are equally likely (because each player's types are equally likely): Game 00, in which Row Type 0 plays against Column Type 0; and Game 01, Game 10, and Game 11, each defined similarly. Table 1 also shows the Nash equilibria of each of the four possible true games.

Jordan's model effectively assumes that at each play of the game the players choose their "stage game" strategies myopically, seeking only to maximize their current-period expected earnings, without considering the effect their decisions will have on play in future stage games. Thus, the model assumes that the first time the players play the game they will simply play a Bayesian equilibrium of the incomplete information game. Each player then observes the action his opponent has just taken, and uses that observation to update his belief via Bayes' Rule. At every subsequent play of the game the players continue in the same way—always playing a Bayesian equilibrium (BE) for their current beliefs about their opponents' types, and each player then using his observation of the other's action to update his own belief via Bayes' Rule. We refer to such a sequence of Bayesian equilibria as a *Jordan Learning Sequence*, or JLS.[5] Jordan shows that under certain regularity conditions on the

[5] Jordan [7] studies a slightly different and more general construction than the JLS, which he refers to as a Bayesian Strategy Process.

players' prior beliefs, the sequence of action profiles that results from any JLS converges to a Nash equilibrium (NE) of the true game the players are playing.

It is relatively straightforward to determine the path of play predicted by Jordan's model in our simple incomplete information environment. Recall that at the first period of play the players play a Bayesian equilibrium of the incomplete information game. It is easy to show that there is a unique pure strategy Bayesian equilibrium: Column always plays Left (thus always choosing his high-payoff action[6]); Row chooses Bottom if Type 0 and Top if Type 1 (thus always choosing his low-payoff action). There are also many mixed strategy Bayesian equilibria; in each of these equilibria one of the Column types and at least one of the Row types play strictly mixed strategies. We ignore the mixed strategy Bayesian equilibria for now, and assume that the unique pure strategy BE is played.

After the first play is completed, each player observes the choice his opponent has made and uses that observation to update his belief about the opponent's type. Specifically, the Row player has fully revealed his type to the Column player by his type-dependent BE play, but the Column player has revealed nothing to the Row player by his BE play of Left, so the Row player still believes the Column player is with equal probability either Type 0 or Type 1. These new beliefs yield the following unique pure strategy BE at the second play of the game (where TR, for example, means that Row plays Top and Column plays Right):

True game:	00	01	10	11
BE at 2nd play:	TR	TL	BL	BR

There are also many mixed-strategy Bayesian equilibria at the second period, even after a pure-strategy BE at the first period. We again ignore the mixed-strategy equilibria for now, but we will return to them later, when they will play an important role in helping us rationalize the experimental data.

Following the second play of the game, each player has learned his opponent's type, and all subsequent play is therefore at a NE of the true game. (The players now have complete information, and therefore the BE for their current beliefs are simply the NE of the true game.) Note, however, in Table 1 that two of the true games have multiple pure strategy equilibria (as well as a mixed-strategy equilibrium), and that the other two games have a unique equilibrium, but it is in strictly mixed strategies. Therefore, the Bayesian model does not predict unique choices in the third and subsequent plays of the game, even if the unique pure-strategy BE have occurred at the first two plays and we assume pure-strategy BE at subsequent periods. Since we want alternative theories to make precise predictions (and for additional reasons as well), we will construct our experiment so that players engage in only two rounds of play after a true game is drawn.

[6] Note that whether a player is Type 0 or Type 1, it is always the case that one of his actions will yield a payoff of either 0 or 1 and his other action will yield a payoff of either 0 or 2. We refer to these actions as the player's *low-payoff action* and his *high-payoff action*, respectively.

With sequences of play restricted to only two periods (the true game repeated only a single time), adaptive models predict that at the second period each player will best respond to his opponent's first-period action. Thus, in the learning environment we study, the best response model is representative of the class of adaptive models. Further, if play at the first period is the pure strategy BE, as it typically was in our experiment, then it is easy to see that best response at the second period always yields exactly the opposite behavior as the Jordan model: in games 00, 01, 10, and 11 we saw above that the Jordan model yields, respectively, TR, TL, BL, and BR; best response yields BL, BR, TR, and TL.

2. The experiment

Each experimental session included 12 subjects, six of whom always played as the Row player, the other six always playing as the Column player. Subjects were randomly matched into six pairs, each containing one Row subject and one Column subject, to play the repeated game described in Section 1. Two drawings were held at the outset of the repeated game, a drawing to determine the Row type (i.e., the Row payoff table) and a drawing to determine the Column type; thus, all six Row subjects were the same type and all six Column subjects were the same type. The game was then played *only twice*—subjects chose their actions at the first play and then each subject's action was revealed to his opponent (but not to the other ten subjects), and then, with their types remaining the same, the subjects chose their actions for the second play and those actions were revealed to their opponents. Then the subjects were rematched, randomly, to again play the game twice, with a new random drawing of Row and Column types. We refer to each such two-play sequence, with a single random drawing of types at the outset, as a *regime*. Each experimental session consisted of 50 such regimes, each regime beginning with a random matching into six Row–Column pairs, followed by the drawing of types, and then by the two repetitions of play.[7]

Eight experimental sessions were conducted in the networked computer facility at the University of Arizona's Economic Science Laboratory. The experiment therefore generated a panel data set with a cross-section consisting of 96 subjects (48 Row subjects and 48 Column subjects), each of whom played 50 two-period regimes. Subjects earned one dollar for every three payoff units they accumulated; i.e., the monetary payoffs in Table 2 were effectively 33 cents, 67 cents, and zero. Each session lasted less than 2 h, and subjects' earnings were between \$14.50 and \$37.00, to which a \$5.00 participation payment was added.

3. Experimental results: the aggregate data

We first investigate the experimental data at the aggregate level, attempting to determine whether there is a single learning model which describes the subjects'

[7] An online demonstration of the experimental environment, via the computer interface actually used in the experiment, is available at *www.u.arizona.edu/~mwalker/exp-demo*.

behavior. In Section 3.1, we determine the degree to which the subjects' first-period play conformed with the first-period pure strategy Bayesian equilibrium. In Section 3.2, we ask how consistent the subjects' second-period play was with either of our benchmark models of learning, or indeed with any one single model of learning.

3.1. Period 1 play

Fig. 1 presents, in time series form, the percentage of Row and Column subjects whose period 1 play was their part of the pure strategy BE. Recall from Section 1 that in the pure strategy period 1 Bayesian equilibrium the Column player selects his high-payoff action in each of his types (i.e., he plays Left in each type). In fact, the Column subjects in the experiment predominantly played Left from the very beginning of the experiment. In contrast, the Row subjects initially played their high-payoff action more frequently than their Bayesian equilibrium low-payoff action. But then, as the experiment progressed and they consistently observed Left play from their opponents, Row subjects came to play predominantly their low-payoff action. During the final 25 regimes (the last half of the experiment), Column subjects played Left nearly 85% of the time, and Row subjects played their low-payoff action nearly 79% of the time.

Table 3 presents the joint and marginal frequencies of observed play in each true game during the final half of the experiment. The cells predicted by the myopic pure strategy BE at Period 1 (these are the cells in which the frequency is underlined) clearly have the highest frequencies—about 79% for Row subjects and 85% for Column subjects, as indicated above. The pure strategy BE predicts that these

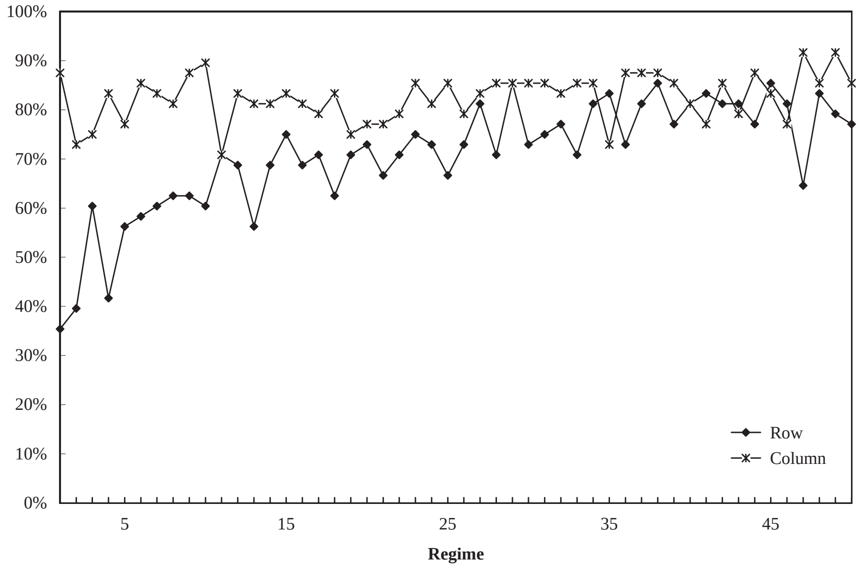

Fig. 1. Percentage of play consistent with period 1 BNE.

Table 3
Cell frequencies for the four games

Period 1
(A single underline indicates Jordan pure strategy prediction)

Game 0 v. 0
N=318

	L	R	Marginals
T	20.13%	2.52%	22.64%
B	62.58%	14.78%	77.36%
Marginals	82.70%	17.30%	

Game 0 v. 1
N=318

	L	R	Marginals
T	17.61%	2.52%	20.13%
B	66.67%	13.21%	79.87%
Marginals	84.28%	15.72%	

Game 1 v. 0
N=294

	L	R	Marginals
T	64.29%	12.93%	77.21%
B	19.05%	3.74%	22.79%
Marginals	83.33%	16.67%	

Game 1 v. 1
N=222

	L	R	Marginals
T	70.72%	9.46%	80.18%
B	16.22%	3.60%	19.82%
Marginals	86.94%	13.06%	

Period 2
(Conditional on Bayes equilibrium play at Period 1, a single underline here indicates the Jordan pure strategy prediction, and a double underline indicates the best response prediction)

Game 0 v. 0
N=318

	L	R	Marginals
T	31.45%	14.15%	45.60%
B	46.86%	7.55%	54.40%
Marginals	78.30%	21.70%	

Game 0 v. 1
N=318

	L	R	Marginals
T	24.53%	12.26%	36.79%
B	33.65%	29.56%	63.21%
Marginals	58.18%	41.82%	

Game 1 v. 0
N=294

	L	R	Marginals
T	30.27%	24.49%	54.76%
B	28.23%	17.01%	45.24%
Marginals	58.50%	41.50%	

Game 1 v. 1
N=222

	L	R	Marginals
T	50.45%	8.11%	58.56%
B	30.18%	11.26%	41.44%
Marginals	80.63%	19.37%	

frequencies will all be 100%, so it might seem that one of the mixed strategy Bayesian equilibria could be a better description of the subjects' behavior. But a straightforward statistical test establishes convincingly that the data are not consistent with any of the mixed strategy equilibria. (Note, for example, in Table 3

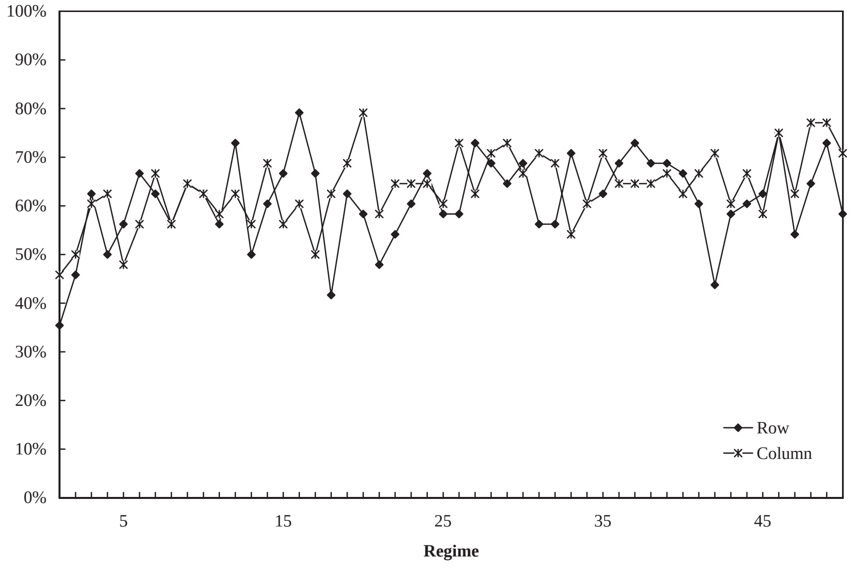

Fig. 2. Percentage of play consistent with period 2 Cournot best response prediction.

that the Column subjects played Left at virtually the same rate whether they were Type 0 or Type 1, while in every mixed strategy equilibrium the Column player's play is highly type-dependent.)

3.2. Period 2 play

To the extent that the pure strategy Bayesian equilibrium is a good description of the observed Period 1 play, the Bayesian and best response models then make specific—and completely opposite—predictions for Period 2 play, as described in Section 1. Thus, if either of our benchmark models is the "right" model of how people learn to play in strategic settings, this will be clearly reflected in the data.

Fig. 2 presents a time series in which, for each two-play regime, the vertical height of the graphs measures the proportion of Row subjects and Column subjects whose Period 2 play was a best response to the observed Period 1 choices of their opponents. The rates are initially close to 50%—second-period play in the early regimes was divided about equally between the best response action and the opposite (Bayesian) action.[8] Play moved slightly in the direction of best response as subjects gained more experience, but it did not exhibit the decisive separation we would observe if either benchmark model were the "true" model describing subjects' behavior.

Table 3 presents the aggregate frequencies with which each profile of play occurred in each true game. These numbers reflect the inconclusiveness of Fig. 2 for our two

[8] The Bayesian action is conditional upon the first-period play conforming with the Period 1 pure strategy Bayesian equilibrium.

benchmark models. No other single behavioral model suggests itself as a way to rationalize the data in this table. One possible explanation of the inconclusive nature of the data is that individual subjects' behavior might have been quite noisy and simply not very systematic. Another is that individual subjects' play might have been relatively systematic, but random—perhaps subjects were in effect playing mixed strategies. A third possible explanation of the aggregate data is that individual subjects may have been playing systematically and deterministically, but there was unobserved heterogeneity across subjects. In the following section we undertake an analysis of the data at the disaggregated, individual level in an attempt to determine whether any one of these explanations can account for the data.

4. Representation of individual behavior

In order to analyze our experimental data at the individual level we must have a way to *represent* individual behavior. In this section we develop a binary representation of individual behavior in our experimental game by characterizing every possible decision rule (or strategy) for an individual as a string of binary digits. This representation will play a central role in subsequent sections.

A decision rule, or *strategy*, specifies what action a player will take in each information state in which he could potentially find himself. In our experiment, there are two kinds of information states: first-period and second-period states. At the first period of play, a player knows only his own payoff function (or "type"), π; there are only two possible states for this information, either Type 0 or Type 1. Thus, a player's first-period decision rule can be represented as a function ξ_1 from the set of his possible types, $\Pi = \{0, 1\}$, to the set of actions he could choose, $A = \{0, 1\}$. There are $2^2 = 4$ such functions, and each can be represented as an ordered pair of members of A, or as a two-digit binary number, in which the first component or digit is the action to be taken when Type 0, and the second is the action to be taken when Type 1. (We assume for the moment that players do not play mixed strategies.)

At the second period of play, a player's information has expanded beyond the single binary bit of information that he had in the first period to now include also his own first-period action and his observation of the first-period action that his co-player chose. Each of those variables is also binary—each player could have chosen either 0 or 1 (Top or Bottom, Left or Right)—and therefore a player's decision rule at the second period is a function ξ_2 from the eight-element domain $\{0, 1\} \times \{0, 1\} \times \{0, 1\}$ (the player's type, and each player's first-period action) to the two-element action set $A = \{0, 1\}$.

The information states, or information sets, and the actions available at each one of them, are depicted in a tree diagram in Fig. 3.[9] Note that each player has 10 information sets altogether, with two actions available at each one, for a total of $2^{10} = 1024$ possible strategies for each player. Each strategy is a function

[9] In order to make the tree more readable, the Column player's second-period decision branches (two at each of the 32 rightmost nodes) are not depicted. Thus, the tree actually has 64 terminal nodes.

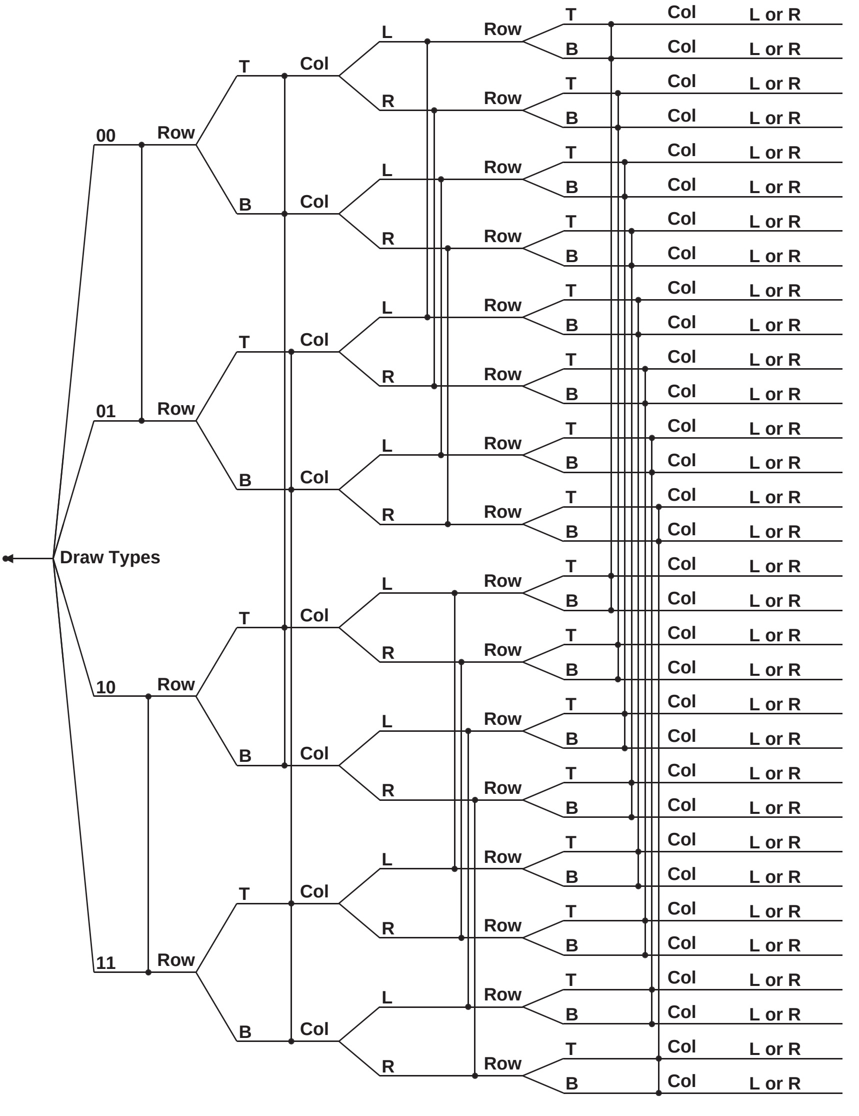

Fig. 3. Game tree.

$\xi = (\xi_1, \xi_2)$, and is naturally represented as a 10-digit binary number, one digit for each information set.

Note that the strategic situation has been represented in Fig. 3 as an extensive form game, except that the payoffs at the terminal nodes have not been specified. (Thus, Fig. 3 actually depicts a *game form*.) We ignore payoffs for now, in order not to impose on the data any restrictions derived from theory (such as, for example, that the data constitute some kind of equilibrium). After we have obtained a satisfactory representation of the data, the payoffs will become important in Section

7, where we will rationalize the data as an equilibrium of the extensive form game defined by the payoffs.

The 1024 strategies we have obtained for each player present some econometric and computational difficulties. The maximum likelihood procedure we will employ is computationally impractical if there is such a large number of strategies for each player. More important, many of the eight second-period information states for each player are likely to be reached only very rarely, if at all, so that econometric estimation would be unreliable. Note, in particular, that a player who is following one of his 1024 strategies will reach only those second-period information sets that are consistent with the strategy's prescription for first-period play. (And indeed, we see in the data that nearly 90% of the typical subject's play did pass through only four of his eight possible second-period information states.)

Thus, we make the obvious simplification, in which we assume that a player's second-period information consists only of his own type and his co-player's first-period play, thereby reducing the number of second-period information sets from eight to four.[10] Each second-period strategy is now a function from $\{0,1\} \times \{0,1\}$ to the set $A = \{0,1\}$, and can therefore be represented as an ordered four-tuple of members of A, or as a four-digit binary number.

Thus, altogether, we represent a player's strategy as a function $\xi = (\xi_1, \xi_2)$, or equivalently as a six-digit binary number: two digits for the strategy's first-period prescription and four digits for its second-period prescription. There are therefore $2^6 = 64$ strategies, or decision rules, available to each player. The specific mapping we will use from digits to actions is as follows, where z denotes a player's observation of his opponent's play at period 1:

1st digit: first-period action when Type 0
2nd digit: first-period action when Type 1
3rd digit: second-period action when Type 0 and $z = 0$
4th digit: second-period action when Type 0 and $z = 1$
5th digit: second-period action when Type 1 and $z = 0$
6th digit: second-period action when Type 1 and $z = 1$

For example, the binary number 001001 (decimal[11] number 9) represents for a Column player the strategy "Play Left at the first period no matter which type I am; and at the second period, play Right if my type 'matches' z, otherwise play Left." It is easy to verify that this particular strategy is the Cournot best response rule for the Column player at Period 2; at Period 1, as we have already seen, this strategy is the Column player's part of the pure-strategy BE. The same number, 001001, of course represents the same decision rule for the Row player if we substitute Top and Bottom for Left and Right in the description of the rule. It happens that this is the

[10] In other words, we do not distinguish between strategies that are *equivalent*—i.e., that play the same as one another against each of the opponent's strategies. Each such equivalence class contains 16 strategies, and there are 64 equivalence classes.

[11] We will often use the decimal (i.e., the usual base-10) equivalents of the binary numbers that represent strategies; the decimal forms seem to be cognitively easier to remember.

best response rule for the Row player, just as it is for the Column player, but, as we have already seen, it is *not* the Row player's part of the pure strategy BE at period 1.

The following rules are worth noting (an asterisk indicates a digit for which the rule does not specify a value):

Player	Decision rule	Binary	Decimal
Row	Best response	$^{**}1001$	$9, 25, 41, 57$
Column	Best response	$^{**}1001$	$9, 25, 41, 57$
Row	BE at $t = 1$	10^{****}	$32, \ldots, 47$
Column	BE at $t = 1$	00^{****}	$0, \ldots, 15$
Row	Jordan model	$100^{*}1^{*}$	$34, 35, 38, 39$
Column	Jordan model	000110	6
Row	Low-payoff action	101010	42
Column	Low-payoff action	111111	63
Row	High-payoff action	010101	21
Column	High-payoff action	000000	0

5. The estimation procedure

The experiment produced a panel data set which contains, for each one of our 48 Row and 48 Column subjects, the detailed record of each of the 50 regimes in which the subject participated: viz., the true game that was being played, the subject's own play, and his opponent's play. The set of 64 possible decision rules that a player or subject might use to condition his play on his information, which we developed in Section 4, provides us with an organizing principle for analyzing this disaggregated set of data. Our objective in this section and the next is to identify, for each subject, the single decision rule, or strategy, that gives the best description of that subject's behavior in the experiment. We will also ask whether we can identify a single decision rule that adequately describes the behavior of all the subjects who were playing as Row players, and a single rule that similarly describes the behavior of all the subjects who played the Column role. And if no single Row and Column decision rules can be identified, can we instead identify the distribution of subjects across the decision rules?

El Gamal and Grether (EG) [5] present an estimation procedure which is ideally suited for our task. Given a panel data set and a set of behavioral rules, the EG procedure uses a maximum likelihood approach to identify a parsimonious subset of the behavioral rules and to identify, for each individual in the panel, which rule in this parsimonious set is the one that best describes his behavior. In our adaptation of the EG procedure, we begin with the set S of 64 six-digit binary decision rules developed in Section 4, and we determine which parsimonious subset of S best characterizes our Row subjects' observed behavior, which subset best characterizes our Column subjects' behavior, and for each individual subject we identify which one of these rules is the best description of his behavior in the experiment.

Our analysis uses only the second half of the experiment (the last 25 regimes): while some subjects exhibited movement across rules over the course of the experiment, there was little migration across rules in the last 25 regimes. We describe the estimation procedure for the Row subjects; estimation of the Column subjects' behavior is identical.

For each Row subject $i \in I = \{1, \ldots, 48\}$, each decision rule $s \in S = \{000000, \ldots, 111111\}$, each regime $r \in \{1, \ldots, 25\}$, and each period $t = 1$ and $t = 2$ of play, define the variable x^s_{irt} as follows:

$$x^s_{irt} = \begin{cases} 1 & \text{if, in period } t \text{ of regime } r, \text{ subject } i\text{'s choice agreed} \\ & \quad \text{with the prescription by rule } s \text{ for period } t, \\ 0 & \text{otherwise.} \end{cases}$$

For each subject i and each rule s, we define the summary statistics X^s_{i1} and X^s_{i2} to be the number of regimes in which subject i's first- and second-period actions were consistent with rule s:

$$X^s_{i1} = \sum_{r=1}^{25} x^s_{ir1} \quad \text{and} \quad X^s_{i2} = \sum_{r=1}^{25} x^s_{ir2} \ .$$

We write X_{i1} and X_{i2} for the 64-tuples $(X^s_{i1})_{s \in S}$ and $(X^s_{i2})_{s \in S}$, and X_1 and X_2 for the 48-tuples $(X_{i1})_{i \in I}$ and $(X_{i2})_{i \in I}$. Thus, the entire Row-subject data set in the second half of the experiment is summarized, for our purposes here, by X_1 and X_2.

We model the stochastic data generating process as follows. We assume that there is a ("small") subset K of the set S of 64 decision rules which adequately describes the behavior of all the Row subjects (i.e., that every one of the 48 subjects uses one of the rules $s \in K$). Every time a subject makes a choice, however, we assume that there is a probability that he makes an error, deviating from the action his rule prescribes for him, and choosing instead the opposite action. We assume that the probability of an error occurring in the first period of a regime is e_1 and that the error probability in the second period is e_2. We further assume that the two error probabilities are independent of each other; that they are independent of the subject's information state (including his own type and his own action at $t = 1$); that each of the 48 Row subjects has the same two error probabilities e_1 and e_2; and that the probabilities of different subjects making an error are independent. Our task is to estimate the set K (one set for the Row subjects, another for the Column subjects) and the distribution of the use of the rules in K across the experimental subjects, as well as the error rates e_1 and e_2.

Under the data generating process we have just described, the likelihood function for a subject i and a rule s is as follows:

$$f_s(s, e_1, e_2; X^s_{i1}, X^s_{i2}) = (1 - e_1)^{X^s_{i1}} \ (e_1)^{25 - X^s_{i1}} \ (1 - e_2)^{X^s_{i2}} \ (e_2)^{25 - X^s_{i2}}.$$

In order to keep track of which rule we are assigning to each subject, we introduce the following construction: let δ_{is} be 1 if subject i uses rule s, and 0 otherwise. Thus, in the matrix $(\delta_{is})_{1 \leqslant i \leqslant 48;\, 0 \leqslant s \leqslant 63}$, the row associated with subject i contains all zeroes, except for a single entry of 1 in the column associated with the strategy s assigned to subject i. (Each subject is assumed to use only one rule.)

We take as given, for the moment, the size of the set K, denoted $|K|$. Our maximum likelihood estimation problem is therefore to find the two error rates $\hat{e}_1$ and $\hat{e}_2$, the set $\hat{K} = \{\hat{s}_1, \ldots, \hat{s}_{|K|}\}$, and the matrix $(\hat{\delta}_{is})$ that maximize the value of the $|K|$-rule likelihood function

$$F_{|K|}(K, (\delta_{is}), e_1, e_2; X_1, X_2) = \prod_{i=1}^{48} \prod_{s=0}^{63} f_s(s, e_1, e_2; X_{i1}^s, X_{i2}^s)^{\delta_{is}},$$

subject to the constraint that $\delta_{is} = 0$ for each $(i, s) \in I \times (S \backslash K)$. This constraint prevents the assignment of rules outside K to any subject. Equivalently, our problem is to maximize, subject to the same constraint, the value of the $|K|$-rule log-likelihood function

$$\ln F_{|K|}(K, (\delta_{is}), e_1, e_2; X_1, X_2) = \sum_{i=1}^{48} \sum_{s=0}^{63} \delta_{is} \ln f_s(s, e_1, e_2; X_{i1}^s, X_{i2}^s).$$

Now let us drop the assumption that $|K|$ is fixed, treating it instead as a parameter whose value must be estimated. Of course, the maximum value of the likelihood function $F_{|K|}$ increases as we increase $|K|$, but if the increase in the value of $F_{|K|}$ is small, we risk overfitting the data when we increase $|K|$. A central feature of the EG procedure is the specification of a penalty function designed to take account of this "cost" of a larger set K. The penalty, which is strictly increasing in $|K|$, is subtracted from the log-likelihood value. A derivation parallel to the one in El-Gamal and Grether [5] yields the following penalized log-likelihood function:

$$\text{PLF}(|K|, K, (\delta_{is}), e_1, e_2; X_1, X_2) = \ln F_{|K|} - |K| \ln(2) - |K| \ln(64) - 48 \ln(|K|).$$

The El Gamal and Grether procedure, applied to our data generating process and our data, consists of selecting the parameter values $|K|, K, (\delta_{is}), e_1, e_2$ that maximize the value of this penalized log-likelihood function, by maximizing $F_{|K|}$ for successively larger values of $|K|$, until the maximum of the function $\text{PLF}(\cdot)$ is found.

6. The estimation results

In this section we carry out the maximum likelihood estimation of subjects' decision rules. We find that nearly all the subjects adhered for the most part to just one decision rule, but there is also unmistakable heterogeneity in the rules used by both Row subjects and Column subjects. We find that almost every Row subject's observed behavior can be characterized by one or the other of just two decision rules for second-period play, and similarly for the Column subjects. The estimation indicates that each of these rules was used by a substantial fraction of the subjects. Each rule has a straightforward cognitive interpretation as a simple way to play in the repeated game.

We have performed the estimation separately for the first and second halves of the experiment, because we expect that subjects may typically require some experience in order to adequately understand their strategic situation, and play in the last half of the experiment is therefore likely to be more informative. Moreover, separate first-half and second-half estimations provide some information about the extent to which the subjects' behavior was changing over time. We report only the estimation for the second half of the experiment,[12] in Tables 4 and 5.

6.1. Estimation of column subjects' behavior

Estimation of the Column subjects' decision rules for the second half of the experiment is summarized in Table 4. The table describes the estimation results as we successively increase $|K|$, the size of the set K, from $|K| = 1$ to $|K| = 8$, as the leftmost column of the table indicates. The remaining columns of the table contain the following information for each $|K|$: (1) the estimated set K of decision rules; (2) the number of subjects assigned to each of the rules in K; (3) the likelihood and penalized likelihood values at the size-$|K|$ estimates; (4) the estimated values of the error rates e_1 and e_2; and (5) the value of the likelihood-ratio test statistic for homogeneity, denoted by λ. The shaded region in each table indicates the maximum penalized likelihood estimates.

Which rules did the Column subjects use? The set of rules that maximizes the value of the penalized likelihood function is the six-rule set $K_{\text{Col}} = \{0, 8, 9, 24, 32, 59\}$, which is displayed in the shaded region of Table 4. All but five of the 48 Column subjects are estimated to have used rules whose first two digits are zero (rules of the form 00^{****}); these are the rules in which the Column player plays his part of the pure strategy Bayesian equilibrium in the first period. Indeed, for all the reported set sizes $|K|$, at least 43 of the Column subjects are estimated to have used such rules; hence, this observation does not depend upon the particular penalty function that has been used to define the penalized likelihood function. In their first-period play, then, the Column subjects were quite homogeneous, as we have already observed.

In light of the homogeneous first-period play, we note that the data nevertheless decisively reject the hypothesis that the Column subjects' overall behavior was homogeneous, i.e., that a single decision rule s can adequately account for the observed behavior. (Thus, second-period play clearly cannot have been homogeneous.) For each value of $|K|$ larger than one, the value of the chi-square homogeneity test statistic λ is far out in the tail of the statistic's distribution under

[12] The separate estimations indicate that only about a third of both Row and Column subjects changed their behavior significantly from the early part of the experiment to the later part. The changes were not simply random: the rules the subjects primarily moved to, rule 9 for the Column subjects and rule 41 for the Row subjects, are the rules that prescribe Bayesian equilibrium play at the first period and best response behavior at the second period.

Table 4
Estimates for column subjects (regimes 26–50)

$\|K\|$	Set K of rules		# Using rule	Likelihood value	Penalized likelihod	e_1	e_2	Homogeneity statistic λ
$\|K\|$=1	0	000000	48	-1272.17	-1277.02	0.17	0.31	
$\|K\|$=2	0	000000	24	-1149.13	-1192.11	0.17	0.21	246.08
	9	001001	24					
$\|K\|$=3	9	001001	23	-1086.77	-1154.06	0.12	0.22	370.80
	0	000000	21					
	56	111000	4					
$\|K\|$=4	9	001001	23	-1054.86	-1140.81	0.12	0.21	434.62
	0	000000	21					
	32	100000	2					
	59	111011	2					
$\|K\|$=5	9	001001	23	-1031.99	-1133.50	0.11	0.21	480.36
	0	000000	20					
	32	100000	2					
	59	111011	2					
	16	010000	1					
$\|K\|$=6	9	001001	20	-1009.25	-1124.37	0.11	0.20	525.84
	0	000000	15					
	8	001000	8					
	32	100000	2					
	59	111011	2					
	24	011000	1					
$\|K\|$=7	9	001001	20	-1002.28	-1129.65	0.11	0.20	539.78
	0	000000	15					
	8	001000	8					
	32	100000	2					
	59	111011	1					
	54	110110	1					
	16	010000	1					
$\|K\|$=8	9	001001	18	-996.61	-1135.24	0.11	0.19	551.12
	0	000000	14					
	8	001000	8					
	1	000001	3					
	32	100000	2					
	59	111011	1					
	54	110110	1					
	24	011000	1					

the null hypothesis of homogeneity. The value $\lambda = 246$, for example, for $|K| = 2$ in Table 4, yields a p-value of approximately 1–10^{-50}. Thus, the homogeneity hypothesis is clearly rejected in favor of heterogeneity. This is reflected too in the estimated error rates, which are about 2/3 as large when K is allowed to contain several rules as they are when K is constrained to contain only a single rule.

It is of course not surprising that homogeneity is rejected. But what *is* surprising, perhaps, is that the Column subjects' heterogeneous second-period play can be

Table 5
Estimates for row subjects (regimes 25–50)

$\|K\|$	Set K of rules		# Using rule	Likelihood value	Penalized likelihod	e_1	e_2	Homogeneity statistic λ
$\|K\|$=1	41	101001	48	-1426.66	-1431.51	0.21	0.39	
$\|K\|$=2	41	101001	41	-1319.31	-1362.29	0.16	0.37	214.70
	19	010011	7					
$\|K\|$=3	41	101001	31	-1255.56	-1322.85	0.16	0.30	342.20
	39	100111	10					
	19	010011	7					
$\|K\|$=4	41	101001	21	-1230.77	-1316.72	0.16	0.27	391.78
	44	101100	11					
	35	100011	9					
	19	010011	7					
$\|K\|$=5	41	101001	21	-1219.51	-1321.02	0.16	0.27	414.30
	44	101100	11					
	35	100011	10					
	27	011011	5					
	54	110110	1					
$\|K\|$=6	41	101001	17	-1211.47	-1326.59	0.15	0.27	430.38
	44	101100	11					
	35	100011	10					
	27	011011	5					
	43	101011	4					
	54	110110	1					
$\|K\|$=7	41	101001	17	-1203.51	-1330.88	0.15	0.26	446.30
	44	101100	11					
	35	100011	9					
	27	011011	5					
	43	101011	4					
	54	110110	1					
	6	000110	1					
$\|K\|$=8	41	101001	17	-1198.23	-1336.86	0.15	0.26	456.86
	44	101100	11					
	39	100111	5					
	27	011011	5					
	43	101011	4					
	34	100010	4					
	6	000110	1					
	54	110110	1					

effectively summarized in terms of just two decision rules, rules 0 and 9. The second-period prescriptions of rules 0 and 9 are **0000 and **1001, respectively: rule 0 prescribes play of Left regardless of the player's information (i.e., ''always choose the high-payoff action''), and rule 9 prescribes best response. Nearly three quarters of the subjects (in the shaded region of Table 4) played either rule 0 or rule 9, slightly more playing 9 than 0. Perhaps even more striking is the following observation: exactly half of all the 64 rules in S that the subjects *could* have used differ at two or more second-period digits (information sets) from *both* rules 0 and 9—we could say that such rules are "not close to either 0 or 9"—but every one of the 48 Column subjects is estimated to have used a rule which differed from either 0 or 9 in no more than one of the four second-period digits, i.e., rules that are "close" to either 0 or 9, or to both. Indeed, note that for *all* the rule-set sizes $|K|$ reported in Table 4, the estimated play by virtually every one of the Column subjects is a rule "close" to either rule 0 or rule 9.

6.2. Estimation of row subjects' behavior

Estimation of the Row subjects' decision rules is summarized in Table 5, which presents the same information for the Row subjects as Table 4 does for the Column subjects. The four-rule set $K_{\text{Row}} = \{19, 35, 41, 44\}$, displayed in the shaded region of Table 5, is the set that maximizes the value of the penalized likelihood function. Describing and interpreting the estimation results for the Row subjects is somewhat more delicate than for the Column subjects because Row subjects visit two of the second-period information sets only rarely (the two that follow play of Right by Column, corresponding to the fourth and sixth digits of the Row strategies). Consequently, estimation of these two digits is not very reliable, and we will therefore be circumspect in making distinctions between rules that differ only in these two digits.

All but seven of the 48 Row subjects are estimated to have used decision rules of the form 10^{****}, rules which prescribe first-period play that is the Row player's part of the pure strategy Bayesian equilibrium. However, just as we saw for the Column subjects, the data clearly reject the hypothesis of homogeneity across Row subjects: the value $\lambda = 214$, for $|K| = 2$ in Table 5, yields a p-value of approximately $1 - 10^{-43}$, so that at any reasonable level of significance the homogeneity hypothesis would be rejected in favor of heterogeneity. Thus, the Row subjects' second-period play must have been heterogeneous.

The shaded region of Table 5 suggests, at first glance, that the Row subjects' play cannot be described by just two decision rules, as we found for the Column subjects. However, note that rules 41 and 44 differ only at the two rarely visited information sets. Rule 41 prescribes best response at the second period, and rule 44 prescribes "choose the low-payoff action," which coincides with best response at the two second-period information sets where nearly all play occurred (the sets visited when the Column player plays his part of the pure strategy Period 1 BE). And rules 19 and 35 coincide exactly with one another at the second period: each prescribes "choose the high-payoff action" at Period 2. Rule 35 is one of the four strategies in which the Row player is playing his part of the unique pure strategy Jordan Learning Sequence. Seven of the Row subjects are estimated, however, to have used rule 19, in which first-period play is opposite to the pure strategy Period 1 BE. Summarizing, just as for the Column subjects, only two rules for second-period play are required for describing the Row subjects' play, and the rules each have a simple cognitive interpretation: "choose the high-payoff action" in one rule, and "choose the low-payoff action" in the other.

7. Equilibrium in the extensive-form game

In this section we provide a game theoretical explanation of the heterogeneous behavior we have documented in the preceding sections. We take a conventional game-theoretic approach, modelling the two-period learning environment in our experiment as an extensive form game and then investigating the game's equilibria to determine whether any of them could have generated the observed data.

The game form in Fig. 3 is a complete description of the two-period learning environment—the twice-played game of incomplete information—except that it contains no payoff information. We assume from now on that a player's payoff at any terminal node in Fig. 3 is the sum of his payoffs at the first- and second-period plays on the path to that node. Fig. 3 thus becomes an extensive form game. (In effect, we are assuming that players do not discount or otherwise weight differently the payoffs they receive at the two periods.)

It is worth pointing out how this is different from the Jordan model, which does not treat the twice-played game of imperfect information as a game in its own right, and does not include a notion called "equilibrium." The Jordan model does not make use of any payoffs at the terminal nodes, but instead assumes that players' decisions are based only on "myopic" payoffs at each of the two stages of play. (In effect, Jordan's players are assumed to look backward but never forward.)

Analyzing the set of equilibria of the game in Fig. 3 is difficult, but we take three steps that simplify the analysis. First, we work with the game's reduced normal form. Second, we concentrate only on equilibria in which first-period play constitutes a Bayesian equilibrium, as the first-period play of our subjects does. And third, we establish that each equilibrium of this kind is equivalent to a Jordan learning sequence, so that we can analyze the equilibria by analyzing Jordan learning sequences.

Recall first of all that each of the $2^6 = 64$ strategies we identified for each player in Section 4, i.e., each element of the set S, is an equivalence class of 16 strategies that play identically in the game depicted in Fig. 3, the game in which each player has 10 information sets. These equivalence classes do not depend in any way on payoffs. But now that we *have* introduced payoffs, those same equivalence classes consist of exactly the strategies that are *payoff*-equivalent—in other words, they are the strategies for the game's reduced normal form. We denote by G the 64×64 reduced normal form game.

The game G has a large and complex set of equilibria. But recall that the experimental subjects played predominantly the unique pure strategy Bayesian equilibrium at Period 1. We therefore restrict our attention to the strategies in which a player's first-period play is his part of the pure-strategy Period 1 Bayesian equilibrium—in other words, to Row strategies of the form 10^{****} (i.e., Bottom if Type 0, Top if Type 1) and Column strategies of the form 00^{****} (i.e., Left in both types). We will refer to these strategies, and to profiles of such strategies, as *PS1 strategies* and *PS1 profiles* (PS1 for "pure strategy at Period 1"). Restricting ourselves to PS1 equilibria of G enables us to carry out the analysis in terms of Jordan learning sequences.

Table 6
Possible period 2 play in PS1 Jordan learning sequences

When Beliefs are $\pi_{R0} = 1$ and $\pi_{C0} = 0.5$ (i.e. when the true game is 00 or 01)					When Beliefs are $\pi_{R0} = 0$ and $\pi_{C0} = 0.5$ (i.e. when the true game is 10 or 11)				
	Strategy profile					Strategy profile			
Strategy	σ_{R0}	σ_{R1}	σ_{C0}	σ_{C1}	Strategy	σ_{R0}	σ_{R1}	σ_{C0}	σ_{C1}
P0	0	—	1	0	P1	—	1	0	1
M1	2/3	—	0	2/3	M1	—	2/3	0	2/3
M2	1/3	—	2/3	0	M2	—	1/3	2/3	0

The following proposition establishes that indeed each PS1 equilibrium of G is a Jordan learning sequence.

PS1 equilibrium proposition[13]: Every PS1 equilibrium (σ_R, σ_C) of G generates a Jordan learning sequence—i.e., the play at Period 2 in (σ_R, σ_C) is a Bayesian equilibrium of the single-period game defined by stage game payoffs combined with beliefs that are updated from initial beliefs by applying Bayes' Rule to Period 1 play that is the (unique) pure strategy Bayesian equilibrium in the Period 1 game.

With this proposition in hand, we do not have to solve directly for all the PS1 equilibria of G. We can instead proceed as follows. We first find all the Jordan learning sequences for G that have pure strategy play at Period 1; it will turn out that there are nine such JLSs. For each of these nine Jordan learning sequences, we will determine which profiles of pure or mixed strategies (mixtures over S) yield that JLS. According to the PS1 Equilibrium Proposition, these are the only profiles which could be PS1 equilibria of G. Finally, we check the strategy profiles we have obtained to determine whether they are in fact equilibria of G.

We begin, then, by determining all the Jordan learning sequences that have pure strategy play at Period 1. Recall from Section 1 that after the pure-strategy BE is played at Period 1, the Period 2 beliefs that are generated by Bayesian updating are as follows: Column is fully informed of Row's type, but Row continues to assign probability 1/2 to each of Column's types. Thus, there are two possible incomplete-information games at Period 2, and which one actually occurs depends upon whether Row's true type is Type 0 or Type 1. A PS1 Jordan learning sequence is therefore completely described by specifying the equilibria that are played in these two information conditions. In each case, there are three possible equilibria—one in pure strategies and two in mixed strategies. The equilibria are described in Table 6: the set E_0 of possible equilibria when Row is Type 0 consists of a pure strategy equilibrium, which we denote by P0, and two mixed strategy equilibria, which we denote by M1 and M2; and the set E_1 of possible Period 2 equilibria when Row is Type 1 also

[13] The strategy profiles (σ_R, σ_C) in the proposition are profiles of mixtures over the 64 pure strategies in S; σ_i is Player i's mixed strategy. Formal definitions of Jordan learning sequence and PS1 equilibrium, and a proof of the proposition, can be found at http://www.u.arizona.edu/˜mwalker/ps1proof.pdf.

consists of one pure strategy equilibrium, denoted P1, and the same two mixed strategy equilibria, M1 and M2, that are possible when Row is Type 0.

Thus, the nine Jordan learning sequences that have pure strategy play in Period 1 correspond to the nine elements (e_0, e_1) in the set $E_0 \times E_1$, as described in Table 7, where each row describes one of these nine JLSs. In order to identify all the equilibria of *G*, then, the PS1 Equilibrium Proposition tells us that we simply need to determine, for each row of Table 7, the set of equilibria that correspond to that row's JLS. Each row of Table 7 describes all the strategy profiles in *G* that yield the JLS in that row. There are four pure strategy profiles in the first row: the Column player plays strategy 6 (binary 000110), and the Row player plays any one of the pure strategies 34, 35, 38, or 39 (binary 100*1*), each of which plays the same against Column's strategy 6. Each of the remaining eight rows of Table 7 describes all the strategy profiles in *G* that yield one of the mixed strategy JLSs.

There is only one pure-strategy Jordan learning sequence, the one in the first row of Table 7—the only JLS in which each player plays a pure strategy in *each* of the two possible Period 2 games, as well as in Period 1. Moreover, it is easily verified in the 64×64 reduced-normal-form payoff table that these four strategy profiles are the *only* pure strategy equilibria of the game *G*. We have already observed that while our subjects played according to this pure strategy JLS in Period 1, they did *not* play in accordance with it in Period 2. Therefore, if we wish to find an equilibrium explanation for the observed data, we must consider mixed strategies.

While almost every subject was estimated in Section 6 to be rather consistently playing one of his 64 pure strategies during the final 25 regimes, the *proportions* of the subjects who were estimated to be playing the various pure strategies are very close to the mixture probabilities in the Jordan learning sequence (M1,M2), the seventh row of Table 7. Table 5 indicates that 32 of the Row subjects played either strategy 41 or 44 (exactly 2/3 of the 48 Row subjects), and that the remaining 16 subjects played strategies 19 or 35 (strategy 19 differs from 35 only at the first period: seven subjects were estimated to be playing the non-PS1 strategy 19). Table 4 indicates that 17 Column subjects played strategies that coincide with strategy 0 at Period 2; 20 played strategy 9; and the remaining 11 subjects played strategies that could be said to be a hybrid of strategies 0 and 9.

8. Bayesian statistical analysis of the equilibria

In this section we carry out a Bayesian statistical analysis to determine which of the equilibria of the game *G* correspond most closely to the experimental data. We find that the Jordan learning sequence (M1,M2) receives virtually all of the posterior probability; all the other equilibria taken together receive only the most minuscule probability.[14]

[14] Recall that the subjects' interactions were governed by random rematching between plays of the game *G*. Thus, if the proportions of Row subjects and Column subjects using various pure strategies coincide with a mixed strategy equilibrium of *G*, then each subject would find that the pure strategy he is using maximizes his expected payoff, as if he were playing against the mixed strategy of a single "representative player."

Table 7
PS1 Jordan learning sequences and equilibria

BE at period two:			Row player			Column player			
	If $\pi_R = 0$	If $\pi_R = 1$	Behavior strategy[a]	Mixed strategy Support set	Mix	Behavior strategy[a]	Mixed strategy Support set	Mix	Number of extreme points
1.	P1	P2	100*x*1*x* :	{34,35,38,39}	1	000110 :	{6}	1	4
2.	P1	M1	100*x*b*x* :	{32,33,36,37}	1/3	0001b0 :	{4}	1/3	16
				{34,35,38,39}	2/3		{6}	2/3	
3.	P1	M2	100*x*a*x* :	{32,33,36,37}	2/3	00b100 :	{4}	1/3	16
				{34,35,38,39}	1/3		{12}	2/3	
4.	M1	P2	10b*x*1*x* :	{34,35,38,39}	1/3	00001b :	{2}	1/3	16
				{42,43,46,47}	2/3		{3}	2/3	
5.	M2	P2	10a*x*1*x* :	{34,35,38,39}	2/3	000b10 :	{2}	1/3	16
				{42,43,46,47}	1/3		{6}	2/3	
6.	M1	M1	10b*x*b*x* :	{32,33,36,37}	1/3	000bb0 :	{0}	1/3	16
				{42,43,46,47}	2/3		{3}	2/3	
7.	M1	M2	10b*x*a*x* :	{34,35,38,39}	1/3	00b00b :	{0}	1/3	16
				{40,41,44,45}	2/3		{9}	2/3	
8.	M2	M1	10a*x*b*x* :	{34,35,38,39}	2/3	000bb0 :	{0}	1/3	16
				{40,41,44,45}	1/3		{6}	2/3	
9.	M2	M2	10a*x*a*x* :	{32,33,36,37}	2/3	00bb00 :	{0}	1/3	16
				{42,43,46,47}	1/3		{12}	2/3	

[a] Each digit in a behavior strategy gives the mixture probability on Bottom for Row (Right for Column) at that digit's information set.
The symbols a and b are shorthand for the probabilities 1/3 and 2/3, respectively.
The value of x is the weighted average of the pure strategies' values at this digit (0 or 1), where the weights are the "Mix" column entries.

Let E be the set of PS1 equilibria of the extensive form game G, as described in Section 7, with generic element e. Recall that E is partitioned into nine equivalence classes, and that each equivalence class is a Jordan learning sequence, as in Table 7. Strictly speaking, the four PS1 equilibria (the equilibria in the first JSL) all require pure strategy play at Period 1. And while most of the Period 1 actions our subjects chose were the pure strategy actions, other actions were chosen occasionally. Thus, if we don't allow for errors in subjects' behavior, then every equilibrium in E will receive a likelihood value of zero, and posterior probabilities would be undefined. We therefore introduce a non-strategic error structure; the error rates are of course unknown to us and must be estimated.

Our specification of the error structure assumes that when a player reaches an information set, there is a positive probability (the error rate) that instead of following the equilibrium's prescription, the player randomly selects one of his two possible actions with equal probability. This assumption ensures that every information set can be reached with positive probability, thereby overcoming the zero-likelihood problem. In order to reduce the computational size of the resulting estimation problem, we posit three error rates instead of six for both the Row player and the Column player. Each of a player's error rates is associated with two of the player's information sets: ε_{R1} is the Period 1 error rate for the Row player, the same error rate for each of his two types (thus, ε_{R1} applies to the first two digits of his strategy); ε_{R2} is the Row player's Period 2 error rate when he is Type 0, the same error rate whichever action he has observed his opponent choose in Period 1 (thus, ε_{R2} applies to the third and fourth digits of the player's strategy); and ε_{R3} is the Row player's Period 2 error rate when he is Type 1 (so that ε_{R3} applies to the fifth and sixth digits of his strategy). The Column player's error rates are specified in the same way.

With this error structure, we can now specify a prior probability distribution over the set $\Theta = E \times [0,1]^6$. We use a uniform prior over Θ: the prior probabilities on the error rates are thus independent, and independent of the prior probabilities of the equilibria; but because of the structure of the set of equilibria, the specification on E requires a bit more care. Recall (cf. Table 7) that there are only four pure strategy equilibria (the first row of Table 7), but that each of the remaining Jordan Strategy Processes (the eight remaining rows of Table 7) corresponds to a six-dimensional continuum of equilibria. A continuous prior over E would therefore place zero prior probability on the four pure strategy equilibria—the pure strategy equilibria would be ruled out a priori. Thus, instead of assigning equal prior probability to each of the infinity of equilibria in E, we instead assign equal prior probability to each equivalence class of equilibria, i.e., to each of the nine JLSs in Table 7. For computational purposes, we replace each continuum (the continuum sets of equilibria, and the error rate intervals $[0,1]$) with uniform finite grids, or lattices. All calculations will therefore involve sums instead of integrals.

We use the data from the last half of the experiment to update the prior we have specified. For our purposes here, the data set is summarized by the frequency distribution of play across the 64 paths, or terminal nodes v, in the extensive form game G. Thus, the data set is described by a list $x = (x_v)_{v \in \mathcal{N}}$ of 64 non-negative integers, where $\mathcal{N}$ is the set of all 64 terminal nodes and where the component x_v indicates how many times in the data the terminal node v was reached. It is convenient to represent the terminal nodes v as pairs (g,h), where g is the game that was drawn and h is the history of play that occurred. Thus, $g \in \mathcal{G} = \{00, 01, 10, 11\}$, where the two components or digits of g indicate the Row and Column players' types; and $h \in \mathcal{H} = \{0000, 0001, \ldots, 1111\}$, where the first two digits of h indicate the Row and Column subjects' observed choices at Period 1, and the third and fourth digits indicate the corresponding play at Period 2.[15] Thus, $x = (x_v)_{v \in \mathcal{N}} = (x_{g,h}) \in \mathbb{N}^{4\times 16} = \mathbb{N}^{64}$, and we write $x_g = (x_{g,0000}, x_{g,0001}, \ldots, x_{g,1111})$ for each $g \in \mathcal{G}$.

We now have (a) a set of potential data-generating models, $E \times [0,1]^6$; (b) a prior belief over the models; and (c) a set of data with which to update our belief. Carrying out the updating to obtain a posterior belief is conceptually straightforward: we simply apply Bayes' Rule. Specifically, given the data x, the posterior probability of any $\theta^* = (e^*, \varepsilon^*) \in E \times [0,1]^6$ is

$$\Pr(\theta^*|x) = \frac{\Pr(x|\theta^*)\Pr(\theta^*)}{\sum_{\theta \in \Theta} \Pr(x|\theta)\Pr(\theta)},$$

[15] Recall that 0 denotes play of T by Row, or L by Column, and 1 denotes play of B by Row or R by Column.

where $\Pr(\theta)$ is the prior probability of θ and $\Pr(x|\theta)$ is the likelihood value for x and θ, the probability of observing x conditional upon the model θ. Because we have assumed a uniform prior on the error rates, $\Pr(\varepsilon)$ is the same for every value of ε; moreover, the priors on error rates ε and equilibria e are independent. We therefore have

$$\Pr(\theta^*|x) = \frac{\Pr(e^*)\Pr(x|\theta^*)}{\sum_{\theta\in\Theta}\Pr(e)\Pr(x|\theta)}.$$

Since the denominator has the same value for every θ^*, we have $\Pr(\theta^*|x) \propto \Pr(e^*)\Pr(x|\theta^*)$, and we therefore need to evaluate $\Pr(e^*)\Pr(x|\theta^*)$ for each value of θ^*. The term $\Pr(e^*)$ is simply the prior probability of e^*; we have described above how these probabilities are specified. In order to evaluate the term $\Pr(x|\theta^*)$, we identify, for each game $g\in\mathcal{G} = \{00, 01, 10, 11\}$, the number of times in the data that the game g was drawn, which we denote by n_g. We take the total number of draws, n, to be fixed, so we have $n_{00} + n_{01} + n_{10} + n_{11} = n$. Thus (and henceforth writing θ instead of θ^*),

$$\Pr(x|\theta) = \Pr(n_{00}, n_{01}, n_{10}, n_{11}) \cdot \prod_{g\in\mathcal{G}} \Pr(x_g|n_g, \theta).$$

Since $\Pr(n_{00}, n_{01}, n_{10}, n_{11})$ is independent of θ, we have

$$\Pr(x|\theta) \propto \prod_{g\in\mathcal{G}} \Pr(x_g|n_g, \theta).$$

For each $g\in\mathcal{G}$, we have

$$\Pr(x_g|n_g, \theta) = M(x_g) \cdot \prod_{h=0}^{15} \Pr(h|g, \theta)^{x_{gh}},$$

where $M(x_g)$ is the multinomial term

$$M(x_g) := \frac{n_g!}{x_0!x_1!\cdots x_{15}!}.$$

Since $M(x_g)$ does not depend upon θ, we have

$$\Pr(x_g|n_g, \theta) \propto \prod_{h=0}^{15} \Pr(h|g, \theta)^{x_{gh}},$$

and we need to evaluate the terms $\Pr(h|g, \theta)$. Assuming for the moment that all error rates are zero, we would have the following, where the $\eta\in v$ are the information sets on the path $v = (g, h)$, and where we take η to include the specification $\rho(\eta)$ of which player is to play at η:

$$\Pr(h|g, \theta) = \prod_{\eta=(g,h)} \beta_{\rho(\eta)}(\eta),$$

where $\beta_{\rho(\eta)}$ is the behavior strategy of player $\rho(\eta)$—specifically,

$$\beta_{\rho(\eta)}(\eta) = \sum_{s\in S} \mu_{\rho(\eta)}(s, \theta) \cdot s_\eta,$$

where $\mu_{\rho(\eta)}(s,\theta)$ is the mixture probability player $\rho(\eta)$ uses for pure strategy s under the mixed strategy equilibrium specified by θ. (Note that for every pure strategy $s \in S$, the value of the component s_η is either 0 or 1, because s_η is one of the six digits of the binary string s.) Thus, $\beta_{\rho(\eta)}(\eta)$ is a number in the real unit interval, the mixture probability with which the player who chooses at η is directed to choose B (if he is Row) or R (if he is Column). Note that on each path v there are exactly four information sets: one for each player at each period.

Next we incorporate the error rates into the calculations. In a slight abuse of notation, we allow ε_η to represent the error rate corresponding to the information set η. In order to account for the vector ε of error rates in $\theta = (e, \varepsilon)$, the mixture probabilities $\beta_{\rho(\eta)}(\eta)$ must be replaced with the "error-adjusted" probabilities

$$\tilde{\beta}_{\rho(\eta)}(\eta) = (1 - \varepsilon_\eta)\beta_{\rho(\eta)}(\eta) + \frac{1}{2}\varepsilon_\eta.$$

Gathering together all the preceding steps, the posterior probabilities $\Pr(\theta|x)$ satisfy

$$\begin{aligned}\Pr(\theta|x) &\propto \Pr(e) \prod_{v \in \mathcal{N}} \left\{ \prod_{\eta \in v} \left[\frac{1}{2}\varepsilon_\eta + (1 - \varepsilon_\eta) \sum_{s \in S} \mu_{\rho(\eta)}(s,\theta) \cdot s_\eta \right] \right\}^{x_v} \\ &= \Pr(e) \prod_{v \in \mathcal{N}} \prod_{\eta \in v} \left[\frac{1}{2}\varepsilon_\eta + (1 - \varepsilon_\eta) \sum_{s \in S} \mu_{\rho(\eta)}(s,\theta) \cdot s_\eta \right]^{x_v}.\end{aligned}$$

Calculation of the posterior probabilities from our experimental data provides extremely strong evidence that only one of the nine equivalence classes of equilibria is consistent with the data. The posterior probabilities on eight of the JLSs in Table 7, including the pure strategy equilibrium in the table's first row, all have orders of magnitude less than 10^{-50}. Virtually all posterior probability is on the mixed strategy JLS (M1,M2), the JLS on line 7 of Table 7.

The posterior distribution depends, of course, on the prior distribution we specify. But this posterior distribution is quite robust to reasonable variation in the prior, because, as shown above, the posterior probabilities are proportional to the prior probabilities. For example, in order to bring the posterior probability of (P0,P1), the pure strategy JLS, to the same order of magnitude as the posterior probability of (M1,M2)—so that each would have about half of the posterior probability—one would have to have a prior belief that (P0,P1), the pure strategy JLS, is at least 10^{104} times more likely to be the correct model of the subjects' behavior than the mixed strategy JLS (M1,M2). The posterior probabilities of the various equilibria are quite robust to variations in the prior distribution over the error rates as well.

Thus, our Bayesian analysis of the equilibria of the game G strongly confirms the informal observation we were led to by our estimation of the subjects' behavioral learning rules, or strategies: the data look very much like the mixed strategy equilibrium (M1, M2), and no other equilibrium (including the pure strategy equilibrium) provides a viable explanation of the data. An important caveat must be emphasized, however: we placed all the prior probability on *equilibria* of the game G. In this respect we are adopting the fully game-theoretic view of Section 7, that we should expect the subjects to play an equilibrium of the game. Within that context, the analysis of this section shows that only the mixed strategy equilibrium (M1, M2) could plausibly describe the subjects' play.

9. Concluding remarks

Several avenues for further study are suggested by the research we have reported here. The simple and systematic structure of the heterogeneity that emerged from our experiment suggests that it will be important, in developing the theory of learning in games, to take explicit account of heterogeneity—either by assuming that players will behave heterogeneously, or, ideally, by deriving the heterogeneity as an implication of the theory. Moreover, the evidence that some of our subjects' learning strategies changed over time (cf. footnote 12), and that the changes were predominantly toward the equilibrium (M1, M2), suggests a theory about "learning how to learn" in which the players' learning strategies evolve over time.

The experiment we carried out was well designed for deciding among alternative learning models in games, and for estimating the characteristics of heterogeneity in learning, but it was relatively ineffective for studying the dynamics of learning in the extensive form game, largely because each observation provided data about only a few of the game's information sets. An experiment designed expressly to tell us about this dynamic process would be important.

The equilibrium that eventually emerged so clearly from our analysis, and which coincided with a Jordan Learning sequence, was in mixed strategies, but appeared via pure strategy play in a population of subjects who were repeatedly rematched to play against one another. This suggests an experiment in which subjects are not rematched but play many regimes in fixed pairings. The issue of myopic play would have to be addressed in designing such an experiment, and it is not clear a priori whether we should expect learning behavior in this context to be homogeneous (i.e., pure strategy play), nor is it clear whether the learning strategies should be expected to constitute an equilibrium, as they apparently did in our experiment.

Acknowledgments

We thank the University of Arizona's Economic Science Laboratory for providing the facilities and staff support necessary to conduct our experiment; and the National Science Foundation, which provided financial support.

References

[1] R.T. Boylan, M.A. El-Gamal, Fictitious play: a statistical study of multiple economic experiments, Games Econom. Behav. 5 (1993) 205–222.
[2] C.F. Camerer, T. Ho, Experience-weighted attraction in games, Econometrica 67 (1999) 827–874.

[3] Y.-W. Cheung, D. Friedman, Individual learning in normal form games: some laboratory results, Games Econom. Behav. 19 (1997) 46–76.
[4] J. Cox, J. Shachat, M. Walker, Experimental test of Bayesian learning in games, Games Econom. Behav. 34 (2001) 11–33.
[5] M.A. El-Gamal, D.M. Grether, Are people Bayesians? Uncovering behavioral strategies, J. Amer. Statist. Assoc. 90 (432) (1995) 1137–1145.
[6] I. Erev, A.E. Roth, Predicting how people play games: reinforcement learning in experimental games with unique, mixed strategy equilibria, Amer. Econom. Rev. 88 (1998) 848–881.
[7] J.S. Jordan, Bayesian learning in normal form games, Games Econom. Behav. 3 (1991) 60–81.
[8] E. Kalai, E. Lehrer, Rational learning leads to Nash equilibrium, Econometrica 61 (1993) 1019–1045.
[9] D. Mookherji, S. Barry, Learning and decision costs in experimental constant sum games, Games Econom. Behav. 19 (1997) 97–132.
[10] E. Roth, E. Ido, Learning in extensive-form games: experimental data and simple dynamic model in the intermediate term, Games Econom. Behav. 8 (1995) 164–212.
[11] D.O. Stahl, Rule learning in symmetric normal-form games: theory and evidence, Games Econom. Behav. 32 (2000) 105–138.