GAMES AND ECONOMIC BEHAVIOR **19,** 46–76 (1997)
ARTICLE NO. GA970544

Individual Learning in Normal Form Games: Some Laboratory Results*

Yin-Wong Cheung and Daniel Friedman

Economics Department, University of California, Santa Cruz, California 95064

Received April 21, 1995

We propose and test a simple belief learning model. We find considerable heterogeneity across individual players; some players are well described by fictitious play (long memory) learning, other players by Cournot (short memory) learning, and some players are in between. Representative agent versions of the model fit significantly less well and sometimes point to incorrect inferences. The model tracks players' behavior well across a variety of payoff matrices and information conditions. *Journal of Economic Literature* Classification Numbers: C72, C73, C92, D83. © 1997 Academic Press

1. INTRODUCTION

Cournot (1838) introduced the first explicit model of learning in games. He assumed that players choose a best response to what they most recently observed. The fictitious play model of Brown (1951) and others makes the somewhat more appealing assumption that players choose a best response to the average of all previous observations, not just the most recent. Both of these learning models have leading roles in recent theoretical treatments of learning in games (e.g., Fudenberg and Levine, 1995). Empirical examination of the two models therefore is in order, whether or not the original authors ever intended to describe behavior literally.

Boylan and El-Gamal (1993) showcase their Bayesian methodology by comparing the two learning models in nine laboratory game sessions. They

* First draft, April 1993. We are grateful to the U.S. National Science Foundation (Grant SES-9023945) for funding the experiments reported here and to Carl Plat for excellent research assistance. Tim Kolar wrote the key computer programs, Thanh Lim helped revise them, and Nicole Bouchez helped with the final data analysis. Audiences at Caltech, UC Irvine, University of Minnesota, University of Pittsburgh, University of Texas, and the 1994 Public Choice Society meetings provided many useful comments. We are indebted to two patient and careful referees and an associate editor of this journal for helping us improve the exposition. E-mail: dan@cash.ucsc.edu.

0899-8256/97 $25.00

conclude that, although the Cournot model does better in a few sessions, the fictitious play model overall offers a far better explanation of the data.

The Boylan and El-Gamal results raise a host of new questions. Do some payoff matrices encourage players to take a short (Cournot) view while other matrices encourage a longer (fictitious play) view? Or are we seeing individual player differences at work, the majority usually favoring longer views? Are institutional details important, such as what players can observe of other players' strategies? Such questions are important if we want to make serious empirical use of learning models. Answers will require an extension of the Cournot and fictitious play models to allow for a wider range of learning behavior, for possible responses to institutional details, and for possible individual differences across players.

In this paper we offer an extended three-parameter learning model and confront it with a variety of laboratory environments. All games are in normal form with binary action sets, but we look at several types of 2×2 payoff matrix (or bimatrix) and vary the number of players, the matching procedure, and the amount of feedback information.

The next section develops a conceptual framework and a three-parameter model for learning processes in normal form games. It explains the need for a simple model that deals explicitly with beliefs and that tracks the distribution of players' behavior across a variety of environments. It also shows where the three-parameter model fits into the recent empirical literature. The more intricate mathematical arguments are sketched in the Appendix. Section 3 introduces the data, which are described more fully in Friedman (1996, henceforth F96) and in our working paper Cheung and Friedman (1994, henceforth CF94). Section 4 collects the main empirical results. On the whole, the distributions of fitted parameters behave sensibly. Key structural parameters remain essentially unchanged when the payoff matrix changes, but move in the appropriate direction when the information conditions change. Perhaps the most striking conclusion is that players are quite heterogeneous in crucial dimensions such as effective memory length and responsiveness to evidence. We show that the heterogeneity can affect basic inferences as well as explanatory power. A brief summary and discussion appear in the last section.

2. LEARNING MODELS

Figure 1 presents a framework for discussing learning models. Proceeding clockwise from the top, we are given a stage game with payoff function g. The stage game produces a current outcome $x \in X$ for each combination of feasible current actions $a_i \in A$ chosen by players $i = 1, \ldots, N$. In particular, the payoff to player i is a function $g(a_i, s_i)$ of her own action a_i

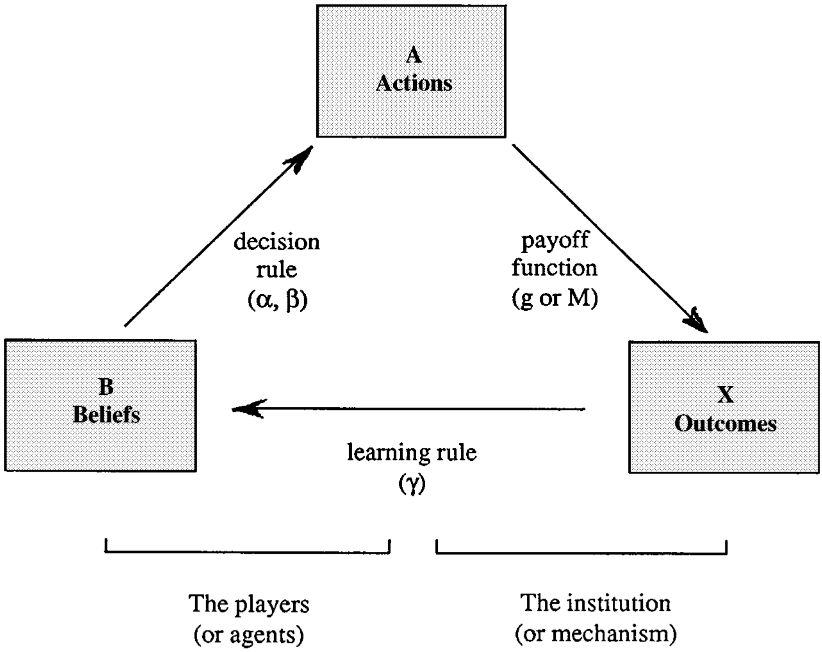

FIG. 1. Conceptual framework for learning in games.

and the payoff relevant state s_i created by other players' actions. Player i observes a portion of the current outcome, perhaps only her own payoff or perhaps the entire action combination and payoff vector, depending on the institutional arrangements of the game. To the extent that the observed outcome is not fully anticipated, the player revised her beliefs $\hat{s}_i \in B$ using some learning rule. The last leg of the diagram is a decision rule (possibly stochastic) that determines an action for each player given her current beliefs.

The stage game is repeated many times. Beliefs typically will change as experience accumulates. Changes in beliefs typically will induce changes in actions and outcomes, hence further changes in beliefs. The process may or may not eventually settle down. Theoretical literature shows under rather general conditions that if beliefs, actions, and outcomes do settle down, then the actions are mutually consistent best responses, i.e., are a Nash equilibrium of the stage game.[1] Whether the process converges to

[1] We refer here to normal form game results such as Proposition 3.3 of Friedman (1991) or several propositions of Fudenberg and Levine (1995). With extensive form stage games, behavior can settle down into self-confirming actions and beliefs that do not quite constitute a Nash equilibrium (e.g., Fudenberg and Levine, 1993).

equilibrium quickly or slowly (or not at all) depends on players' learning and decision rules and on the institution through which they interact.

2.1. *Empirical Learning Rules*

Empirical learning models must deal with the fact that beliefs are not directly observable. Hence in empirical work we must postulate a set B of possible beliefs and estimate the decision rule jointly with the learning rule. A natural specification for B is the set of action distributions the player could face, i.e., (opponents') mixed strategies. Boylan and El-Gamal use this specification to compare two venerable learning rules. The Cournot rule takes current beliefs to be the most recently observed action distribution. The fictitious play rule takes current beliefs to be the simple average of all previously observed action distributions (possibly also averaging in some unobserved prior).

We want to construct a one-parameter class of learning rules that includes Cournot and fictitious play as special cases. Fictitious play assumes that players have long memories and consider all previous observations, a not unreasonable assumption for the 10–16 period repeated trial experiments we shall examine. Cournot assumes that players pay much more attention to recent observations than to older observations. To some extent this also is reasonable, because if other players are learning then the action distribution changes over time and the more recent evidence indeed is more representative of the current state. We shall assume that each player i uses some discount factor γ_i on the older evidence. Specifically, a player who has observed the historical states $h_{it} = (s_{i1}, \ldots, s_{it})$ by the end of period t will begin the next period with beliefs

$$\hat{s}_{it+1} = \frac{s_{it} + \sum_{u=1}^{t-1} \gamma_i^u s_{it-u}}{1 + \sum_{u=1}^{t-1} \gamma_i^u}. \tag{2.1}$$

As noted in CF94, Eq. (2.1) is a slight variant on the standard partial adjustment (or adaptive) model, and the denominator simply renormalizes so that the weights on historical observations sum to 1. We apply the learning rule beginning at $t = 1$ and (to reduce the number of free parameters) we neglect any beliefs held prior to period 1.

Setting $\gamma = 0$ in Eq. (2.1) yields the Cournot learning rule, and setting $\gamma = 1$ yields fictitious play. We have adaptive learning when $0 < \gamma < 1$; in this case all observations influence the expected state but the more recent observations have greater weight. There are, of course, two other logical possibilities. Values of $\gamma > 1$ imply that older observations have greater weight. Such values would characterize a player who relies on first impressions or (like one of Konrad Lorenz' famous ducklings) is susceptible to imprinting. Values of $\gamma < 0$ are highly counterintuitive in that they imply that the influence of a given observation changes sign each period.

The data we will examine come from binary choice games (i.e., $\#A = 2$) in which each player faces a single strategically homogeneous population of n opponents. In this case the historical observations in Eq. (2.1) are just real numbers $s_{it} = n^{-1}\sum_{j=1}^{n} a_{jt} \in [0, 1]$, where $a_{jt} = 1$ if player j chooses the first action in period t and $= 0$ if he chooses the other action. The current belief $\hat{s}_{it}$, the expected fraction of opponents who will choose their first action, is a weighted average of the historically observed fractions.

2.2. *Empirical Decision Rules*

The decision rules we consider are based on differences in expected payoffs. A player with 2×2 payoff matrix $M = ((m_{ij}))$ who believes her opponent will choose his first action with probability s has expected payoff $(1, 0)M(s, 1 - s)'$ for her first action $(1, 0)$ and expected payoff $(0, 1)M(s, 1 - s)'$ for her other action action $(0, 1)$. The difference in expected payoffs (i.e., the perceived advantage to the first action) therefore is

$$R(s) = (1, -1)M(s, 1 - s)' = b + cs, \tag{2.2}$$

where $b = (1, -1)M(0, 1)'$ and $c = (1, -1)M(1, -1)' = m_{11} + m_{22} - m_{12} - m_{21}$.

Note for later reference that 2×2 matrices come in three major types, depending on the sign of c and the solution s^* to $0 = R(s) = b + cs$:

(1) $c < 0$ and $0 < s^* < 1$, in which case s^* is the unique Nash equilibrium (NE) of the symmetric game where all players face the same matrix M;

(2) $c > 0$ and $0 < s^* < 1$, in which case the NE of the symmetric game are $s = 0, 1$, and s^*; and

(3) there is no solution $s^* \in (0, 1)$, in which case either $s = 0$ or $s = 1$ is a dominant strategy and hence the unique NE of the symmetric game.

There is also the trivial case that $c = 0$ and $b = 0$, in which case a player is always indifferent between her actions.[2]

[2] The classification has a fairly long tradition and can be extended to nonsymmetric bimatrix or population games; see, e.g., Zeeman (1980) or F96 for a recent discussion. For present purposes it does not matter whether the population game is symmetric; all that matters are the given player's payoff matrix M and her beliefs $\hat{s}_{it}$ regarding potential opponents, whether or not they face the same payoff matrix.

The input to the decision rule of player i at time t is the expected advantage of the first action $\hat{r}_{it} = R(\hat{s}_{it})$, given current beliefs $\hat{s}_{it}$ from Eq. (2.1) and given the payoff matrix M defining R in Eq. (2.2). The first decision rule that occurs to most game theorists is deterministic best response, but stochastic decision rules are more natural in empirical work. Boylan and El-Gamal's decision rule is noisy best response, with noise parameter $\varepsilon \in [0, 1]$. For binary choice their decision rule is to choose action 1 with probability $1 - \varepsilon/2$ if its expected advantage $\hat{r}_{it}$ is positive, with probability $\varepsilon/2$ if $\hat{r}_{it}$ is negative, and with probability $1/2$ if $\hat{r}_{it} = 0$.

We propose a stochastic decision rule that is continuous in the expected payoff difference $\hat{r}_{it}$ and that allows for some persistent individual idiosyncracies. Let F be a convenient cumulative distribution function on $(-\infty, \infty)$ centered at 0, such as the logistic function $F(x) = (1 + e^{-x})^{-1}$ or the unit normal CDF. Then according to our decision rule the first action will be chosen with probability

$$P(a_{it} = 1 \mid \hat{r}_{it};\, \alpha_i, \beta_i) = F(\alpha_i + \beta_i \hat{r}_{it}). \tag{2.3}$$

Each player i has her own degree of responsiveness β_i to the perceived payoff advantage $\hat{r}_{it}$ and her own idiosyncratic tendency α_i to favor the first action. The larger positive value of β she uses, the more likely she is to best respond, i.e., to choose action 1 (action 0) when the perceived advantage $\hat{r}$ of action 1 is positive (negative). Her α will be positive (negative) if she is more likely to pick action 1 (action 0) when she expects both actions to give the same payoff.

The responsiveness parameter β is crucial. A fully rational player completely confident in her estimate $\hat{s}$ would choose the best response with certainty, and therefore would have an infinitely large positive β. Large values of β can create convergence problems, however. When all players face the same type 1 payoff matrix M with interior Nash equilibrium s^*, then r_{it} and $s_{t-1} - s^*$ have opposite sign. With large βs most players will best respond and the state s_t will overshoot s^*, producing unstable oscillations. The Appendix shows that the inequalities $0 < \beta < \sqrt{2\pi}/|c|$, where again c is $(1, -1)M(1, -1)'$, eliminate such overshooting for nonnegative values of the learning parameter γ when F is the unit normal CDF. Finite (but positive) estimates of β permit trembles, or "experiments" in the sense of Fudenberg and Kreps (1993), possibly reflecting a lack of confidence in the estimate of the expected state. Negative values can be interpreted either as perverse behavior or as anticipatory behavior in the sense of Selten (1991b). A player who anticipates that other players will use (2.1)–(2.3) with sufficiently high positive

βs will find it advantageous near a unique, interior NE to use a negative β. We see no rationalization for negative β when the state converges to a corner NE.

The decision parameter α is a catch-all intercept coefficient. It would, for example, pick up an average impact of prior beliefs.[3] We believe that the main influence on α is a player's idiosyncratic response to the payoff matrix. In the Hawk–Dove game described in the next section, for example, some players just like to be Hawkish and others like to be Dovish to some degree. Such idiosyncracies are especially important in the neighborhood of an interior NE s^*. Here both actions are best responses and so at $\hat{s}_{it+1} = s^*$ we have $\hat{r} = 0$ and $F(\alpha + \beta\hat{r}) = F(\alpha)$. Hence convergence to an interior NE s^* implies that $F(\alpha)$ reproduces the mixing probability s^* on average. We will test the hypothesis that players' idiosyncratic preferences adapt to a payoff matrix with an interior NE in a manner that allows convergence to the NE.[4]

2.3. *The Three-Parameter Model*

Equations (2.1)–(2.3) define an empirical model whose three parameters (α, β, and γ) can be estimated for each player from trial-by-trial decisions in binary choice games. For α we can test the point null hypotheses $\hat{\alpha} = 0$ of unbiased choice and $\hat{\alpha} = F^{-1}(s^*)$ of mixed Nash strategies. For β we have the null hypotheses $\hat{\beta} = 0$ of unsystematic choice (neither responsive to expected payoff nor anticipatory) and $\beta = \sqrt{2\pi}/|c|$ of borderline stability. More importantly, for γ we have $\hat{\gamma} = 0$ (Cournot) and $\hat{\gamma} = 1$ (fictitious play).

Unlike many authors, we allow for the possibility that players may differ. Perhaps the Harsanyi doctrine (e.g., Kreps, 1990, p. 110) is correct in that all players use the same learning rule and the same decision rule and have the same prior beliefs at birth. Even so, people have different experiences in life and therefore bring different priors to an experiment. As players, then, they may well differ in their discount factor (or effective memory length) γ, their responsiveness to (or confidence in) the perceived payoff advantage β, and their idiosyncratic preference α for the first action.

[3] Suppose that s_o is the initial prior in a sequence of periods of duration T. Then α would contain a term of the form $\beta\gamma^T R(s_o)$, where $R(s_o) = b + cs_o$ as in Eq. (2.2).

[4] That is, we will compare the empirical distribution of α_i with $F^{-1}(s^*)$ for matrices with various mixed strategy NE s^*. We do not attempt to model (or test) an adaptation process *per se* for α. Our intuition is that if there are initially too many Hawks, say, then low payoffs will quickly discourage this idiosyncratic preference and enough of them will turn Dovish to allow closer approach to the NE. See F96 for a related discussion on persistent idiosyncratic preferences.

Our main concern in empirical work will be to track the *distribution* of parameter estimates *across institutions* or environments. The reason is twofold. First, as economists we are interested in individual differences mainly for their outcome consequences, e.g., their effect on convergence to equilibrium. The consequences depend only on the population distribution; e.g., the stability of an interior NE depends on the group mean of $F(\alpha_i)$ and the bounds on the β_is. The outcomes do not depend on details that might interest psychologists such as what sort of life experiences lead to a low β_i. Second, as applied theorists we are interested in laboratory games mainly for the insight they provide into actual or possible field institutions. If changes in the institution produce unpredictable changes in the distribution of learning and decision parameters, then even good fits of laboratory data will offer little insight into field environments. For the model to be useful in practical applications, we need to be able to say something about how its parameters change as we change the payoff matrix, the information conditions, and other environmental variables.

According to the model, the payoff matrix should affect behavior only through α and the b and c coefficients in Eq. (2.2), i.e., only through a player's idiosyncratic action preferences and through the way the payoff matrix connects the expected payoff difference $\hat{r} = b + c\hat{s}$ to beliefs $\hat{s}$. The model does not allow for systematic effects of the payoff matrix on estimates of the structural parameters β and γ.

We can think of two *a priori* reasons why, contrary to the model's predictions, there might be an empirical relation between the payoff matrix M and the structural parameter β. First, there may be a selection bias in the data. We just noted that large positive values of β destabilize interior NE. Convergence to a corner NE s^*, by contrast, implies that β is large (relative to α and the payoff differential $R(s^*)$ at the corner NE) and positive.[5] Consequently β estimates may tend to be larger when the state converges to a corner NE than when it converges to an interior NE. Second, the model does not allow for anticipatory players in the sense of Selten (1991b). Such players may use negative βs when the payoff matrix has an interior NE. With these considerations in mind, the empirical prediction is that β as measured might be larger for matrices that promote convergence to corner equilibria (type 2) than in matrices that promote convergence to interior equilibria (type 1), but that γ will be unaffected by any matrix that allows for nontrivial payoff differences r.

Perhaps the most important predictions are for information conditions. We predict that γ will decrease and β will increase when the institution

[5] For example, if you regard $\bar{s} < 0.05$ as representing convergence to the NE $s^* = 0$ and use the unit normal CDF for F in (2.3), then for convergence you need the joint restriction $\alpha + \beta r^* < -1.96$, where $r^* = R(s^*) \leq 0$ since $s^* = 0$ is a NE.

provides better information on outcomes. The intuition on β is simply that people will respond more strongly to evidence when they know the evidence is higher quality. The intuition on γ is that when contemporaneous evidence is weak, people will rely more heavily on older evidence. The Appendix contains a simple theoretical sketch to support the intuition.

2.4. *Other Learning Models*

In the present paper we do not test the three-parameter model against nonnested alternatives, but a brief discussion of alternative learning models will provide some useful perspectives. In the terminology of Simon (1957) and Selten (1991a), we so far have dealt entirely with *belief learning*. The alternative is *rote learning*, in which experience affects behavior directly as, for example, in Roth and Erev (1995). In a formal sense, belief learning and rote learning are equivalent. Given a belief learning model as in Fig. 1, you can construct an observationally equivalent rote learning model by composing the decision rule with the learning rule, i.e., mapping outcomes directly to decisions by (decision rule)$\circ$(learning rule). Conversely, given a rote learning model you can construct an observationally equivalent belief learning model by defining the set B of beliefs as accumulated experience, so the "learning rule" is simply experience updating and the "decision rule" is simply the rote action mapping.

The point of belief learning may be clearer after discussing Mookherjee and Sopher (1994), who investigate a two-player repeated game ("matching pennies") using both rote and belief learning models. Their belief learning model is similar to ours, except that it imposes the representative agent restrictions $\alpha_i = \alpha_o$ and $\beta_i = \beta_o$ and the fictitious play restriction $\gamma_i = 1$ for all i. For the stochastic decision rule they use the logit function $F(x) = (1 + e^{-x})^{-1}$, while we will use the probit function $F(x) =$ cumulative unit normal distribution function. Their empirical results (confirmed in our data) support a natural symmetry in the coefficients that allows their main rote learning model (their Eqs. (2.1) and (2.2)) to be expressed as

$$\Pr[a_{it} = 1] = F(\alpha + \beta \tilde{r}_t), \tag{2.4}$$

where $\tilde{r}$ is not the difference in expected payoff, but rather the difference in average payoff actually experienced so far. Mookherhjee and Sopher recognize that a different version of rote learning (their Eqs. (2.3)–(2.6)) is equivalent to belief learning. To see that (2.4) above also is a special case of our belief learning model, assume for the moment (very counterfactually!) that the player somehow managed to play both strategies each period. Then

$$E\tilde{r}_t = (t-1)^{-1} \sum_{u=1}^{t-1} (1, -1) M(s_{it-u}, 1 - s_{it-u})'$$
$$= (1, -1) M(\hat{s}_{it}, 1 - \hat{s}_{it})' = \hat{r}_{it}, \qquad (2.5)$$

where $\hat{s}_{it} = (t-1)^{-1} \sum_{u=1}^{t-1} s_{it-u}$ is the expected state under fictitious play. Now drop the counterfactual and note that the player i's choice of action 0 or 1 each period defines two disjoint subsamples of $\{s_{t-u}: u = 1, \ldots, t-1\}$ from which she estimates $\tilde{r}$, but the iid assumption underlying fictitious play implies that the subsamples and the estimates are unbiased. Hence we still have $E\tilde{r}_t = \hat{r}_{it}$ so up to some noise (which can be absorbed into F) Eq. (2.4) above reduces to our Eq. (2.3). With a bit more work, the conclusion extends to values of γ other than 1.0. The underlying reason is the invertible linear relation $r = b + cs$ between expected states and payoff differences, assuming $0 \neq c = (1, -1)M(1, -1)' = m_{11} + m_{22} - m_{12} - m_{21}$. In this case, forming beliefs about states is equivalent to forming beliefs about payoff differences directly.[6]

Rote learning and belief learning part company following any change in the institution. Suppose, for example, that action 0 is a dominant strategy before a change in the payoff matrix and action 1 is dominant after the change. Then a researcher using a rote model faces an acute dilemma. If he sticks with the old parameter estimates, then he has a parsimonious but probably grossly inaccurate prediction of postchange human behavior. (Perhaps the prediction will be accurate for ants.) If he drops the old estimates, then he has no prediction at all until sufficient data accumulates. By contrast, a researcher who has properly specified beliefs B in a belief learning model has predictions that are just as sharp and (one hopes) just as accurate after the switch as before. In our three-parameter model, for example, beliefs s_{it} refer to the fraction of players expected to choose (say) the dominant action, and the coefficients b and c in Eq. (2.2) can pick up any more subtle changes in the payoff matrix. The structural parameters β and γ remain unchanged.[7]

[6] Roth and Erev (1995) and many other rote learning models are based on payoff levels rather than payoff differences, in which case an equivalent belief learning model will look quite different than ours. Rote learning models based on payoff levels implicitly assert that *nonsalient* changes in payoffs can substantially affect behavior. For example, if each player always receives an extra dollar each period, independent of actual decisions, then such models predict noisier behavior (closer to uniform random over the action set).

[7] The intercept parameter α also remains unchanged in this example, because there is no interior NE before or after the switch. In general, the natural null hypotheses are as stated in the beginning of the last subsection: $\alpha = 0$ or (if there is an interior NE s^* then also) $\alpha = F^{-1}(s^*)$. Using the second hypothesis, we predict a shift in α when the interior NE shifts.

To make the same point a different way, belief learning models seek "deep parameters" that remain invariant under institutional change. Rote learning models seek only convenient reduced forms.[8] Hence, for the reasons discussed in the previous subsection, the applied economist who wants to predict behavior in field environments will find belief learning models more useful than rote learning models.

McKelvey and Palfrey (1995) independently develop a rather different empirical model with a parametrization similar to our belief learning model. For binary choice their model can be written

$$P[a_{it} = 1] = F(\beta r_t^*), \tag{2.6}$$

where $r_t^* = (1, -1)M(P[a_{it} = 1], 1 - P[a_{it} = 1])' = b + cP[a_{it} = 1]$ is the rational expectation of the payoff difference given the choice technology (2.6). In empirical work, they take the distribution F to be the logistic function, $F(x) = (1 + e^{-x})^{-1}$ for the two-action case. Obviously the intent here is not to explain how players learn to acquire mutually consistent beliefs, but rather to fit laboratory data to an equilibrium model incorporating noisy rational expectations.

Despite its different purposes, the McKelvey–Palfrey model does imply some parametric restrictions for our model, namely that there is only a single free parameter $\beta_o = \beta_i$ and that $\alpha_i = 0$ for all i. The payoff difference r^* is not estimated from historical data, but rather is the solution to the transcendental equation $r = b + cF(\beta_o r)$. Using the values $b = 4$ and $c = -6$ (from the standard Hawk–Dove matrix presented in the next section), one can verify that the implicit function $r^*(\beta)$ is strictly decreasing, that $r^*(0) = 1$ and that $\beta r^*(\beta) \to \ln 2$ as $\beta \to \infty$ so $r^*(\infty) = 0$. Our Eqs. (2.1) and (2.2) therefore do not apply but, given the iid fluctuations the model envisions around equilibrium, the most efficient estimator in (2.1) is $\gamma = 1$. Thus their model suggests the representative agent fictitious play restriction $\gamma_i = 1$ for all i.[9]

McKelvey and Palfrey use their model on games with more than two alternative actions. Their decision rule is essentially the standard logit model. Stahl and Wilson (1995) use the same decision rule in their study of

[8] Of course, sometimes a reduced form is well suited for one's purpose. For example, Roth and Erev's rote learning model is quite effective in driving home their point that payoffs to strategies not used in equilibrium can affect convergence.

[9] In their concluding remarks, McKelvey and Palfrey suggest dropping the representative agent restriction and allowing "learning," defined as an increase over time in the parameter β (or λ in their notation). In our terminology a specification such as $\beta_t = \beta_o \exp(-\delta t)$ is not even rote learning because it is not responsive to experience, just to clock time. This dynamic stochastic equilibrium model (reminiscent of the model in McKelvey and Palfrey, 1992) still has no direct role for our α or γ parameters, although it seems that efficient estimates of the state using historical data would involve γ a bit less than 1.0.

individual differences in initial beliefs. It is beyond the scope of the present paper to consider larger actions sets, but the natural generalization of our decision rule would be based on expected payoffs r_{jit} for action j assessed by player i at time t relative to some baseline payoff and would assign to each action $j \in A$ the probability $F(\alpha_{ji} + \beta_i r_{jit})/\Sigma_{k \in A} F(\alpha_{ki} + \beta_i r_{kit})$.

Finally, Crawford (1995) uses a belief learning model to investigate a class of coordination games with an ordered action set. Decisions are myopically optimal, i.e., they maximize expected current period payoff given current beliefs. Beliefs are adaptive and characterized by a parameter β similar to our γ. This parameter is assumed identical across individuals, up to idiosyncratic error with mean zero and a variance that declines over time. Crawford shows that in his setting beliefs and actions converge to one of the multiple equilibria, and that the model tracks convergence rather well across changes in the payoff matrix.

3. THE DATA

A proper test of the three-parameter learning model requires a variety of information conditions and a variety of payoff matrices. Table Ia displays the main matrices we use. The first two entries define symmetric or single population games in which all players face the same 2×2 matrix M, of type 1 (“Hawk−Dove” or HD) and type 2 (“Coordination” or Co). HD has a unique NE that calls for 2/3 of the players to choose Hawk and the other 1/3 to choose Dove in each period. Co is a bit special in that its two pure strategy NE satisfy conflicting selection criteria (Harsanyi and Selten, 1988): $s = 1$ is payoff dominant (all players get 5 per period versus 1 per period in the other pure NE), while $s = 0$ is risk-dominant (an opponent’s deviation actually increases a player’s payoff by 3 versus a decrease by 6 in the first NE). The third entry, which defines a single population game of type 3, is shown for completeness but it will not be used in the data analysis because players with a dominant strategy provide little evidence on their beliefs about their opponents’ actions.

The other two entries in Table Ia are two population games: players in one population have a payoff matrix M^1 that may differ from the payoff matrix M^2 for players in the other population. Equally important, players’ opponents always come from the other population, never their own. (Think of the two populations as row players and column players in the bimatrix game $(M^1; M^{2\prime})$.) Buyer−Seller (B−S) is analogous to Hawk−Dove in that it has a single NE in mixed strategies; the first population has mixing probability $p = 1/4$ and the second has mixing probability $q = 1/2$. We

TABLE Ia
Some Payoff Matrices

Name	Matrix	Type	NE	EE
1. Hawk–Dove (HD)	$\begin{matrix} -2 & 8 \\ 0 & 4 \end{matrix}$	1	$s = 2/3$	$s = 2/3$
2. Coordination (Co)	$\begin{matrix} 5 & -1 \\ 4 & 1 \end{matrix}$	2	$s = 2/3, 0, 1$	$s = 0, 1$
3. Weak Prisoners' dilemma (WPD)	$\begin{matrix} 4 & 0 \\ 5 & 1 \end{matrix}$	3	$s = 0$	$s = 0$
4. Buyer–Seller (B-S)	$\mathbf{M}^1$: $\begin{matrix} 2 & 0 \\ 3 & -1 \end{matrix}$ $\mathbf{M}^2$: $\begin{matrix} 2 & 3 \\ -1 & 4 \end{matrix}$	1a	$(p, q) = (1/4, 1/2)$	$(p, q) = (1/4, 1/2)$
5. Battle of the Sexes (BoS)	$\mathbf{M}^1$: $\begin{matrix} 1 & -1 \\ -1 & 2 \end{matrix}$ $\mathbf{M}^2$: $\begin{matrix} 3 & -1 \\ -1 & 1 \end{matrix}$	2a	$(p, q) = (1/3, 3/5)$, $(1, 0), (0, 1)$	$(p, q) = (1, 0)$, $(0, 1)$

Note. Matrix types are defined in the text. The NE column lists all Nash equilibria for the symmetric two-player game using that matrix. The EE column lists the NE to which convergence is supposed to be possible (see F96). The data from the third entry, a variant of the standard Prisoner's dilemma, are not analyzed in the present paper.

also examine a type 2 analogue, Battle of the Sexes (BoS), with two pure strategy NE and one mixed strategy NE. As noted in the results section below, we sometimes consider minor variants on the matrix entries listed in Table Ia.[10]

3.1. *Laboratory Procedures*

Variety in the information conditions (or the institution) is achieved via two alternative matching procedures and two alternative computer screen displays for the players. Under the *random pairwise* (RP) matching procedure, the computer randomly picks a matching scheme independently in each period, each admissible scheme being equally likely. For example, in a 2×6 Buyer–Seller game, population 1 players (the Buyers) with identification numbers 0, 1, 2, 3, 4, and 5 might be matched respectively with population 2 players (the Sellers) numbered 4, 0, 2, 1, 5, and 3 in the first

[10] The working paper CF94 also examines several other two population payoff matrices of various types. The results are omitted here to conserve space.

period and with Sellers numbered 2, 5, 0, 3, 4, and 1 in the second period. In a 1×12 Hawk–Dove game, the single population of players might be paired $\{0,2\}$, $\{1,9\}$, $\{3,11\}$, $\{4,5\}$, $\{6,10\}$, and $\{7,8\}$ in the third period.

Under random pairwise matching, a player with payoff matrix M has *expected* payoff advantage $(1,-1)M(s,1-s)'$ for the first action if the distribution of actions by potential opponents is $(s,1-s)$. However, the *actual* payoff depends on the action taken by the actual opponent, and so has some variance around its expectation. The variance is eliminated in the alternative matching procedure, called *mean matching* (MM). Here each player is matched once against each possible opponent in each round and receives the average (mean) payoff. For example, if 6 of 12 players in Hawk–Dove (the first entry of Table I) choose the first action then the state is $(s,1-s)=(0.5,0.5)$ and the payoff differential is $(1,-1)\cdot M(0.5,0.5)'=3.0-2.0=1.0$.

The other information variable is the amount of historical data that appears on each player's screen. In the *No History* (NH) condition, the player receives no historical information other than what she could tabulate herself: her own action and actual payoff in previous periods. In the alternative condition, *History* (H), the screen displays a list of the actual states of the relevant population in previous periods. For example, the player in period 4 might see that 9 of 12 players took the first action in period 1, then 8 in period 2, and 6 in period 3. Of course, the payoffs in mean matching implicitly reveal this historical information, but in random matching the information is entirely new.

Two other procedures deserve brief mention. The number of players varies across sessions—e.g., perhaps 12 players in one session and 16 in another. Some sessions employ *split groups* in some periods—e.g., all 16 players belong to a single population in the first 40 periods, then are divided into two separate 8-player groups (no pairing or mean-matching across the two groups) for the next 80 periods and reunited into a single group for the last 40 periods of a 160 period session.

Each session consists of several *runs*, sequences of periods in which the institution is held constant in order to give players the opportunity to learn. Runs are separated by obvious changes in the player population, the information and matching conditions, and/or the payoff matrix. The history screen also is restarted at the beginning of a new run. If runs are too short then there will not be enough observations to fit a learning model, but if runs are too long then players may respond to boredom rather than to payoffs. Typical run lengths are 10 or 16 periods.

A more complete description of experimental procedures can be found in F96 and CF94. Instructions to subjects and detailed lists of treatments employed in each session are available on request.

3.2. *Estimation Procedures*

The raw data consist of the actions chosen by each player in each period of each run, together with the treatments employed in the run. Commercial statistical packages applied to these data can easily estimate the decision parameters α and β in a representative agent version of our parametric model (2.1)–(2.3). The probit (or, if you prefer, the logit) regression procedures applied to any run or set of runs will generate the desired estimates once the anticipated payoff differentials $\hat{r}$ are specified. We modified the standard probit procedures for two reasons. First, for each game type (payoff matrix) we estimate a separate learning model for each subject in each session. That is, we impose the restriction that the decision and learning parameters are identical across runs within a session but may vary across individual subjects.

A second modification required more work. In Eq. (2.1) the learning parameter γ enters in a recursive and nonlinear fashion in each run. One can exploit the linearity of (2.2) to take exponential averages of r's instead of s's, and we did so to reduce computation times. We wrote a FORTRAN program that, in combination with the optimization routines in the GQOPT package, provided the necessary point estimates and standard errors, and used grid search results as well as Mathematica to confirm that the program worked properly.

Table Ib counts the total number of individuals to our experiments and the number for which we can estimate the learning model. We exclude sessions with less than three runs of the given payoff matrix and sessions

TABLE Ib
Players By Game

Game	Number of sessions	Number of players	Number of estimable players	Percentage of estimable players
Hawk–Dove (HD)	11	138	127	92
Coordination (Co)	9	110	77	70
Buyer–Seller (B-S)	8	116	92	79
Battle of the Sexes (BoS)	9	118	97	82

Note. All sessions with group size larger than six and at least three runs are included. The second column reports the number of individual subjects (players) in these sessions. The third column excludes subjects whose choices did not permit estimation of the learning model. In most cases, such individuals chose the same action in all periods, except perhaps the first.

with groups smaller than six players. (As explained in F96, some players apparently attempt to influence the behavior of other players in these small group sessions. Such attempts are outside the scope of the learning models we are presently investigating.) For example, HD runs are found in 11 different sessions involving 138 individual players. The choices of 11 of the 138 players did not permit estimation of the model, typically because all choices (except perhaps the first) in each run were the same, e.g., all "Dove." Thus we estimate the three parameter learning model for 127 of 138 individual HD subjects.

In each session and type of game we also estimate the learning model for two aggregate players. The first aggregate, which we refer to as "player 99," consists of the choices of all players in that session. The second aggregate, which we refer to as "player 98," excludes the choices of nonestimable players. Thus we have $2 \times 11 = 22$ aggregate HD players, $2 \times 9 = 18$ aggregate Co players, etc. Later we will compare learning model fits of individuals and aggregates.

4. RESULTS

The histograms in Fig. 2 show the distributions of point estimates across individual players. The α estimates always are quite dispersed and usually unimodal, with modestly positive mode in some data and modestly negative in order. The exception is in the Co data, where negative tail is truncated. It seems Co players tend to prefer the payoff dominant action to the risk dominant action. There is a sprinkling of negative β estimates in the type 1 (HD and B−S) data, but almost none in the type 2 (Co and BoS). Since we expect convergence to interior NE in type 1 but not type 2 games, the histograms suggest a sprinkling of anticipatory players. Negative γ estimates are not rare, but most estimates seem to lie between 0 and 1 as we had hoped.

The standard errors are not shown here; some estimates are quite precise but others are not. The dispersion in the histograms arises partly from imprecise estimates and partly from true heterogeneity across players. The hypothesis tests presented below will have to deal with both heterogeneity and with estimation error.

4.1. *Parameter Estimates*

Table II reports median parameter estimates across individual players. The first line (labeled 3 par) comes from the basic three-parameter model estimated from all relevant runs of each payoff matrix. The other two lines report separate parameter estimates for different information and match-

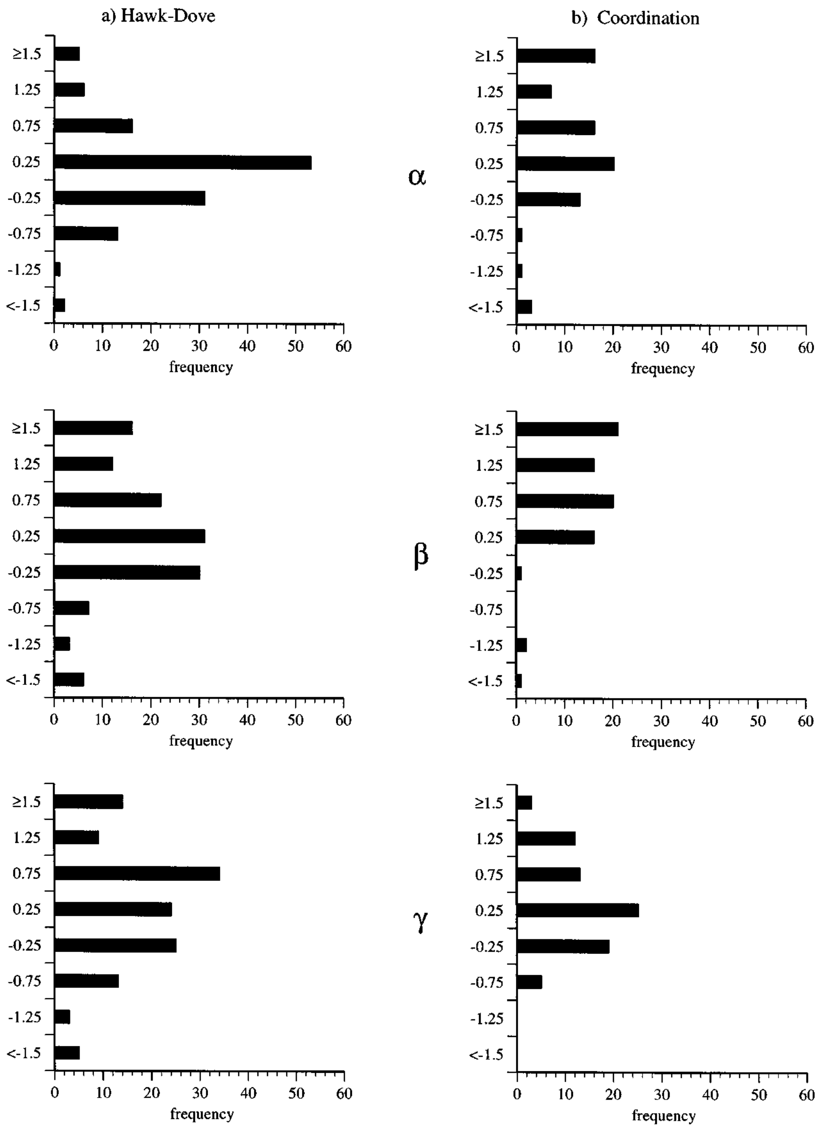

FIG. 2. Parameter histograms.

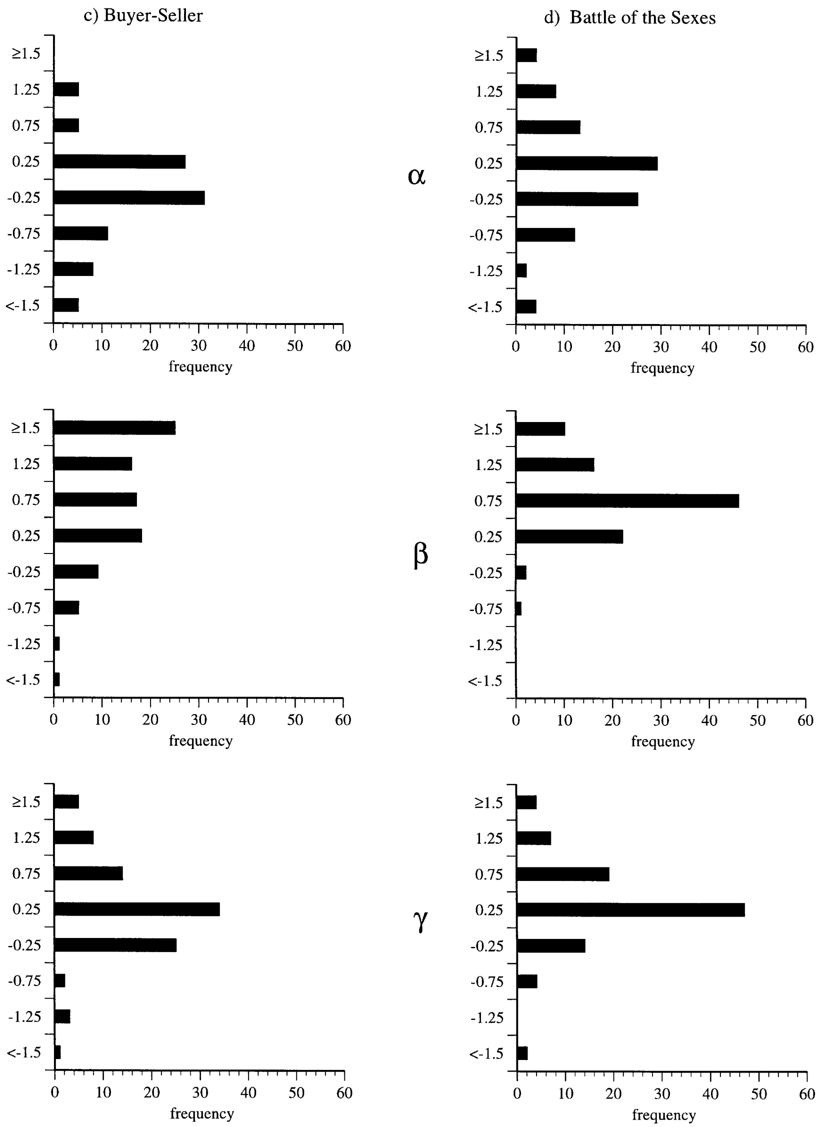

FIG. 2.—*Continued*

TABLE II
Median Parameter Estimates

	$F(\alpha)$	α	β	γ	$F(\alpha)_A$	α_A	β_A	γ_A
HD: 3 par	0.58	0.20	0.17	0.32				
HD: 6 M/R	0.63	0.32	0.56	0.46	0.51	0.02	0.29	0.39
HD: 6 H/N	0.62	0.31	0.33	0.44	0.55	0.13	−0.10	0.57
Co: 3 par	0.70	0.51	0.93	0.41				
Co: 6 M/R	0.88	1.18	1.65	0.13	0.72	0.58	0.47	0.48
Co: 6 H/N	0.65	0.40	0.90	0.25	0.64	0.36	0.38	0.40
B-S: 3 par	0.45	−0.13	0.87	0.26				
B-S: 6 M/R	0.40	−0.24	1.90	0.17	0.45	−0.14	0.55	0.50
B-S: 6 H/N	0.43	−0.18	1.87	0.18	0.48	−0.04	0.55	0.52
BoS: 3 par	0.53	0.09	0.72	0.33				
BoS: 6 M/R	0.61	0.27	0.88	0.12	0.47	−0.06	0.62	0.30
BoS: 6 H/N	0.64	0.36	0.87	0.18	0.52	0.04	0.48	0.25

Note. The number of observations used to calculate each median value are found in Table III. For the six-parameter model, the medians for Mean Matching and History are reported in the columns on the left, and the medians for Random Pairwise and No History are reported in columns on the right.

ing conditions. For example, for each player an α, β, and γ is estimated from mean matching (MM) runs and another α, β, and γ from random pairwise (RP) runs, and the medians are reported in the lines labeled 6 M/R. Similarly, the lines labeled 6 H/N reports separate estimates from History runs and from No History runs.

Table II indicates that the $\hat{\alpha}_i$'s indeed center near $F^{-1}(2/3)$ for the single-population (symmetric) games HD and Co with mixed NE $s^* = 2/3$. In the two-population games the $\hat{\alpha}_i$'s center on slightly negative values as often as positive values, especially (as expected) in B−S where the mixed NE $(1/4, 1/2)$ is at or below $F^{-1}(0) = 1/2$.

The $\hat{\beta}_i$'s mostly center below the critical value $\beta_s = \sqrt{2\pi}/|c| \approx 0.42$ for the standard HD matrix and a mostly bit above $\beta_s \approx 0.84$ for the standard Co matrix.[11] Median values of $\hat{\beta}_i$ are always positive except in the No History HD data.

[11] The relevance of β_s is less in type 2 and 2a games such as Co and BoS because their convergence is expected to the corner NE, not to the interior NE. See F96 for an explanation of this point, and for listings of slight variants on the standard matrices used in some of the runs. In any case, the comparisons of the $\hat{\beta}_i$s and β_s are tangential to the main issues and we will neglect them in the discussions below. The tables often retain the standard HD β_s value of 0.42 as a rough benchmark for interested readers.

The distribution of $\hat{\gamma}_i$s seems fairly consistent across payoff matrices. The median values generally fall in the lower half of the "adaptive" range, sometimes approaching the Cournot value of 0.0. CF94 shows that the parameter estimates are robust to specification error due to omitted variables that change slowly and finds generally similar results for the sparser data from other two population games.[12]

4.2. *Tests of Location and Treatment Effects*

The first pair of lines in each part of Table III reports tests of the point null hypotheses $F(\alpha) = 2/3$ or 0, $\beta = 0$ or 0.42, and $\gamma = 0$ (Cournot) or 1 (fictitious play). The data tested are individual players' estimated coefficients (e.g., 127 of them for HD and 77 for Co). Since the standard errors of the coefficients vary greatly, and since the distribution of estimates has some large outliers, the table relies on the simple nonparametric signs test. CF94 discusses other possible tests and points out that they yield similar results. The signs tests are powerful enough to reject the null hypotheses, α, β, and $\gamma = 0$ (in favor of the positive alternatives) at the 0.005 level or better for both the HD and the Co data. We do not reject the hypothesis that $F(\alpha)$ is centered at $2/3$ for the Co data. For the HD data we confidently infer that $F(\alpha)$ is centered between 0 and $2/3$ and that β is centered (slightly) below 0.42. The fictitious play hypothesis that γ is centered at 1.0 is rejected firmly in both data sets, so we conclude that it indeed is centered in the adaptive zone between 0 and 1.

The two-population data reinforce the main single-population results. The median β estimate is near or above 0.42 and the median γ is between

[12] Another sort of specification robustness was not discussed in CF94 but may be worth summarizing here. Given either History or Mean Matching (or both) players can infer the exact historical states s_{it-u}, but in RP/NH players observe only a (presumably unbiased) sample of the historical states. To maintain data consistency across treatments we always report parameter estimates based on the exact states (ES). But there are enough RP/NH runs in the HD data to estimate the parameters using the actual sample (AS) of observations to check robustness. There are 44 individual players whose parameters can be estimated both from ES and AS, usually on about 30 observations each. The median α estimate was 0.26 lower under AS than under ES (P value $= 0.08$), β was 0.21 higher (P value $= 0.24$), and γ was 0.05 higher (P value $= 1.0$). Changing the RP/NH specification from ES to AS therefore would have no noticeable effect on the γ estimates, would (as one might expect) increase β but not dramatically and not significantly (the one-tailed level is $0.24/2 = 12\%$) and would (idiosyncratically) increase α somewhat (given the absence of a predicted direction, the significance level is 8%). Quantitatively, the differences of AS from ES estimates seem too small to affect the inferences drawn from the histograms or tables.

TABLE IIIa
Hypothesis Tests

	Nobs	α	β	γ
Hawk–Dove				
H_o: $\alpha = 0$, $\beta = 0$, $\gamma = 0$	127	16.5	17.5	17.5
		(0.00)	(0.00)	(0.00)
H_o: $F(\alpha) = 0.67$, $\beta = 0.42$, $\gamma = 1$	127	−32.5	−10.5	−40.5
		(0.00)	(0.08)	(0.00)
H_o: $(M$ par$) - (R$ par$) = 0$	91	12.5	9.5	3.5
		(0.01)	(0.06)	(0.53)
H_o: $(H$ par$) - (N$ par$) = 0$	84	3.0	7.0	0.0
		(0.59)	(0.16)	(1.00)
Coordination				
H_o: $\alpha = 0$, $\beta = 0$, $\gamma = 0$	77	20.5	34.5	14.5
		(0.00)	(0.00)	(0.00)
H_o: $F(\alpha) = 0.67$, $\beta = 0.42$, $\gamma = 1$	77	2.5	24.5	−23.5
		(0.65)	(0.00)	(0.00)
H_o: $(M$ par$) - (R$ par$) = 0$	49	2.5	11.5	−5.5
		(0.57)	(0.00)	(0.15)
H_o: $(H$ par$) - (N$ par$) = 0$	28	−1.0	5.0	−1.0
		(0.85)	(0.09)	(0.85)

Note. The signs test statistic is $k - 0.5n$, where k is the number of cases in which the estimated value exceeds the value in the null hypothesis, and n is the number of cases in which the estimated value differs from the value in the null hypothesis. Two-tailed P values are reported in parentheses.

0 and 1. The main difference from the single-population data is that the α estimates center near or below 0 in B–S and in BoS. This finding is consistent with the view that α responds mainly to the mixed NE, because for these matrices the NE mixing probabilities (1/4 and 1/2, and 1/3 and 3/5) mostly are at or below $F(0) = 1/2$.

The other lines in Table III address a crucial question. Do the information and matching treatments affect the parameter estimates? Consistent with theoretical predictions, α is not significantly affected in most cases, the only exception being a higher median estimate under MM than under RP matching in the HD data. More importantly, and again as predicted, the median β estimate is invariably higher under the better information treatments (MM and History); the difference is usually significant at the 0.005 level or better. Even in the least favorable case, History vs No History in the HD data, we can reject the null hypothesis of no effect in favor of the research hypothesis at the one-tailed confidence level

TABLE IIIb
Hypothesis Tests

	Nobs	α	β	γ
Buyer−Seller				
H_o: $\alpha = 0$, $\beta = 0$, $\gamma = 0$	92	−9.0 (0.08)	30.0 (0.00)	15.0 (0.00)
H_o: $F(\alpha) = 0.67$, $\beta = 0.42$, $\gamma = 1$	92	−33.0 (0.00)	13.0 (0.01)	−33.0 (0.00)
H_o: $(M$ par$) - (R$ par$) = 0$	82	−3.0 (0.58)	14.0 (0.00)	−10.0 (0.04)
H_o: $(H$ par$) - (N$ par$) = 0$	76	−1.0 (0.91)	14.0 (0.00)	−11.0 (0.02)
Battle of the Sexes				
H_o: $\alpha = 0$, $\beta = 0$, $\gamma = 0$	97	5.5 (0.31)	45.5 (0.00)	28.5 (0.00)
H_o: $F(\alpha) = 0.67$, $\beta = 0.42$, $\gamma = 1$	97	−21.5 (0.00)	34.5 (0.00)	−37.5 (0.00)
H_o: $(M$ par$) - (R$ par$) = 0$	72	3.0 (0.56)	11.0 (0.01)	−4.0 (0.41)
H_o: $(H$ par$) - (N$ par$) = 0$	75	0.5 (1.00)	11.5 (0.01)	−0.5 (1.00)

$0.16/2 = 0.08$.[13] Recall that we also predict a lower γ under the better information treatments, and when significant (in B−S at 2 and 1% one-tailed levels, and in Co M−R at the marginal $0.15/2 = 7.5\%$ level) the signs tests confirm the prediction.[14] In view of the discussion in Section 2.4, these results also support belief learning over rote learning.

4.3. *Tests of Consistency across Payoff Matrices*

How did the payoff matrix affect the parameter estimates? The question is crucial, because a learning model is not worth much if its parameters

[13] The P values in the tables are for two-tailed tests. Again the tables do not show results for AS estimates as discussed in the previous footnote, but for 39 of the 44 HD players estimated AS in RP/NH we were also able to obtain MM/H estimates. In 24 of the 39 cases the MM/H estimates were higher, consistent with the theoretical prediction and significant at the 0.10 one-tailed level in the matched pair binomial (signs) test. We are grateful to a referee for encouraging us to pursue this robustness question and a bit surprised to get a marginally significant result despite the small number of observations per player and the small number of relevant players.

[14] When revising the theoretical section we noticed that the argument that gave us predictions for the information treatments also suggests the same impact for population size. Unfortunately the limited variation in population size in our data permits only weak tests. The most useful data are 24 Co players participating in several split group (and whole group) runs. The median estimates of β and γ are not significantly different even at the 20% level, but α is significantly higher in the split groups. The effect on α is consistent with ‘‘Kantian’’ play, a nonlearning model sketched in F96.

TABLE IV
Wilcoxon Tests

Comparison	Nobs	α	β	γ
Co vs HD	77, 127	3.87	5.51	0.39
		(0.00)	(0.00)	(0.70)
B−S vs BoS	92, 97	−2.36	0.55	−0.61
		(0.02)	(0.59)	(0.54)
B−S: Pop B vs A	92, 92	1.28	1.35	−0.07
		(0.20)	(0.18)	(0.95)
BoS: Pop B vs A	97, 97	−7.11	1.83	0.18
		(0.00)	(0.07)	(0.86)

Note. The null hypothesis is that both samples in the comparison have the same distribution. The test statistic (and the two-sided P values in parentheses) are for the normal approximation to the Wilcoxon rank-sum test.

change unpredictably when payoffs change. Table IV reports Wilcoxon rank-sum tests of the null hypothesis that the distribution of parameter estimates is the same across various data sets. (Signs tests are inappropriate since we cannot pair estimates across the different data sets.) We find that α is very significantly higher in Co data than in HD data; an interpretation is that players' idiosyncratic preferences for the payoff-dominant action over its alternative is stronger than preferences for Hawk over Dove. The smaller NE mixing fractions in B−S than in BoS are a plausible explanation for the α differences reported in the second line of Table IV. The same sort of explanation would account for the third line (marginally significant at the $0.20/2 = 10\%$ one-tailed level). We did not expect to see the highly significant difference in α between the two populations in the Battle of the Sexes; apparently a strong convention emerged favoring the second population.

The β comparisons have greater theoretical interest. Recall that basic learning theory suggests no effect, consistent with the results reported in the second and third lines. Recall also that the selection bias discussed in footnote 4 (as well as anticipatory learning) correctly predicts the significantly higher βs for type 2 than type 1 matrices reported in the first line. The marginally significant effect in the last line is puzzling. Perhaps the emergent convention requires players in the second population (with more to gain from getting their preferred NE) to pay more attention to payoffs and less to idiosyncratic preferences.

The most important theoretical prediction here is that γ is invariant to the payoff matrix. It is gratifying to see that in every case the estimates are entirely consistent with that prediction.

TABLE V
Contingency Tables for Gamma Classifications

a. Individual Players

Type	Fickle	Cournot	Adaptive	Fictitious	Imprint	Uninformative	Total
BoS	4	51	6	18	0	18	97
B−S	4	37	9	17	0	25	92
Co	5	25	6	18	1	22	77
HD	15	31	8	31	1	41	127
Total	28	144	29	84	2	106	393

$\chi^2_{15} = 27.4$, P-value $= 0.03$

b. Individual Players

Type	Cournot	Adaptive	Fictitious	Total
BoS	51	6	18	75
B−S	37	9	17	63
Co	25	6	18	49
HD	31	8	31	70
Total	144	29	84	257

$\chi^2_{6} = 10.7$, P-value > 0.10

c. Individual Players

Type	Cournot	Adaptive	Fictitious	Total
Co	25	6	18	49
HD	31	8	31	70
Total	56	14	49	119

$\chi^2_{2} = 0.7$, P-value $= 0.71$

d. Aggregate Players

Type	Cournot	Adaptive	Fictitious	Total
Co	5	2	1	8
HD	0	2	6	8
Total	5	4	7	16

$\chi^2_{2} = 8.6$, P-value $= 0.01$

4.4. *Cournot vs Fictitious Play vs Adaptive Learning*

We now take a closer look at the distribution of γ estimates. Table V classifies the individual players using (one-sided, heteroskedasticity-consistent) t tests at the 5% level of the null hypotheses that γ_i is 0 or 1. The player is called fickle if we accept the alternative hypothesis $\gamma_i < 0$ and is called imprintable if we accept $\gamma_i > 1$. When we can reject neither null hypothesis the player is called uninformative. The remaining players are called Cournot if we accept the alternative hypothesis $\gamma_i < 1$ and do not reject the null $\gamma_i = 0$; fictitious if we accept $\gamma_i > 0$ but do not reject the null $\gamma_i = 1$; and adaptive if we accept both $\gamma_i > 0$ and $\gamma_i < 1$.

The first contingency table shows that even under these rather exacting definitions almost $2/3$ of the players (257 of 393, or 65.4%) can be classified as Cournot, adaptive, or fictitious. The null hypothesis that the classification proportions are the same across all payoff matrix types is rejected at the 3% level by the χ^2 statistic for the first contingency table. However, a look at the cell entries raises the suspicion that the rejection arises not from fundamental differences, but rather from varying proportions of the undesirable classifications.[15] Restricting attention to the desirable classifications, we see in Table Vb that we cannot quite reject the null hypothesis of equal frequencies even at the 10% level, an impressive result for a contingency table of this size.

The remaining contingency tables in Table V tell a cautionary tale about our next theme, aggregation. The χ^2 statistic in Table Vd indicates a significant difference at the 1% level between HD and Co for the aggregate (“98”) players, while Table Vc shows that there actually is no difference for individual players. A closer look at the cell counts shows that the discrepancy arises mainly because in the HD data individual Cournot players and fictitious players are equally common (31 each) but the Cournot players disappear in the aggregate data.

The point is important. The empirical power of our learning model stems largely from the hypothesis that people do not change their learning process when the payoff function changes (or, more precisely, that the γ distribution is invariant to M). If had we relied on aggregate estimates, we would have erroneously concluded that the fictitious play model characterized the HD data and the Cournot model better characterized the Co data. The individual estimates actually confirm the invariance hypothesis rather strongly.

[15] Moreover, the usual rule of thumb for a reliable χ^2 is that each cell in the contingency table contain at least five observations. This condition holds in Table Vb but not Va.

4.5. *Individual vs Aggregate Learning Models*

Estimating the learning model on an individual basis may occasionally keep us from making incorrect inferences, but the question still remains whether the complication of individual estimation really provides much additional explanatory power. After all, as economists we are interested in individual differences mainly to the extent that those differences affect aggregate behavior, and individual estimation is "costly" in terms of statistical degrees of freedom. Does the improved fit to the aggregate data justify the use of separate parameters for each individual player?

Table VI shows that the answer is strongly affirmative. For each of the main payoff matrices and each relevant laboratory session, we compare the individualized fits of the three-parameter learning model to representative agent fits of the same model. We use a likelihood ratio test for the null hypothesis that the representative agent (player 98) model is correct. The null hypothesis implies that the statistic $-2(L_{98} - \sum_{i=1}^{n_{98}} L_i)$ has the asymptotic $\chi^2(d)$ distribution, where $d = 3(n_{98} - 1)$ is the degrees of freedom and L_i, $i = 1, \ldots, n_{98}$ are the maximum log likelihoods of the individual player model fits, while L_{98} is the maximum log likelihood of the representative agent model fit.

TABLE VIa
Loglikelihood Aggregation Tests for Three-Parameter Model

	Hawk–Dove				Coordination			
Session	Nobs	*df*	LRStat	*P* value	Nobs	*df*	LRStat	*P* value
3	720	27	41	0.04	252	18	37	0.00
6	1320	30	308	0.00	270	27	71	0.00
9	1200	24	233	0.00	720	21	133	0.00
10	1170	42	332	0.00	—	—	—	—
15	594	27	84	0.00	1188	30	207	0.00
20	—	—	—	—	1188	30	217	0.00
21	540	27	69	0.00	486	24	84	0.00
22	756	27	78	0.00	—	—	—	—
23	600	27	124	0.00	—	—	—	—
24	675	42	149	0.00	315	18	64	0.00
25	495	30	144	0.00	792	30	127	0.00
26	756	33	107	0.00	108	9	34	0.00

Note. The LRStat uses the likelihood ratio test to compare likelihood values computed at the experiment level with the sum of likelihood values for all individuals in the corresponding experiment. The degrees of freedom (*df*) is defined by: (the number of individuals in the experiment − 1)∗(the number of parameters estimated).

TABLE VIb
Loglikelihood Aggregation Tests for Three-Parameter Model

	Buyer−Seller				Battle of the Sexes			
Session	Nobs	*df*	LRStat	*P* value	Nobs	*df*	LRStat	*P* value
5	864	21	146	0.00	—	—	—	—
7	—	—	—	—	1440	27	63	0.00
13	440	27	153	0.00	1144	33	105	0.00
17	—	—	—	—	648	33	81	0.00
19	216	21	85	0.00	—	—	—	—
23	660	30	170	0.00	720	33	123	0.00
27	540	42	112	0.00	1008	39	110	0.00
28	504	39	111	0.00	1152	45	111	0.00
29	504	39	120	0.00	432	33	73	0.00
30	432	33	87	0.00	288	21	37	0.02

Note. The LRStat uses the likelihood ratio test to compare likelihood values computed at the experiment level with the sum of likelihood values for all individuals in the corresponding experiment. The degrees of freedom (df) is defined by: (the number of individuals in the experiment − 1) $*$ (the number of parameters estimated).

The entries in Table VI show that in 34 of 36 cases we reject the null hypothesis at the 0.005 level. Even in the exceptional cases (HD session 3 and BoS session 30) we can reject the null at the 5% level. We conclude that heterogeneity across subjects in terms of learning and decision parameters is an important feature of our data.

5. DISCUSSION

Belief learning models offer the prospect of predicting aggregate performance when people interact, even in novel environments with novel institutions. In this paper we proposed a very simple belief learning model that generalizes the classic models of Cournot and fictitious play. The model has three fitted parameters, one for learning rate or memory length (γ) and two for the decision process (β and α). The model succeeded in capturing many aspects of behavior observed in a variety of laboratory environments. In particular,

1. We find that players are quite heterogeneous. Allowing for parameter differences across individuals greatly increases explanatory power and improves the reliability of inferences.

2. Estimates of the parameter γ allow us to classify the majority of players as short memory (Cournot), or intermediate (adaptive), or long memory (fictitious play).

3. We argue on theoretical grounds that the distribution of Cournot, adaptive and fictitious players should be invariant to the payoff function. The individual data from our simple payoff matrices are consistent with the invariance prediction. (Relying on a representative agent model would have lead us to the incorrect inference that the distribution depends largely on the payoff matrix.)

4. When significant, the data are consistent with the theoretical prediction that γ will decrease in more informative environments.

5. The decision sensitivity parameter β is positive for most players, consistent with direct but approximate optimization. It is negative for a few players, as suggested by Selten's (1991b) anticipatory model.

6. Simple (nonanticipatory) learning predicts that the distribution of β is invariant to changes in the payoff function, but a selection bias suggests that observed values of β will be lower for matrices that encourage convergence to an interior equilibrium. Most of the data are consistent with the selection bias (or with anticipatory play).

7. The data are quite consistent with the theoretical prediction that β will increase in more informative environments.

8. The decision bias parameter α varies across individual players but centers near (not precisely on) values that imply stability for interior Nash equilibria.

9. The distribution of α generally responds as predicted to changes in the payoff matrix. The data are also consistent with the prediction that the α distribution is invariant to changes in the information environment.

10. Taken together, the results (especially 3, 4, 6, 7, and 9) support our arguments for belief learning over rote learning, which takes no explicit account of beliefs.

Despite these successes, we do not believe that our simple three-parameter model captures all important aspects of learning in games. There is some scattered evidence of anticipatory play in the sense of Selten (1991b) and other signs that the model is misspecified for some players. Fortunately, there is no shortage of ideas for modifying the model. Theoretical learning models continue to proliferate in the economics (and psychology) literature.

We hope our paper will help focus professional creativity on empirical learning models that (a) are sensitive to the institutional context, especially to the information conditions, (b) allow for individual diversity, (c) represent beliefs explicitly, (d) account for the decision process, and, of course, (e) are sufficiently tractable and parsimonious for empirical work. We look forward to the day, preferably neither very soon nor very late, when such a model improves on our current explanation of the data.

APPENDIX: MATHEMATICAL DETAILS

Stability at Interior NE

Let s^* be an interior NE of the symmetric 2×2 game with payoff matrix M and let $c = (1, -1)M(1, -1)'$. For the moment assume a representative agent, i.e., assume that for a fixed value of each parameter ($\gamma \geq 0$, α, and β) Eqs. (2.1)–(2.3) describe the behavior of every subject. Conditioned on the realization of previous actions h_t, the state next period has a binomial distribution with mean $F(\alpha + \beta R(\hat{s}))$, where F is the cumulative unit normal distribution with density $F'(x) = (2\pi)^{-1/2} \cdot \exp[-x^2/2]$. Recall that $\hat{s}$ is defined from h_t by Eq. (2.1), that $R(s) = b + cs$ by (2.2) and that $R(s^*) = 0$.

The stochastic approximation argument, summarized, e.g., in Chapter 4 of Sargent (1993), shows that for (local asymptotic) stability it suffices to show that the deterministic part of the dynamics at $s = s^*$ is error reducing. Accordingly, let $s = s^* + e$ be the current state and let the preceding states be error free, $s = s^*$. Then from (2.1) we have $\hat{s} = as + (1 - a)s^* = ae + s^*$, where $a^{-1} = \sum_{u=0}^{t-1} \gamma^u$; note that the weight a is between 0 and 1 for all $t > 0$ as long as $\gamma \geq 0$. The deterministic part of the response to the current error e therefore is $f(e) = F(\alpha + \beta R(ae + s^*)) = F(\alpha + \beta(b + cae))$. The error reduction condition is $|f'(0)| < 1$. Since $|f'(0)| = |\beta c| aF'(\ ) \leq |\beta c| F'(\ ) \leq |\beta c|(2\pi)^{-1/2}$, it follows that $|\beta| < \sqrt{2\pi}/|c| = \beta_s$ ensures stability.

Extensions of the argument show that the uniform bound $|\beta_i| < \beta_s$ suffices even when the β_is differ across individuals, and that the bound is sharp in the sense that if all individuals have $\beta_i > \beta_s$ then (for sufficiently small positive γ and t) we have an overshooting instability at s^*.

Comparative Statics of Information Conditions

Consider a static binary decision where action $a = 1$ is chosen if a summary statistic $y \geq 0$, and $a = 0$ is chosen if $y < 0$. The statistic y comes from four sources.

1. ‘‘old’’ evidence, an unbiased sample of size N from a distribution of precision $\tau_1 \equiv \sigma_1^{-2}$;
2. ‘‘new’’ evidence, an unbiased sample of size N from a distribution of precision $\tau_2 > \tau_1$;
3. unbiased prior evidence of precision τ_o; and
4. a random impulse y_α, normalized to unit precision.

Iterated application of Theorem 9.5.1 of DeGroot (1970, p. 167) tells us that if all relevant variables are independent and normal then the Bayesian (optimal) summary statistic is $y = (y_\alpha + \tau_o \mu + N\tau_1 \bar{x}_1 + N\tau_2 \bar{x}_2)/(1 + \tau_o + N\tau_1 + N\tau_2)$, where μ, $\bar{x}_1$, and $\bar{x}_2$ are the prior and the sample means. It is clear that the weight $w_\alpha = (1 + \tau_o + N\tau_1 + N\tau_2)^{-1}$ of the impulse is decreasing in sample size N, so the weight $w_\beta = 1 - w_\alpha$ of the evidence is increasing in N. Finally, let $\Gamma = (\tau_o + N\tau_1)/(N\tau_2) = \tau_1/\tau_2 + N^{-1}\tau_o/\tau_2$ be the weight of the older evidence relative to the "new" evidence. Then Γ decreases in N, although the effect is relatively weak when priors are imprecise.

The static model should not be taken literally. Players may not have the sophistication to use Bayes' theorem, and even sophisticated players may not know enough about other players' beliefs, decision rules, and payoffs to use a correct likelihood function. Nor are the distributional assumptions satisfied in laboratory data. Yet we believe that the comparative statics of the simple model capture the natural response of most humans to increases in the quality of information. If so, we obtain the following predictions when we take the sample size N as a proxy for the quality of information.

- The evidence weight β in the three-parameter empirical model, like w_β in the static model, will increase when the quality of information improves.

- The learning parameter γ in the three-parameter empirical model, like Γ in the static model, will decrease when the quality of information improves. This effect should be weaker since the impact of N on Γ dies out as we increase the number of periods or decrease the prior precision.

REFERENCES

Boylan, R., and El-Gamal, M. (1993). "Fictitious Play: A Statistical Study of Multiple Economic Experiments," *Games Econ. Behav.* **5,** 205–222.

Brown, G. (1951). "Iterated Solution of Games by Fictitious Play," in *Activity Analysis of Production and Allocation* (T. Koopmans, Ed), pp. 377–380. New York: Wiley.

Cheung, Y-W., and Friedman, D. (1994). "Learning in Evolutionary Games: Some Laboratory Results," UCSC Economics working paper No. 303.

Cournot, A. (1838). *Researches sur les Principes Mathematiques de la Theorie des Richesses*. (N. Bacon, Ed.), English ed. *Researches into the Mathematical Principles of the Theory of Wealth*. New York: Macmillan, 1897.

Crawford, V. (1995). "Adaptive Dynamics in Coordination Games," *Econometrica* **63,** 103–144.

DeGroot, M. (1970). *Optimal Statistical Decisions*. New York: McGraw–Hill.

Friedman, D. (1991). "Evolutionary Games in Economics," *Econometrica* **59,** 637–666.

Friedman, D. (1996). "Equilibrium in Evolutionary Games: Some Experimental Results," *Econ. J.* **106,** 1–25.

Fudenberg, D., and Kreps, D. (1993). "Learning Mixed Equilibria," *Games Econ. Behav.* **5,** 320–367.

Fudenberg, D. and Levine, D. (1993). "Steady State Learning and Nash Equilibria," *Econometrica* **61,** 547–575.

Fudenberg, D., and Levine, D. (1995). *Theory of Learning in Games*, draft manuscript, UCLA Economics Department.

Harsanyi, J. C., and Selten, R. (1988). *A General Theory of Equilibrium in Games*. Cambridge, MA: MIT Press.

Kreps, D. (1990). *A Course in Microeconomic Theory*. Princeton, NJ: Princeton Univ. Press.

McKelvey, R. D., and Palfrey, T. R. (1992). "An Experimental Study of the Centipede Game," *Econometrica* **60,** 803–836.

McKelvey, R. D., and Palfrey, T. R. (1995). "Quantal Response Equilibria for Normal Form Games," *Games Econ. Behav.* **10,** 6–38.

Mookherjee, D., and Sopher, B. (1994). "Learning Behavior in an Experimental Matching Pennies Game," *Games Econ. Behav.* **7,** 62–91.

Roth, A., and Erev, I. (1995). "Learning in Extensive Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term," *Games Econ. Behav.* **8,** 164–212.

Sargent, T. (1993). *Bounded Rationality in Macroeconomics*. Oxford: Clarendon Press.

Selten, R. (1991a). "Evolution, Learning, and Economic Behavior," *Games Econ. Behav.* **3,** 3–24.

Selten, R. (1991b). "Anticipatory Learning in Games," in *Game Equilibrium Models, Vol. 1*, Chapter 3. New York: Springer-Verlag.

Simon, H. (1957). *Models of Man*. New York: Wiley.

Stahl, D., and Wilson, P. (1995). "On Players' Models of Other Players: Theory and Experimental Evidence," *Games Econ. Behav.* **10,** 218–254.

Zeeman, E. C. (1980). "Population Dynamics from Game Theory," in *Global Theory of Dynamical Systems: Lecture Notes in Mathematics, 819* (Z. Nitecki and C. Robinson, eds.). Berlin: Springer-Verlag.