GAMES AND ECONOMIC BEHAVIOR **16,** 303–330 (1996)
ARTICLE NO. 0088

Boundedly Rational Rule Learning in a Guessing Game

Dale O. Stahl*

Department of Economics, University of Texas, Austin, Texas 78712

Received March 1, 1995

We combine Nagel's "step-*k*" model of boundedly rational players with a "law of effect" learning model. Players begin with a disposition to use one of the step-*k* rules of behavior, and over time the players learn how the available rules perform and switch to better performing rules. We offer an econometric specification of this dynamic process and fit it to Nagel's experimental data. We find that the rule of learning model vastly outperforms other nested and nonnested learning models. We find strong evidence for diverse dispositions and reject the Bayesian rule-learning model. *Journal of Economic Literature* Classification Numbers: C70, C52, D83. © 1996 Academic Press, Inc.

1. INTRODUCTION

While there is ample evidence that people rarely choose Nash equilibrium strategies the first time they play a given game, there is evidence that some people tend to learn to play a Nash equilibrium.[1] How do people learn to play? What do people learn (simple behaviors or complicated rules)? How do people make their initial choice, and how is the information they accumulate incorporated into the dynamic learning process? A growing literature on learning in games is addressing these questions.[2]

* Partial funding of this research was provided by grant SBR-9308914 from the National Science Foundation. The author thanks Rosemarie Nagel for permission to use her experimental data for this study and for many stimulating conversations, Xioahua Lu for substantial programming and research assistance, and Paul Wilson for statistical consultation. The author is also indebted to Daniel Friedman for many helpful suggestions on earlier drafts. All errors and omissions are the author's sole responsibility. E-mail: stahl@mundo.eco.utexas.edu.

[1] Cheung and Friedman (1994), El-Gamal *et al*. (1993, 1994), Mookherjee and Sopher (1994), Roth and Erev (1995), Van Huyck *et al*. (1995) to mention just a few.

[2] Crawford (1995), Friedman *et al*. (1995), Fudenberg and Kreps (1993), Fudenberg and Levine (1993), Jordan (1991), Kalai and Lehrer (1993), Milgrom and Roberts (1991), Rosenthal (1993), Selten (1990, 1991). Selten and Buchta (1994), etc.

0899-8256/96 $18.00
Copyright © 1996 by Academic Press, Inc.
All rights of reproduction in any form reserved.

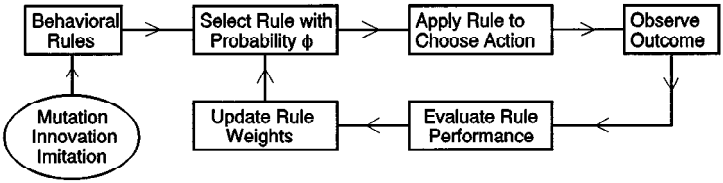

FIGURE 1

In a simple adaptive behavior model, an individual is endowed with an initial "propensity" to choose a particular behavior, and the observed payoffs for various choices affect future propensities: behaviors which lead to more favorable outcomes are more likely to be repeated in the future. This is the "Law of Effect" in psychology (see Bush and Mosteller, 1955, and Roth and Erev, 1995). Simple adaptive behavior models can predict reasonably well and lead to optimal behavior in stationary, single-decision-maker problems. However, in games that generate nonstationary sequences of observations, such simple models are severely limited.

The main distinction between a single decisionmaker problem and a game is that an intelligent human player of a game will attempt to anticipate the actions of the other players. Therefore, a successful learning model for games must entail a model of other players, as well as best-response computations.

The Bayesian approach to learning hypothesizes that players have *priors* (beliefs) of the dynamic strategies of other players, update these *priors* using Bayes theorem, and choose a best response given these beliefs (e.g., Blume and Easley, 1992; Jordan, 1991; Kalai and Lehrer, 1993; and Nyarko, 1991). This extremely complicated process is highly unrealistic and there is no evidence that real people use it (Holt, 1993).

Rosenthal (1993) and others (e.g., Holland, 1976; Arthur, 1990, 1993; and Aumann, 1986) have proposed that players begin with a set of "rules-of-thumb" and that the propensity to employ a particular rule-of-thumb is reinforced as in the adaptive behavior model. In other words, the definition of "behavior" is broadened to include rules-of-thumb as well as simple actions.

More generally, we could consider a measurable class of functions (or rules) that map from a player's information to the set of feasible actions. Propensities would be defined as a measure on this class, and the reinforcement principle would define a learning dynamic, as illustrated in Fig. 1. Kuan and White (1994) and Chen and White (1994) investigate the general stochastic properties of such learning dynamics. A wide variety of behavioral rules could be encompassed in such a general model. For example, maximin, maximax, and best response to uniform.

However, to operationalize such a general theory, we must address three sub-

stantive issues. *First*, what is the empirically relevant class of rules? *Second*, what are the initial propensities: Do all individuals have the same initial propensities, or is there significant diversity among individuals? *Third*, what are the specific dynamics: how is reinforcement quantified, how much weight is given to old information and new information, how much experimentation/search is employed, how are new rules introduced, and how significant are computational errors?

In this paper, we suggest an approach to these issues for strategic situations in which each player is given a summary statistic of the other player's choices in the previous period of play. Our approach yields a tractable empirical model of rule learning, and we confront this model with data from a guessing game experiment. Crawford (1995) also addresses this class of games but using a simple adaptive behavior model.

Nagel's (1995) experiment and model motivate our approach. There is a finite number of players. Each chooses a real number in the [0,100] interval. The mean of all the submitted numbers is computed, and the player that is closest to ρ times the mean wins a prize; all others get nothing. We will focus on two treatments by Nagel in which $\rho = 1/2$ and 2/3 respectively. The game is dominance solvable with 0 as the solution. However, in actual experiments, the mean is always well above 0 and only slowly decreases over several trials. Figure 2 displays a kernel estimate of the density function for Nagel's data using a bandwidth of 2.[3] One of the striking features is that in the first period there appear to be modes in the distribution near 50ρ and $50\rho^2$, suggesting that the participants used rules-of-thumb corresponding to different depths of reasoning as suggested by Nagel.[4] Nagel computed a nonparametric test of this hypothesis and found that the period-1 choices were significantly concentrated in intervals around 50, 50ρ, $50\rho^2$, and $50\rho^3$. Further, the written comments from the participants lend support to this "depth-of-reasoning" explanation.

Nagel proposed the following model of boundedly rational behavior. The *step*-0 *rule* chooses the previous mean, with 50 being the default mean for the first period.[5] A player using the step-0 rule essentially has no understanding of the game, but tends to follow the crowd. The *step*-1 *rule* chooses a best response to the previous mean. A player using the step-1 rule essentially believes that all other players will use the step-0 rule. The *step*-2 rule chooses a best response to the step-1 rule. A player using the step-2 rule essentially believes that all others will use the step-1 rule, etc.

For periods 2 to 4, it appears from Fig. 2 that the distribution becomes more

[3] We present this figure as a convenient and suggestive display of the data, not as a formally tested kernel estimate. We used a bandwidth of 2 because the resulting distribution best reveals the modes that Nagel's nonparametric tests showed were significant.

[4] See also Stahl (1993) and Stahl and Wilson (1994, 1995).

[5] Nagel tested and rejected 100 as the default value in favor of 50.

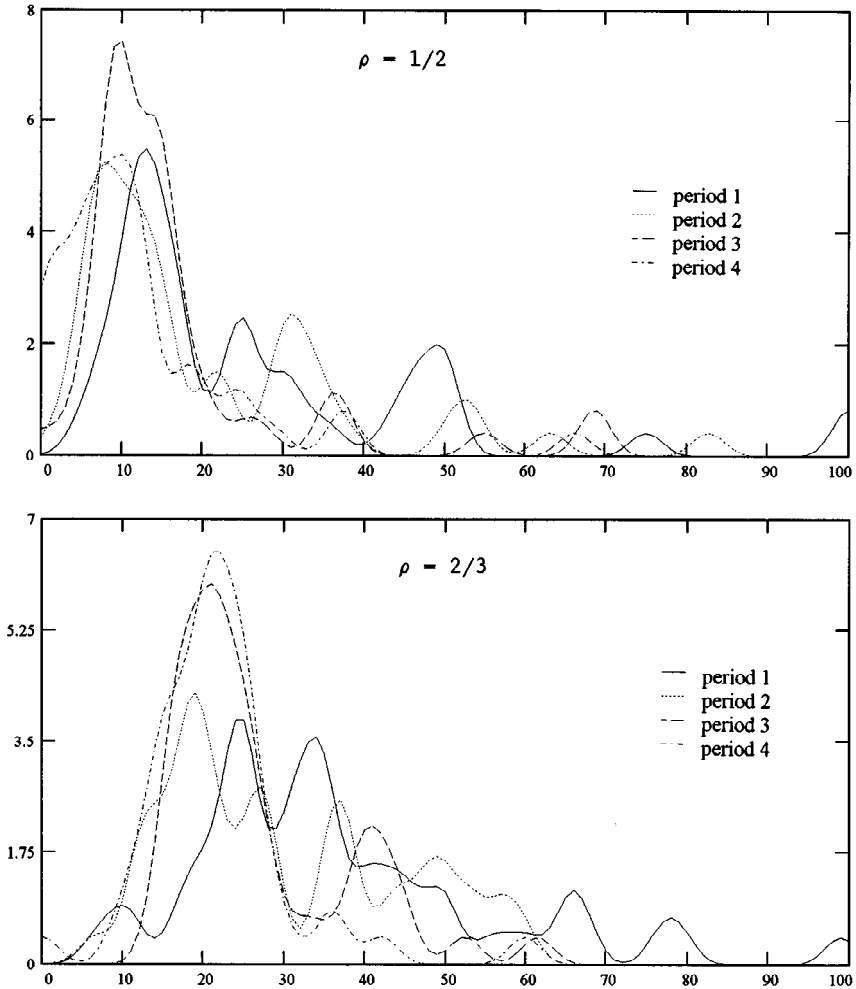

FIG. 2. Kernel Estimates of Nagel Data: Bandwidth = 2.

concentrated around $50\rho^2$, with no strong tendency for convergence to the Nash equilibrium at 0.[6] Nagel performed statistical tests showing that the shift of the distribution toward the step-2 rule is significant. On the other hand, Nagel's test

[6] For periods 2 to 4, the data are divided by the previous mean and multiplied by 50, so $50\rho^k$ can be interpreted as the behavioral rule of choosing ρ^k times the previous mean in every period—i.e., a step-k rule. This interpretation implicitly assumes that 50 is the default value for the previous mean in period 1.

of the prevalence of the step-k rules in periods 2 to 4 fails to reject the hypothesis that the choices were not significantly concentrated around the predictions for the step-0 to step-3 rules inclusive. It is possible, as Nagel argues, that this test failed because different behavioral rules enter after the initial period. However, it also could be that computational error or individual diversity increase, or that learning behavior itself increases the likelihood of type-II errors for this test.

Nagel's hierarchial models suggests not only a set of behavioral rules but also suggests a set of initial propensities (or archetypal "dispositions") for our learning model: each archetypal disposition has the strongest propensity for one of the rules. We define a one-parameter class of such dispositions, and estimate that parameter from the data.

We also hypothesize computational errors and individual diversity within each type category, so the choice of a player employing the step-k rule is a random variable distributed normally with expectation equal to ρ^k times the mean of all players' choices in the previous period and a standard deviation estimated from the data.

Our model hypothesizes that the propensities to use the various rules change from period to period due to reinforcement according to relative performance. At the beginning of each period after the first, the participants know the previous mean. Therefore, they can assess how each rule would have done, taking account of the potential for computational errors. Specifically, we assume that the relevant reinforcement variable for a rule is proportional to the expected payoff of the rule, which is equal to the probability density that the rule would have generated the winning number. Then, the current period's propensity is a weighted average of last period's propensity and the recent reinforcement. These weights are estimated from the data and interpreted later.

Nagel finds support for an alternative "directional" learning model (similar to that of Selten and Buchta, 1994), in which each individual's choice is a multiple of the previous mean, and this multiplicative factor is adjusted up (down) if it was too low (high) in the previous period. We will compare this alternative with our model in Section 4.2. Nagel also interprets the evidence as providing only weak support for increasing depth of reasoning.

When confronted with the data, we find that our learning model vastly outperforms the simple adaptive behavior model and other nested models. We find persuasive evidence for heterogeneous initial dispositions. We strongly reject the hypothesis that a player with a step-k disposition uses the step-k rule in *every* period. In other words, a player learns more than just the previous mean which gets incorporated into the step-k rules; he learns which rules perform better and switches to them. We also formulate, test, and subsequently reject a Bayesian learning model.

In Section 2 we present the formal modeling and in Section 3 we present the econometric methodology. In Section 4 we present the results and tests, and we conclude in Section 5.

2. THE MODEL

2.1. *The Game*

Suppose there are N players. A parameter $\rho \in (0, 1)$ is announced at the beginning of the first period and is fixed for all periods. At the beginning of a period, a player chooses a real number in $A \equiv [0, 100]$. Let $\bar{x}_t$ denote the mean of all the choices for period t. The player whose choice is closest to $\rho\bar{x}_t$ wins a prize. If there are ties, the winner is chosen randomly with equal probability from those who chose the closest number; otherwise, a player receives nothing. After the first period, every player is fully informed about the mean and ρ times the mean of the previous period. There are four periods in all.

2.2. *The Behavioral Rules*

Let $k \in 0, 1, \ldots, K$ index a set of $K + 1$ behavioral rules, and let $f_k\colon A \to \Delta(A)$ denote the *rule* that maps the previous mean of the choices of all players into a probability density on the set of current actions. If the mean in period t was $\bar{x}_t$, then the probability density of x_{t+1} under rule k is denoted $f_k(x_{t+1}; \bar{x}_t)$. We specify that the mean of $f_k(\cdot; \bar{x}_t)$ is $\rho^k \cdot \bar{x}_t$, corresponding to Nagel's "step-k" rule described in the Introduction. Further, we define $\bar{x}_0 \equiv 50$, since 50 is the mean of the uninformed prior.

Thus, $f_k(\rho\bar{x}_t; \bar{x}_{t-1})$ is the probability density for rule k evaluated at the *ex post* best choice for period t. Since the probability that rule k would have been the best rule to use is directly related to this density, it is an obvious candidate for a *performance measure*. However, since making a choice and evaluating past performance are two different operations, we want to allow for the possibility of a different error process as well as the possibility of spillover reinforcement of "nearby" rules.

To this end, let $g_k\colon A \to \Delta(A)$ denote a probability density that has the same expected value as f_k but perhaps a different standard deviation. We will use g_k as the performance measure for rule k.

We must specify a family of distributions that we can identify from the experimental data. Since several participants chose extreme points (0 and 100), we cannot use a family with a zero density at the extreme points; this rules out the beta distribution. For econometric simplicity, we choose the family of normal distributions truncated to the [0,100] interval, with mean m_k and a standard deviation given by

$$\sigma_k \equiv s_1 + r_1 \cdot \theta \cdot m_k \qquad \text{for } k > 0. \tag{1}$$

This specification of the standard deviation with absolute and relative components allows for the possibility that the standard deviation decreases as the mean decreases but does not go below a minimum level. While the theory specifies that $m_k = \rho^k \cdot \bar{x}_{t-1}$ for period t, the (s_1, r_1) parameters will be estimated from the

data; θ is a constant function of the treatment parameter $\rho : \theta = (1-\rho)/(1+\rho)$. This value of θ solves the equation $(1+\theta)\rho^{k+1} = (1-\theta)\rho^k$. By measuring the relative standard deviation in units of θ, one relative standard deviation below m_k is the same point as one relative standard deviation above m_{k+1}. Consequently, r_1 measures the relative gap between successive behavioral rules in a manner that is invariant to ρ.

Since a step-0 type has no real understanding of the game, we allow for different parameter values for the step-0 standard deviation: namely, $\sigma_0 = s_0 + r_0 \cdot m_0$. It turns out that in all models in which we included the s_0 term, the maximum likelihood estimate of s_0 was exactly 0, so in the interest of efficient estimation of the other parameters of the model we set $s_0 = 0$ for all models reported in this paper. We also tested and strongly rejected the hypothesis that $s_0 = s_1$ and $r_0 = r_1$.

For the performance measure g_k, we specify the standard deviation similarly as $s_2 + r_2 \cdot \theta \cdot m_k$ for $k > 1$, and $r_3 \cdot m_0$ for $k = 0$. We can interpret this performance "bandwidth" as computational error in computing performance, and/or as a neighborhood reinforcement effect (i.e., second-best rules get reinforced somewhat also).

2.3. *Propensities and Dispositions*

A player begins a period with a vector of propensities toward the $K+1$ rules: $\omega \equiv (\omega_0, \ldots, \omega_K)$, with the interpretation that the probability of choosing rule k is given by $\exp(\omega_k)/\sum_{j=0}^{K} \exp(\omega_j)$. Since ultimately only the probabilities matter, the equivalence classes of initial propensities can be associated with the K-dimensional simplex. Propensities for later periods will be determined by the dynamics specified in the next subsection. Each individual's behavior will be uniquely determined by his/her initial propensities and the dynamics. We will call an individual's initial vector of propensities a *disposition*.

Unfortunately, with only four periods and $K \geq 3$, it will not be possible to identify the disposition of each individual. Instead, we consider only a finite set of dispositions. The evidence presented by Nagel suggests that a reasonable set of dispositions is the set corresponding to the $K+1$ rules.

In specifying a disposition corresponding to the step-k rule, we clearly want the step-k rule to be the most likely rule to be chosen, but we want the other rules to have some positive chance of being chosen, because if they have a zero chance of being chosen, then a wide class of learning dynamics (including Bayesian updating) would never give them a positive chance. With this in mind, we define a one-parameter class of prior dispositions.

Let $w(k, j, t)$ denote the propensity of a player to use rule j in period t given he/she has a type-k disposition. Then, we define $w(k, k, 1) \equiv \mu > 0$, and $w(k, j, 1) \equiv 0$ for $j \neq k$. With this specification, a *type-k disposition* will have a $\exp(\mu)/[K + \exp(\mu)] > 1/(K+1)$ probability of choosing the step-k rule, and

a $1/[K + \exp(\mu)] < 1/(K + 1)$ probability of choosing a step-j rule ($j \neq k$). The parameter μ, which will be estimated from the data, can be interpreted as the strength of the initial propensities: the larger μ, the more likely it is that the designated rule will be chosen over the other rules.

To summarize, we specify a one-parameter family of dispositions corresponding to the $K + 1$ rules. A disposition specifies the initial vector of propensities to use each of the $K + 1$ rules. An individual with a type-k disposition is more likely to use rule k than any other rule. Thus, an individual is fitted into one of $K + 1$ types. To the extent that there is more diversity among individuals than these $K + 1$ types, we hope that the noise parameters $(r_0, s_1, r_1, s_2, r_2, r_3)$ will capture that diversity.

2.4. *The Dynamics*

The law of effect asserts that good performing rules should be more likely to be used in the future and bad performing rules should be less likely to be used in the future. In the guessing game, there are only two outcomes (win or lose), so the appropriate performance measure is the likelihood of winning conditional on rule j: g_j.

For periods $t > 1$, we define the dynamic on propensities by:

$$w(k, j, t) = \beta_0\, w(k, j, t-1) + \beta_1\, g_j(\rho \bar{x}_{t-1}; \bar{x}_{t-2}). \tag{2}$$

The β_0 coefficient represents the weight given to the most recent propensity. If $0 \leq \beta_0 \ll 1$, then the propensity is discounted heavily (or forgotten readily). In the absence of performance information, the propensities would decay toward zero, and all rules would tend to be chosen with equal likelihood. On the other hand, if $\beta_0 > 1$, then the propensity is accentuated in the future, and in the absence of performance information the rule with the highest initial propensity would tend to be used almost always. The β_1 coefficient represents the weight given to current performance. The rule for which $\rho \bar{x}_{t-1}$ has the largest probability density will receive the most reinforcement. If β_1 is large, then the new propensity will concentrate on the most recent best rule, while if β_1 is small then new performance information will be incorporated slowly.

Given these propensities, the probability that a player with a type-k disposition uses rule j in period t is defined as

$$\varphi(k, j, t) \equiv \exp[w(k, j, t)] / \sum_{\ell} \exp[w(k, \ell, t)]. \tag{3}$$

Note that because of the logarithmic specification of propensities, the probabilities would not be affected by the addition of a constant term to Eq. (2) for renormalization purposes.

2.5. *The Log-Density Mixture Model*

Let $i \in \{1, \ldots, N\}$ index players, and let $x(i, t)$ be the choice of player i in period t. Then, conditional on having a type-k disposition, the probability density of player i's choice is

$$p(i, k, t) \equiv \sum_{j=0}^{K} \varphi(k, j, t) \cdot f_j(x(i, t); \bar{x}_{t-1}). \tag{4}$$

Hence, the probability density of player i's four choices (one for each period), conditional on having a type-k disposition, is

$$P(i, k) \equiv \prod_{t=1}^{4} p(i, k, t). \tag{5}$$

Let α_k denote the proportion of the population of players that have a type-k disposition for $k \in \{0, 1, \ldots, K\}$. Each of these types has the potential to learn that other rules are better.

However, it is conceivable that some players do not learn at all. Indeed, the data reveal that several players chose 100 in rounds 3 and 4, which cannot be rationalized by any rule. Rather than throw out these "outliers" from the data set, we introduce another type that chooses randomly from a uniform distribution in every period; we call such players "-1" types and let α_{-1} denote the proportion in the population.[7] We also define $P(i, -1) = 0.01$, the density of the uniform distribution over A. Naturally, each α_k is nonnegative and $\sum_{k=-1}^{K} \alpha = 1$.

Then, the unconditional probability of player i's choices is

$$P_i^* \equiv \sum_{k=-1}^{K} \alpha_k P(i, k), \tag{6}$$

and the log-density of the whole data set, x, is

$$L(x) \equiv \sum_{i} \ln(P_i^*). \tag{7}$$

3. THE STATISTICAL METHODOLOGY

The log-density L defined by Eq. (7) is a function of the parameters of the model: the α_k's, β_0, β_1, μ, r_0, s_1, r_1, s_2, r_2, and r_3. We seek to find values for

[7] These "irrational" choices suggest that these players were more interested in aggravating their peers than in winning. While such behavior is not itself a random draw from a uniform distribution, the -1 type behavior is a parsimonious way to capture this as well as all other behavior that is not consistent with any of the step-k rules.

these parameters that maximize L, and we will call these parameter values the *maximum density estimates* and denote them as $\hat{\beta}$.

When we proceeded to maximize L, we encountered a heteroscedasticity problem. Specifically, because the choices of the individuals decrease substantially over the periods, and because the data indicate a substantial decrease in the standard deviation over the periods, the densities (being divided by the standard deviation) are an order of magnitude larger in the last period than in the first period. This imbalance means that the fourth period data have a disproportionate influence on the coefficient estimates.

To correct for this problem, we transformed the data by multiplying the choice in period t by $50/\bar{x}_{t-1}$. Thus, in terms of the transformed data, a step-k rule will generate a mean choice of $50\rho^k$ in every period and a standard deviation of $50s_1/\bar{x}_{t-1}+r_1\rho^k$ for period t. Note that the relative standard deviation component becomes independent of the period, while the absolute component becomes significant only in the later period. All results reported below are based on the transformed data.

We maximize L using the simplex algorithm of Nelder and Mead (1965), using a variety of starting values to increase our confidence that a global maximum was achieved. The simplex method requires only function evaluations. Although it is not very efficient in terms of the number of function evaluations required, the method is easier to implement than other algorithms sometimes used with mixture models, such as the EM-algorithm.

Nonparametric confidence intervals for parameter estimates were estimated using the bootstrap percentile method described by Efron (1982, Chap. 10). While it is possible to obtain conventional standard error estimates by evaluating the information matrix derived from the log-likelihood function, interpretation of t-ratios obtained from these estimates is problematic due to the nonnormality of the underlying distributions.

The bootstrap method is based on the notion of replicating error processes by resampling estimated residuals. Since our model is a probabilistic choice model, residual terms are not explicitly estimated, and so the simulation step requires some modification. We first maximize the log-likelihood, Eq. (7), using the actual dataset to obtain a vector of parameter estimates, $\hat{\beta}$. Then, to generate pseudo-data x_s for each player, a uniform $[0, 1]$ pseudorandom deviate is generated via the multiplicative congruential method and compared to the estimates $\hat{\alpha}_k$, $k \in \{-1, 0, \ldots, 3\}$, to determine the player's disposition type. If the type is -1, then a uniform [0,100] pseudorandom deviate is generated for every period as the player's choice. Otherwise, a uniform [0,1] pseudorandom deviate is generated and compared to the $\varphi(k, j, l)$, using $\hat{\mu}$, to determine the rule used by the player for period 1. If the rule is step-k ($k \geq 0$), then normal pseudorandom deviates (with mean $50\rho^k$ and standard deviation σ^k) are generated until one lies in the [0,100] interval, and this one becomes the player's choice.

Sequentially, for periods $t = 2, 3$, and 4 and dispositions $k \geq 0$, the propensity

weights $w(k, j, t)$ are computed according to the dynamic, Eq. (2), using $\hat{\beta}$. Then, a uniform [0,1] pseudorandom deviate is generated and compared to the $\varphi(k, j, t)$ to determine the rule used by the player for period t, given the disposition type already determined (above). If the rule is step-k ($k \geq 0$), then normal pseudorandom deviates (with mean m_k equal to ρ^k times the mean of period $t-1$ of the pseudodata and standard deviation σ_k) are generated until one lies in the [0,100] interval, and this one becomes the player's choice.

Since the original data consist of seven sessions with differing sizes and treatments, we simulate seven sessions of pseudo-data with the same number of players and the same treatment as in the original. Once seven sessions of pseudo-choices x_s have been simulated, the model is reestimated using these pseudo-data to obtain a bootstrap estimate $\hat{\beta}_s$. Then the process is repeated a large number of times to produce S bootstrap estimates, $\{\hat{\beta}_s\}_{s=1}^{S}$, which approximate the sampling distribution of the original estimator, $\hat{\beta}$. Let $\hat{\beta}(j)$ and $\hat{\beta}_s(j)$ denote the jth elements of $\hat{\beta}$ and $\hat{\beta}_s$, respectively. Nonparametric 95%-confidence intervals for $\hat{\beta}(j)$ are obtained by sorting $\{\hat{\beta}_s(j)\}_{s=1}^{S}$ by algebraic value and then deleting $0.025 \times S$ values from each end of the sorted array; the new endpoints give the confidence interval. In the results reported below, we chose $S = 1000$ to ensure adequate coverage.

The bootstrap estimates are also used to compute standard error estimates and "t-ratios" which are reported below. The bootstrap standard-error estimates are obtained by computing the sample standard deviation of $\{\hat{\beta}_s(j)\}_{s=1}^{S}$. As noted earlier, the interpretation of t-ratios and standard errors is problematic due to the nonnormality of the underlying distributions.

Since the model presented in Section 2 explicitly includes the treatment parameter ρ, we maximize L for the pooled data from two treatments: one with $\rho = 1/2$ and the other with $\rho = 2/3$. We present tests for treatment effects in Section 4.4. The first treatment consisted of three separate experimental sessions with a total of 48 players, while the second treatment consisted of four sessions with a total of 67 players. (See Nagel for the details of the experiment.)

4. RESULTS

We first present the model with $K = 3$, i.e., step-0 to step-3 behavioral rules. To simplify the presentation, we first consider the test of the hypothesis that the parameters of the f_k and g_k functions are the same: i.e., $s_2 = s_1$, $r_2 = r_1$, and $r_3 = r_0$. The maximized value of the unrestricted log-density function was 304.528, while the maximized value of the restricted log-density function was 301.928. Twice the difference is 5.200, which has a chi-square distribution with 3 degrees of freedom, giving a p-value of 0.158. Thus, we fail to reject the hypothesis. In other words, we have discovered that a single error process can

TABLE I
Parameter Estimates of Model with $K = 3$

	Estimate	Std Dev	t-Ratio	95% Conf. Interval	
α_{-1}	0.0566	0.0240	2.362	0.0153	0.1118
α_0	0.4153	0.0713	5.825	0.2749	0.5520
α_1	0.2109	0.0672	3.142	0.0839	0.3375
α_2	0.2787	0.0633	4.403	0.1477	0.4038
α_3	0.0384	0.0342	1.125	0.0000	0.1177
r_0	0.5168	0.0612	8.444	0.4086	0.6607
r_1	0.9469	0.1080	8.769	0.7182	1.1471
s_1	0.1356	0.1333	1.017	0.0000	0.4811
β_0	0.6037	0.1331	4.536	0.3391	0.8630
β_1	0.1721	0.0340	5.060	0.1201	0.2549
μ	3.154	1.132	2.787	2.015	6.237

Estimated Maximum Log-Density = 301.928.

account for diversity in both the choice mechanism (f_k) and the reinforcement mechanism (g_k). Our remaining analysis imposes the restriction that $s_2 = s_1$, $r_2 = r_1$, and $r_3 = r_0$.

Table I presents the maximum density estimates of the parameters for this model. The estimated proportions of the initial disposition types in the population (the α's) are all positive. For types -1, 0, 1, and 2, the lower bound of the 95% (and 99%) confidence interval for the α's is strictly positive; hence, we can conclude that these α's are significantly different from zero at all normal significance levels. (We will return to the issue of the statistical significance of $\hat{\alpha}_3$ momentarily.) The relative standard deviation estimate for the step-0 rule, $\hat{r}_0$, is 51.7%, while the relative standard deviation estimate for the higher step rules, $\hat{r}_1 \cdot \theta$, is 31.6% and 18.9% for $\rho = 1/2$ and 2/3 respectively. The absolute standard deviation estimate, $\hat{s}_1$, is 0.2, and so becomes a factor only in periods 3 and 4 when the average choices become small also. To test the significance of $\hat{s}_1$, we estimated the model under the restriction that $s_1 = 0$, and twice the resulting log-likelihood ratio was 14.774, which has a p-value less than 1.2×10^{-4}; thus, we conclude that s_1 is significantly different from 0.

The estimate $\hat{\mu} = 3.15$ means that the probability of a player behaving according to his/her true disposition in period one is 0.887, while the probability of any other behavior is 0.038. Thus, with an initial disposition toward step-k rule, that rule is 23 times more likely to be used than any other rule.

The dynamic parameters (β_0 and β_1) are both positive and less than 1. Thus, the past propensity tends to be discounted and learning is slow. Noting that the maximum density of a normal distribution with standard deviation σ is $1/\sqrt{2\pi}\sigma$, and given the parameter estimates for s_1 and r_1, the maximum density for g_k

ranges from 4.97 to 18.9, so $\beta_1 g_k$ ranges from 0.86 to 2.22, allowing substantial switching to the best-performing rule.

Table II displays the estimated $\varphi(k, j, t)$ functions by session and by period. It is evident from this table that there is substantial learning of which rules perform best. In the first period, each dispositional type has a strong propensity toward its own type behavior. In periods 2 and 3, since the step-2 rule performed best in the previous period, there is a strong tendency for each type to switch to rule 2. In period 4, this trend toward rule 2 continues for sessions 1, 4, 6, and 7, while there emerges some tendency toward rule 3 for sessions 2, 3, and 5.[8]

We turn now to the issue of whether the α_3 parameter estimate with zero as its left-hand endpoint of the 95% confidence interval is significantly positive or not. First, we separately estimate the respective restricted model (with the parameter set to zero) and calculate the likelihood-ratio statistic. Twice the difference of the maximized log-density function is 1.320. If this statistic were distributed χ^2 with one degree of freedom, then we would conclude that it is not significantly different from zero. Unfortunately, the likelihood-ratio statistic has unknown distribution under the null hypothesis since the null values of the parameters are on the boundary of the parameter space (see Everitt and Hand, 1981, and Titterington *et al*., 1985). Conventional Wald and Lagrange multiplier tests also have unknown distributions at the edge of the parameter space.

To circumvent this problem, we use the bootstrap procedure to approximate the sampling distribution of the likelihood-ratio statistic. The choice data were simulated as outlined above for the bootstrap procedure (except $\hat{\beta}_{\text{restricted}}$ was used to simulate the pseudo-data under the null hypothesis). Both the restricted and unrestricted models were estimated on these psuedo-data yielding maximized log-likelihood values $L^*_{\text{restricted}}$ and L^*_{free} respectively. These values were then used to compute a bootstrap estimate of the likelihood-ratio statistic: $\chi^* \equiv 2(L^*_{\text{free}} - L^*_{\text{restricted}})$. The entire process was repeated 1000 times to produce bootstrap estimates $\{\chi^*_s\}_{s=1}^{1000}$. Since these values approximate the sampling distribution of the original likelihood-ratio statistic, it is straightforward to determine the significance of the original likelihood-ratio statistic by first sorting the values and then determining the percentile of the original statistic.[9]

The bootstrap results for the $\alpha_3 = 0$ hypothesis indicate that the 5% critical value for the likelihood-ratio statistic is 2.844, and that 1.320 has a p-value of

[8] Therefore, it did not seem reasonable to consider $K < 3$.

[9] Our bootstrapped hypothesis test is methodologically identical to the Monte Carlo approach employed by Aitken *et al*. (1981), who in effect used only 19 bootstrap replications. Hall (1986) provides theoretical results which show that using a small number of replications in the bootstrap may increase the probability of type-II errors. Several authors have suggested using at least 100 replications for testing null hypotheses regarding parameter values, and as many as 1000 replications for constructing confidence intervals. Note that merely examining the confidence interval ignores the variation in the other parameters; thus, the bootstrapped likelihood-ratio statistic provides a more powerful test.

TABLE II
Rule Probabilities $\varphi(k, j, t)$

Disposition	Step 0	Step 1	Step 2	Step 3
All Sessions: Period 1				
0	0.8865	0.0378	0.0378	0.0378
1	0.0378	0.8865	0.0378	0.0378
2	0.0378	0.0378	0.8865	0.0378
3	0.0378	0.0378	0.0378	0.8856
Session 1				
Period 2				
0	0.4917	0.0834	0.3517	0.0732
1	0.0692	0.5292	0.3324	0.0692
2	0.0283	0.0322	0.9113	0.0283
3	0.0732	0.0834	0.3518	0.4916
Period 3				
0	0.1861	0.0723	0.6757	0.0659
1	0.0573	0.2219	0.6567	0.0640
2	0.0253	0.0310	0.9154	0.0283
3	0.0581	0.0713	0.6657	0.2049
Period 4				
0	0.0895	0.0595	0.7974	0.0537
1	0.0440	0.1174	0.7857	0.0529
2	0.0255	0.0339	0.9100	0.0306
3	0.0443	0.0590	0.7902	0.1065
Session 2				
Period 2				
0	0.5842	0.1265	0.2124	0.0770
1	0.0710	0.6929	0.1733	0.0629
2	0.0507	0.0737	0.8308	0.0449
3	0.0923	0.1341	0.2252	0.5485
Period 3				
0	0.2440	0.1014	0.4003	0.2543
1	0.0735	0.3042	0.3805	0.2418
2	0.0456	0.0597	0.7448	0.1499
3	0.0559	0.0734	0.2897	0.5810
Period 4				
0	0.0728	0.0442	0.2223	0.6608
1	0.0361	0.0877	0.2205	0.6557
2	0.0306	0.0372	0.3751	0.5571
3	0.0224	0.0272	0.1367	0.8138

TABLE II
Continued

Disposition	Step 0	Step 1	Step 2	Step 3
		Session 3		
Period 2				
0	0.4904	0.0839	0.3545	0.0712
1	0.0688	0.5305	0.3337	0.0670
2	0.0280	0.0322	0.9125	0.0273
3	0.0738	0.0848	0.3582	0.4832
Period 3				
0	0.1954	0.0738	0.6405	0.0904
1	0.0603	0.2272	0.6244	0.0881
2	0.0275	0.0328	0.8995	0.0402
3	0.0583	0.0695	0.6034	0.2687
Period 4				
0	0.0780	0.0434	0.2563	0.6223
1	0.0388	0.0866	0.2552	0.6195
2	0.0320	0.0357	0.4213	0.5110
3	0.0246	0.0274	0.1618	0.7862
		Session 4		
Period 2				
0	0.5955	0.1349	0.1958	0.0737
1	0.0701	0.7167	0.1549	0.0583
2	0.0550	0.0837	0.8156	0.0457
3	0.0970	0.1475	0.2141	0.5414
Period 3				
0	0.2791	0.1408	0.5117	0.0684
1	0.0795	0.3993	0.4598	0.0615
2	0.0462	0.0736	0.8444	0.0358
3	0.0924	0.1471	0.5348	0.2257
Period 4				
0	0.1321	0.1045	0.7134	0.0500
1	0.0636	0.2014	0.6869	0.0481
2	0.0400	0.0634	0.8663	0.0303
3	0.0671	0.1062	0.7250	0.1017

0.262. Thus, at the 5% confidence level, we cannot reject the hypothesis that $\alpha_3 = 0$ (i.e., that there are no step-3 dispositions in the sample population). Note that, despite this conclusion, step-3 behavior emerges from learning in three out of seven sessions.

TABLE II
Continued

Disposition	Step 0	Step 1	Step 2	Step 3
		Session 5		
Period 2				
0	0.5654	0.1132	0.2504	0.0710
1	0.0722	0.6521	0.2148	0.0609
2	0.0432	0.0581	0.8624	0.0364
3	0.0911	0.1225	0.2710	0.5154
Period 3				
0	0.2673	0.0950	0.3755	0.2622
1	0.0828	0.2933	0.3673	0.2565
2	0.0520	0.0584	0.7285	0.1611
3	0.0612	0.0687	0.2715	0.5986
Period 4				
0	0.1049	0.0520	0.2254	0.6177
1	0.0524	0.1042	0.2255	0.6179
2	0.0447	0.0443	0.3846	0.5264
3	0.0335	0.0332	0.1439	0.7894
		Session 6		
Period 2				
0	0.5207	0.0830	0.3212	0.0751
1	0.0752	0.5404	0.3115	0.0729
2	0.0324	0.0347	0.9015	0.0314
3	0.0786	0.0842	0.3257	0.5116
Period 3				
0	0.2228	0.0745	0.6166	0.0862
1	0.0700	0.2332	0.6114	0.0855
2	0.0324	0.0342	0.8938	0.0396
3	0.0683	0.0721	0.5964	0.2633
Period 4				
0	0.1746	0.1492	0.5943	0.0818
1	0.0821	0.2813	0.5596	0.0770
2	0.0579	0.0990	0.7889	0.0543
3	0.0877	0.1500	0.5976	0.1647

4.1. *Comparison with Alternative Nested Models*

4.1.1. *The No-Rule-Learning Hypothesis*. We can formally test the hypothesis of no "rule learning" by restricting $\beta_0 = 1$ and $\beta_1 = 0$. The maximum density estimate of μ is 1.81, implying initial dispositions that have 67% probability of true type behavior and a 11% probability of each other type of behavior. Note that the restricted model still permits learning in the sense of incorporating new

TABLE II
Continued

Disposition	Step 0	Step 1	Step 2	Step 3
		Session 7		
Period 2				
0	0.5459	0.1017	0.2829	0.0695
1	0.0728	0.6116	0.2534	0.0622
2	0.0378	0.0473	0.8827	0.0323
3	0.0872	0.1091	0.3034	0.5004
Period 3				
0	0.2483	0.1134	0.5767	0.0616
1	0.0732	0.3331	0.5364	0.0573
2	0.0379	0.0547	0.8777	0.0279
3	0.0817	0.1178	0.5987	0.2018
Period 4				
0	0.1175	0.0761	0.7489	0.0575
1	0.0577	0.1497	0.7361	0.0565
2	0.0347	0.0450	0.8863	0.0340
3	0.0588	0.0762	0.7498	0.1153

information into a fixed rule, but no systematic switching that depends on past rule performance.[10] The maximized value of the restricted log-density function was 263.210. Twice the difference is 77.436, which given a χ^2 distribution with 4 degrees of freedom would have a p-value less than 10^{-16}; hence, under the χ^2 assumption, we would strongly reject the hypothesis of no rule learning.[11]

4.1.2. *Simple Adaptive Behavior.* The simple adaptive behavior model, mentioned in the Introduction, is nested within our model: $\beta_0 = 1$, $\beta_1 = 0$, $\mu = \infty$, and $\alpha_k = 0$ for $k \neq 1$. That is, there is only the step-1 rule, hence no learning of rules, only updating of this step-1 rule: $x_{t+1} = (1-\lambda)\rho\bar{x}_t + \lambda x_t$, where $\bar{x}_0$ and x_0 are initialized to 50.[12] Because of the outliers, we also include a -1 type in the model. The maximized log-density function is 115.359, which is enormously less than our model. Thus, we can strongly reject the simple adaptive behavior model in favor of our model.

[10] The random iid switching allowed by the μ parameter can be suppressed by setting $\mu = \infty$; doing this, the maximized log-density function drops to 224.364.

[11] Given our experience, bootstrapping will not alter this conclusion, so we did not expend the considerable resources to bootstrap this test.

[12] If $\lambda > 0$, then this adaptive expectations model is not nested; however, the maximized log-density estimate $\hat{\lambda}$ is 0, so the estimated model is nested. Since the corresponding likelihood ratio statistic is so large, the issue of the number of degrees of freedom can be resolved liberally without affecting our conclusion.

4.2. *Comparison with Alternative Non-nested Models*

4.2.1. *Homogeneous vs Heterogeneous Dispositions*. Consider a *homogeneous rule-learning model* which differs from our mixture model only in that each participant is assumed to be one of two types: type -1 (defined as before) or type ℓ for "learner." All learners are assumed to have the same initial disposition. Let ν_k denote the probability that the latter type uses the step-k rule in period one. After period one, learning is assumed to occur according to the same dynamic (β_0, β_1) and the same noise parameters (r_0, s_1, r_1). (The four new parameters replace the μ parameter of our mixture model, and we also eliminate $\alpha_0, \ldots, \alpha_3$, leaving a total of eight parameters to be estimated.) The maximized log-density value was 294.635, which is less than the maximized log-density value of the heterogeneous mixture model by 7.293. Since these models are not nested, we cannot use a χ^2 test of significance.

Therefore, we conducted a Monte Carlo simulation of the likelihood ratio statistic (Schork, 1992). Similar to the bootstrap procedure discussed above, we generated 1000 pseudo-data sets using the parameter estimates of the alternative homogeneous rule-learning model. For each pseudo-data set, we estimated the original model and the alternative model and computed the log-likelihood difference. These Monte Carlo results indicate that the 5% critical value of the log-likelihood difference is 1.110, and since the largest simulated log-likelihood difference was 5.774, the statistic 7.293 has a p-value less than 0.001. Thus, we have strong evidence in favor of heterogeneous dispositions as opposed to homogeneous dispositions.

4.2.2. *Adaptive Ratio-Forecasting*. The simple adaptive behavior model considered in Section 4.1.2 did not perform well for reasons similar to why similar price-level forecasting predicts badly in an inflationary economy. Since the mean $\bar{x}_t$ is falling, it would be better to attempt to forecast the rate of fall: $\gamma(t+1) \equiv \bar{x}_{t+1}/\bar{x}_t$. We call this *adaptive ratio forecasting*, and specify a one-parameter forecasting rule:

$$\gamma(t+1) = \lambda\, \bar{x}_t/\bar{x}_{t-1} + (1-\lambda)\gamma(t), \qquad \text{with } \gamma(1) = 1 \text{ and } \bar{x}_0 = 50.$$

The choice is the best response to this forecast: $x_{t+1} = \rho\gamma(t+1)\bar{x}_t$. We also include the -1 type.

The maximum log-density function for this model is 190.132, which is a substantial improvement over simple adaptive expectations. However, this is still substantially less than our model, and while this alternative is not nested within our model, the difference is large enough to persuade us (without conducting Monte Carlo simulations) to reject adaptive ratio forecasting in favor of our model.

4.2.3. *A Two-Lag Forecasting Rule*. Recognizing that the class of rules in our model use only the most recent mean choice, we decided to investigate a model consisting of a type that operates like the step-1 rule in the first period, but thereafter uses the two most recent means. For period $t \geq 2$, the expected mean of the population for period $t+1$ is given

$$x^e(t+1) \equiv \lambda_1 \bar{x}_t + \lambda_2 \bar{x}_{t-1}.$$

In addition, we included a -1 type. The maximized log-density function was 216.805, an improvement over the above alternative, but still substantially less than our model.

4.2.4. *Nagel's Learning Hypothesis*. Nagel (1995) suggested a qualitative "directional" learning model in which each individual's choice is a multiple of the previous mean, and this multiplicative factor is adjusted up (down) if it was too low (high) in the previous period. This notion is similar to our learning model in that players who start by using the step-0 rule will eventually use a lower multiplicative factor, and players who start by using the step-3 rule will tend to use a higher multiplicative factor in periods 2 and 3 (and sometimes in period 4 as well). To make a rigorous comparison, we must first develop a parameterized version of Nagel's suggestion. We will include a type -1 to accommodate outliers, as we did in our model.

For all other players, we assume that in the first period they use one of the step-k rules, $k = 0, 1, \ldots, 3$, and let α_k denote the proportion using the step-k rule. Then, the probability of player i's first period choice is

$$\tilde{p}(i, 1) \equiv \sum_{k=0}^{K} \alpha_k \cdot f_k(x(i, 1); 50), \tag{8}$$

where $f_k(\,)$ is the same as in our model. For $t = 2$, 3, and 4, the expected choice of player i is $\gamma(i, t)\bar{x}_{t-1}$, where

$$\gamma(i, t) \equiv \lambda\gamma(i, t-1) + (1-\lambda)\rho\bar{x}_{t-1}/\bar{x}_{t-2} \qquad \text{and} \qquad \gamma(i, 1) \equiv x(i, 1)/50. \tag{9}$$

In other words, the player adjusts his/her multiplicative factor in the direction of the most recent best factor. We specify that the choice $x(i, t)$ is distributed as a normal random variable, truncated to the [0,100] interval, with mean $\gamma(i, t)\bar{x}_{t-1}$ and standard deviation $\sigma = s + r \cdot \gamma(i, t)\bar{x}_{t-1}$.

There are 10 parameters to estimate (four α's, r_0, s_1, r_1, s, r, and λ). The maximized log-density is 267.705, which is a substantial improvement over the other alternative models, but still 34.223 less than our model. To investigate how significant this difference is, we conducted a Monte Carlo simulation of this likelihood ratio statistic (as described in 4.2.1). These Monte Carlo results indicate that the 5% critical value of the log-likelihood difference is -11.063,

and since the largest simulated log-likelihood difference was 10.679, the statistic 34.223 has a p-value less than 0.001. Thus, we reject the directional learning hypothesis in favor of our model.

4.2.5. *Testing for Step-4 Behavior*. We also considered the model with a step-4 behavioral rule (i.e., $K = 4$). For that model the maximized value of the log-density L was 301.895, which is slightly less than the log-density with $K = 3$ (301.928). Furthermore, the maximum density estimate of α_4 was identically zero. Note that these models are not nested since the set of rules as well as the number of dispositions changes; hence, L does not have to decrease in going from $K = 4$ to $K = 3$. On the contrary, by concentrating the φ functions on a smaller set of rules, the data are fit better, and L increases by 0.033.

To test the significance of this difference, we conducted a Monte Carlo simulation of this likelihood ratio statistic (as described in 4.2.1). These Monte Carlo results indicate that the 5% critical value of the log-likelihood difference is 13.08, and since the smallest simulated log-likelihood difference was 0.017, the statistic -0.083 has a p-value less than 0.001. Thus, we can reject the hypothesis that $K = 4$ in favor of the hypothesis that $K = 3$.

4.3. *Comparison with Bayesian Rule Learning*

By modifying the reinforcement function of our model, we can transform the model into one with Bayesian rule learning. Specifically, if instead of g_j in Eq. (2) we use $\log(f_j)$, then the rule probabilities become

$$\tilde{\varphi}(k, j, t) \equiv \frac{\tilde{\varphi}(k, j, t-1)^{\beta_0} f_j(t-1)^{\beta_1}}{\sum_\ell \tilde{\varphi}(k, \ell, t-1)^{\beta_0} f_\ell(t-1)^{\beta_1}}. \tag{8}$$

Then, specifying $\beta_0 = \beta_1 = 1$, Eq. (8) becomes Bayes formula for the posterior probability that rule j is best conditional on the period $t-1$ observation and the prior belief for period $t-1$. In other words, this modified model allows players to have different priors but then update according to Bayes rule.

When we estimate this alternative model with β_0 and β_1 unrestricted, we find a maximized log-density of 293.284, which is less than our original model (301.928), indicating that our specification of the reinforcement function fits the data better.[13] The maximum density estimates of β_0 and β_1 are 0.406 and 0.717 respectively. Thus, relative to Bayes rule, the participants in this experiment discounted the past too much and did not give adequate weight to current performance.

Restricting $\beta_0 = \beta_1 = 1$, the maximized log-density falls to 280.993. Twice the difference is distributed χ^2 with 2 degrees of freedom, and has a p-value less

[13] Our Monte Carlo simulation indicates that this difference is statistically significant at all commonly accepted levels.

TABLE III
Treatment Effects

	$\rho = 1/2$	$\rho = 2/3$	Combined
α_{-1}	0.0643	0.0464	0.0566
α_0	0.3168	0.5200	0.4153
α_1	0.1855	0.2185	0.2109
α_2	0.3765	0.2022	0.2787
α_3	0.0568	0.0129	0.0384
r_0	0.4996	0.5301	0.5168
r	0.8132	1.052	0.9469
s	0.1753	0.0001	0.1356
β_0	0.4426	0.7076	0.6037
β_1	0.1520	0.1842	0.1721
μ	5.132	2.500	3.154
$\hat{L}$	140.930	167.075	301.928
No. of Subjects	48	67	115

than 5×10^{-6}. Therefore, we strongly reject the Bayesian rule learning model.

One might object to this formulation of a Bayesian rule learning model for the following reason. Suppose players experiment in the first period. Then, our identification of their prior would be contaminated, and the restriction that $\beta_0 = 1$ would be unwarranted, especially for the first period. To investigate this possibility, we reestimated the model allowing β_0 to be less than or equal to one for the update from period 1 to period 2, but restricted to equal one for all subsequent updates. The maximum density estimate $\hat{\beta}_0$ was equal to one, clearly demonstrating that this modification does not improve the Bayesian rule learning model. We therefore reject both formulations of Bayesian rule learning as an explanation of this data.

4.4. *Treatment and Session Effects*

Does our model fully account for the effect of ρ? Since the data involve 48 players with the $\rho = 1/2$ treatment, and 67 players with the $\rho = 2/3$ treatment, we can formally test treatment effects. The results are displayed in Table III. Since $2 \times [L_{(1/2)} + L_{(2/3)} - L] = 14.156$ has a χ^2 distribution with 10 degrees of freedom and a p-value of 0.166, we cannot reject the hypothesis of no treatment effects. Thus, the effect of the treatment parameter of behavior appears to be adequately captured by the theory.

We can ask whether all sessions in one treatment were statistically indistinguishable. To do this, we estimated the model on each session separately. A priori we might expect some differences to show up since (1) there are fewer ob-

servations per session upon which to estimate the 10 parameters to the efficiency of the estimates declines, and (2) extensive data mining enhances the chances of type-II errors.

Indeed, we can reject at the 5% level the hypothesis that all the $\rho = 1/2$ sessions are identical; on the other hand, we cannot reject any one of the three $\rho = 1/2$ sessions as being statistically different from the pooled $\rho = 1/2$ group.[14] Further, we cannot reject at the 5% level the hypothesis that all the $\rho = 2/3$ sessions are identical, and we cannot reject any one of the four $\rho = 2/3$ sessions as being statistically different from the pooled $\rho = 2/3$ group.

Summing the session-by-session maximum density estimates and comparing this sum to the pooled data for both treatments (all 7 sessions), we can reject at the 5% level the hypothesis that all the sessions are identical; on the other hand, on an individual session basis we can reject only Session 1 as being statistically different from the pooled data.

Because Nagel's data include only four observations per player, it would not be possible to obtain useful estimates of individual player differences (such as learning rates), so the theoretical model proposed incorporates individual differences only through the four archetypal initial dispositions. Given this constraint, we would expect that individual differences among the players might manifest themselves as differences among sessions. Thus, we are comfortable with these test results.

4.5. *Relative Expected Performance of Disposition Types*

Given our estimated model, what would be the expected payoff to a player with a disposition type k in a large population of players? To answer this, we simulated 100,000 runs of the experiment (like the bootstrap procedure) and recorded the winners by disposition type. Table IV shows the results. The type-2 disposition which uses the step-2 rule initially has the highest expected payoff. The type-1 disposition is in second place; type-0 and type-3 dispositions are essentially tied for third place.

4.6. *Descriptive and Predictive Performance*

While our model achieves the maximum log-likelihood density of any model, that still does not tell us how well the model fits the data, and while Table II shows the estimated behavior of the rule probabilities (φ), it does not show us the corresponding differences in the real behavior. If we know the true initial dispositions of each participant, then we could average the choices for all step-k types and plot these averages over time to get a visual picture of the behavioral types. However, the best we can do is compute the Bayesian posterior probability

[14] We can reject Session 1 at the 10% level.

TABLE IV
Expected Payoff by Disposition Type

Type	No. by Type	Wins by Type	Exp. Payoff
−1	650539	33119	0.0509
0	4774720	898157	0.1880
1	2425533	585512	0.2414
2	3203913	1190313	0.3715
3	442543	92899	0.2099
Totals:	11,500,000	2,800,000	

that i had a step-k disposition:

$$\tilde{\alpha}_{ik} \equiv \hat{\alpha}_k \hat{P}(i,k)/\hat{P}_i^*, \tag{9}$$

where $\hat{P}(i,k)$ and $\hat{P}_i^*$ are given by Eqs. (5) and (6) evaluated at the estimated parameter values. Since there are five values for $k \in \{-1, 0, \ldots, 3\}$, but only four observations per participant, these calculated posterior probabilities contain little information individually. However, consider the following aggregation. As in Section 4.2.2 and 4.2.4, let $\gamma(i,t) \equiv x_{it}/\bar{x}_{t-1}$ denote participant i's multiplicative adjustment factor relative to the previous mean. Then, let

$$\tilde{\gamma}_k(t) = \left(\sum_i \tilde{\alpha}_{ik} \cdot \gamma(i,t)\right) / \left(\sum_i \tilde{\alpha}_{ik}\right), \tag{10}$$

which is the calculated population-averaged multiplicative adjustment factor for a step-k disposition, using the posterior probabilities as weights. Note that the denominator is an estimate of the number of participants with a step-k disposition. If the model fits perfectly, then in period one we would have $\tilde{\gamma}_k(1) = \rho^k$ (for $k \geq 0$), with gradual convergence with experience.

Figure 3 shows these plots disaggregated by treatments ($\rho = 1/2$ and $2/3$) for $k \in \{0, 1, 2, 3\}$. In period 1, the $\tilde{\gamma}_k(1)$ values correspond well to their theoretical values; noting this, it is easy to trace out each step-k curve. While step-2 and step-3 behaviors are not dramatically different, step-0, step-1, and the mean of steps-2&3 behaviors are quite distinct for the first two periods, and these differences become less distinct for periods three and four as predicted by the model.

We also computed the theoretical expected values and standard deviations of $\hat{\gamma}_k(t)$, given our parameter estimates and the observed means $\bar{x}_t$. In comparing these theoretical values with those of Fig. 3, since $\tilde{\gamma}_k(t)$ is also a random variable of unknown distribution, we use a 1% significance level to compensate for the increased likelihood of false rejection. Each of the 32 points plotted in Fig. 3 corresponds to an hypothesis that $\tilde{\gamma}_k(t) = \hat{\gamma}_k(t)$; only two are rejected at the 1%

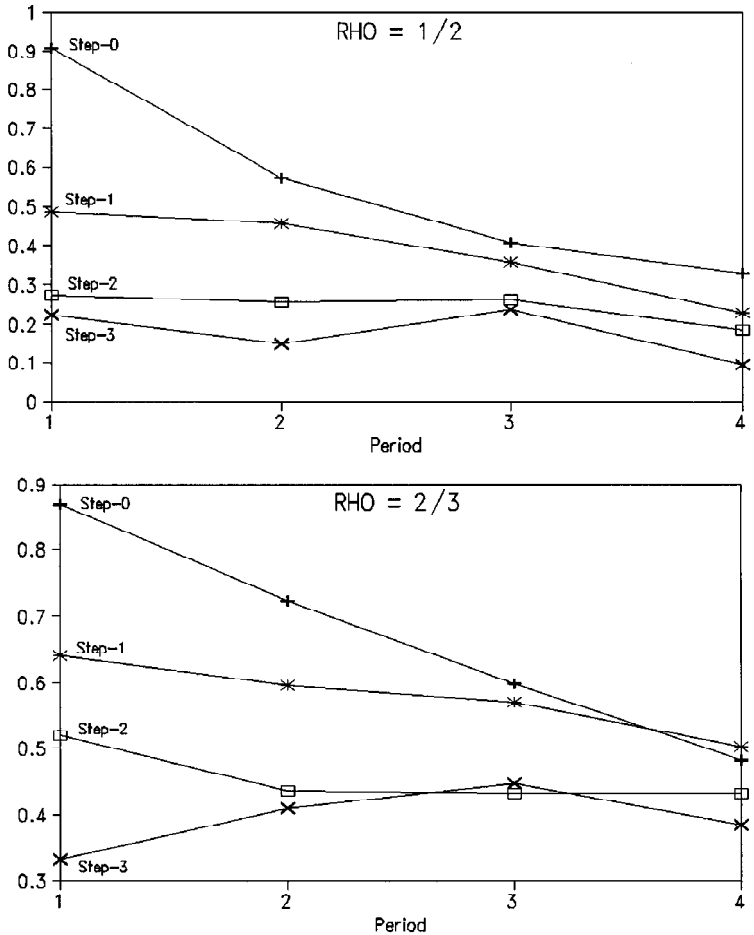

FIG. 3. Descriptive Performance $\tilde{\gamma}_k(t)$.

level. If these tests were independent, the probability of two or more rejections would be 0.041. Therefore our model appears to fit the data well.

5. CONCLUSIONS

We have specified a "law-of-effect" learning model in which players use one of a finite number of behavioral rules from Nagel's step-k hierarchy. The initial vector of probabilities of using the rules is called a disposition, and we

TABLE V
Summary of Maximized Log-Likelihood Density Estimates

Model	Section	$\hat{L}$
Rule-learning ($K = 3$)	4.0	301.928
Rule-learning ($K = 4$)	4.2.5	301.895
Homogeneous rule-learning	4.2.1	294.635
Unrestricted "Bayesian"	4.3	293.284
Bayesian rule-learning	4.3	280.993
Directional learning	4.2.4	267.705
No-rule-learning	4.1.1	263.210
Two-lag forecasting	4.2.3	216.805
Adaptive-ratio forecasting	4.2.2	190.132
Simple adaptive behavior	4.1.2	115.359

hypothesize that associated with each rule is a disposition type for which the associated rule is the most probable rule. The learning dynamic specifies that the propensities of the better performing rules increase while the propensities of the worse performing rules decrease. We then confronted this model with Nagel's experimental data.

We found that a model with step-0 to step-3 rules plus a completely random nonlearning type fits the data better than a model with step-0 to step-4 rules plus a completely random nonlearning type. Further, we could not reject the hypothesis that the initial dispositions were confined to the step-0 to step-2 rules. Nonetheless, over time some participants tend to switch to the step-3 rule. The estimated dynamic parameters indicate that the initial disposition is discounted substantially. Thus, while over a third of the population begins with random play, they quickly abandon that rule the better performing step-2 rule. On the other hand, a type-2 disposition has the greatest expected payoff over four periods of play.

There is strong evidence of rule learning as opposed to simple updating adaptation. We examined a model with a homogeneous disposition for all participants and rejected it in favor of heterogeneous initial dispositions. Our model performs vastly better than the simple adaptive behavior model and better than two variants of adaptive behavior. We developed a parameterized specification of Nagel's directional learning hypothesis, tested and subsequently rejected it in favor of our model. We also tested and rejected the Bayesian rule-learning hypothesis in favor of our model. For ease of comparison, Table V presents the maximum density estimates for the models considered.

In contrast to Nagel (1995), we interpret the evidence as providing strong support for increasing depth of reasoning. In the early periods, a sufficient proportion of players use step-0 and step-1 rules, so the step-3 and higher rules do not perform well; thus, the learning dynamic decreases their likelihood, giving the appearance of no increasing depth of reasoning in the aggregate. On the other

hand, we find that type-0 and type-1 players do learn to switch to the step-2 rule, and in some sessions, by the fourth period the proportion of players using the step-3 rule increases. Theoretically, our learning model will eventually drift toward higher and higher step-k rules, but there are insufficient periods in this data for a conclusive test of this prediction.[15]

One shortcoming of our model is that, due to having only four periods of data, we did not allow for heterogeneous rates of learning (β_0 and β_1) in the population. When confronting data over many more periods, it will be possible to allow for heterogeneous learning rates within each disposition type and to look for correlation between dispositions and learning rates.

In conclusion, we have found (1) that individuals are heterogeneous in their initial behavior which can be characterized by Nagel's (1995) step-k hierarchy of rules and (2) that a law-of-effect learning model for these step-k rules explains individual behavior over time far better than extant alternatives.

The approach of this paper may be applicable to a wide class of games, especially those in which each player need only predict a single summary statistic of what everyone else is doing (e.g., Van Huyck *et al*., 1990, 1991). It would be interesting to contrast this approach with that of Crawford (1995). In ongoing research (Stahl, 1996), our learning model is being extended to apply to typical normal-form games.

REFERENCES

Aitken, M., Anderson, D., and Hinde, J. (1981). "Statistical Modelling of Data on Teaching Styles," *J. Roy. Statist. Soc. A* **144**, 419–461.

Arthur, B. (1990). "A Learning Algorithm that Mimics Human Learning." Stanford University.

Arthur, B. (1993). "On Designing Economic Agents that Behave Like Human Agents," *J. Evolutionary Econ.* **3**, 1–22.

Aumann, R. (1986). "Rationality and Bounded Rationality," Nancy L. Schwartz Lecture. Kellogg School of Management, Northwestern University.

Blume, L., and Easley, D. (1992). "Rational Expectations and Rational Learning." Department of Economics, Cornell University.

Bush, R., and Mosteller, F. (1955). *Stochastic Models of Learning*. New York: Wiley.

Chen, X., and White, H. (1994). "Nonparametric Adaptive Learning with Feedback." University of California at San Diego.

Cheung, Y-W., and Friedman, D. (1994). "Learning in Evolutionary Games: Same Laboratory Results." Economics Department, Univ. of Calif., Santa Cruz.

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

[15] Extrapolating from this data, we would predict that the mean of a fifth period would be less than 1, and given that human subjects tend to round off real numbers to the nearest integer, we believe the signal-to-noise ratio would drop drastically after the fourth period. If we want quality data for six or seven periods, we should try enlarging the choice set to [0,10000].

Efron, B. (1982). *The Jackknife, the Bootstrap, and Other Resampling Plans*. Philadelphia: Society for Industrial and Applied Mathematics.

El-Gamal, M., McKelvey, R., and Palfrey, T. (1993). "A Bayesian Sequential Experimental Study of Learning in Games," *J. Amer. Statist. Assoc*. **88**, 428–435.

El-Gamal, M., McKelvey, R., and Palfrey, T. (1994). "Learning in Experimental Games," *Econ. Theory* **4**, 901–922.

Everitt, B., and Hand, D. (1981). *Finite Mixture Distributions*. London: Chapman and Hall.

Friedman, D., Massaro, D., Kitsis, S., and Cohen, M. (1995). "A Comparison of Learning Models," *J. Math. Psychology* **39**, 164–178.

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 Equilibrium," *Econometrica* **61**, 547–573.

Hall, P. (1986). "On the Number of Bootstrap Simulations Required to Construct a Confidence Interval," *Ann. Statist*. **14**, 1453–1462.

Holland, J. (1976). "Adaptation," in *Progress in Theoretical Biology IV*, (R. Rosen and F. Snell, Eds.), pp. 263–293. New York: Academic Press.

Holt, D. (1993). "An Empirical Model of Strategic Choice with an Application to Coordination Games." Department of Economics, Queen's University.

Jordan, J. (1991). "Bayesian Learning in Normal Form Games," *Games Econ. Behav*. **3**, 60–81.

Kalai, E., and Lehrer, E. (1993). "Rational Learning Leads to Nash Equilibrium," *Econometrica* **61**, 1019–1045.

Kuan, C., and White, H. (1994). "Adaptive Learning with Nonlinear Dynamics Driven by Dependent Processes," *Econometrica* **62**, 1087–1114.

Milgrom, P., and Roberts, J. (1991). "Adaptive and Sophisticated Learning in Normal Form Games," *Games Econ. Behav*. **3**, 82–100.

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

Nagel, R. (1995). "Unraveling in Guessing Games: An Experimental Study," *Amer. Econ. Rev*. **85**, 1313–1326.

Nelder, J., and Mead, R. (1965). "A Simplex Method for Function Minimization," *Comput. J*. **7**, 308–313.

Nyarko, Y. (1991). "Bayesian Learning without Common Priors and Convergence to Nash Equilibrium," New York University.

Rosenthal, R. (1993). "Rules of Thumb in Games," *J. Econ. Behav. Organ*. **22**, 1–13.

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** (Special Issue: Nobel Symposium), 164–212.

Schork, N. (1992). "Bootstrapping Likelihood Ratios in Quantitative Genetics," in *Exploring the Limits of the Bootstrap* (R. LePage and L. Billard, Eds.), pp. 389–396. New York: Wiley.

Selten, R. (1990). "Anticipatory Learning in Games," in *Game Equilibrium Models*, Vol. I, *Evolution and Game Dynamics* (R. Selten, Ed.). New York/Berlin: Springer-Verlag.

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

Selten, R., and Buchta, J. (1994). "Experimental Sealed Bid First Price Auction with Directly Observed Bid Functions," Discussion Paper 270. Universität Bonn.

Stahl, D. (1993). "The Evolution of Smart$_n$ Players," *Games Econ. Behav*. **5**, 604–617.

Stahl, D. (1996). "Evidence-Based Rule-Learning in Normal-Form Games." Department of Economics, Working Paper 9601, Univ. of Texas.

Stahl, D., and Wilson, P. (1994). "Experimental Evidence of Players' Models of Other Players," *J. Econ. Behav. Organ*. **25**, 309–327.

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

Titterington, D., Smith, A., and Makov, U. (1985). *Statistical Analysis of Finite Mixture Distributions*. Chichester: Wiley.

Van Huyck, J., Battalio, R., and Beil, R. (1990). "Tacit Coordination Games, Strategic Uncertainty, and Coordination Failure," *Amer. Econ. Rev*. **80**, 234–248.

Van Huyck, J., Battalio, R., and Beil, R. (1991). "Strategic Uncertainty, Equilibrium Selection, and Coordination Failure in Average Opinion Games," *Quart. J. Econ*. **106**, 885–910.

Van Huyck, J., Battalio, R., Mathur, S., Ortmann, A., and Van Huyck, P. (1995). "On the Origin of Convention: Evidence from Symmetric Bargaining Games." *Int. J. Game Theory* **24**, 187–212.