

# Experience-Weighted Attraction Learning in Coordination Games: Probability Rules, Heterogeneity, and Time-Variation

Colin Camerer

*California Institute of Technology*

and

Teck-Hua Ho

*The Wharton School, University of Pennsylvania*

---

In earlier research we proposed an “experience-weighted attraction (EWA) learning” model for predicting dynamic behavior in economic experiments on multiperson noncooperative normal-form games. We showed that EWA learning model fits significantly better than existing learning models (choice reinforcement and belief-based models) in several different classes of games. The econometric estimation in that research adopted a representative agent approach and assumed that learning parameters are stationary across periods of an experiment. In addition, we used the logit (exponential) probability response function to transform attraction of strategies into choice probability.

This paper allows for nonstationary learning parameters, permits two “segments” of players with different parameter values in order to allow for some heterogeneity, and compares the power and logit probability response functions. These specifications are estimated using experimental data from weak-link and median-action coordination games. Results show that players are heterogeneous and that they adjust their learning parameters over time very slightly. Logit probability response functions never fit worse than power functions, and generally fit better. © 1998 Academic Press

---

## 1. INTRODUCTION

Most game theorists now agree that equilibration arises in games because players learn or evolve, rather than figuring equilibria out by introspection. An important

Direct correspondence and reprint requests to Teck H. Ho, Marketing Department, The Wharton School, University of Pennsylvania, Philadelphia, PA 19004. E-mail: hoteck@marketing.wharton.upenn.edu.

This research was supported by NSF grants SBR-9511001 and SBR-9511137. Hongjai Rhee and Dana Peele provided excellent research assistance.

empirical challenge is explaining how this learning occurs, preferably with a model that is psychologically plausible and fits data well.

In Camerer and Ho (1997), we propose an “experience-weighted attraction” (EWA) learning model and show that it accounts for dynamic behavior well in three classes of experimental games. The EWA learning model has three desirable properties:

- EWA satisfies principles of actual, simulation, and declining effects. The principle of actual effect states that successes will increase choice probability of chosen strategies. This principle corresponds to the “law of effect” widely discussed in behaviorist psychology (Thorndike, 1911; Herrnstein, 1970). The principle of simulated effect states that unchosen strategies which would have yielded high payoffs (i.e., simulated successes) are more likely to be chosen. This principle suggests that players move to reduce *ex post* difference (or error) between the actual and foregone payoffs, and is also widely used in cognitive psychology (e.g., connectionist neural networks). The principle of declining effect states that the effect of payoffs on choices diminishes over time. This principle captures the fact that dynamic behavior in economic experiments often converge to stable behavior over time.

- EWA contains two well-known and very different approaches as special cases. One approach, belief-based models, starts with the premise that players keep track of the history of previous play by other players and form some belief about what others will do in the future based on past observation. Then they choose a best-response strategy which maximizes their expected payoffs, given the beliefs they formed (Brown, 1951; Fudenberg & Levine, *in press*). The other approach, choice reinforcement, assumes that strategies are “reinforced” by their previous payoffs, and the propensity to choose a strategy depends in some way on its stock of cumulative reinforcement. Reinforcement models are belief-free: Players care only about the payoffs strategies yielded in the past, not about the history of play that created those payoffs (Bush & Mosteller, 1955; Harley, 1981; Arthur, 1991; Roth & Erev, 1995). The EWA model includes the belief-based and reinforcement approaches as special cases, but it is not simply a convex combination of those cases. Instead, it is a kind of hybrid or composite which can blend whichever features of each model are most useful for explaining paths of data.

- EWA captures learning situations in which subjects use history of plays by opponents, and full information about their own prospective payoffs, in adjusting their choice behavior. Some existing models (e.g., reinforcement) assume that subjects do not use such information. These simpler models cannot explain why learning is different when information is and is not available (e.g., Van Huyck, Battalio, Rankin, 1996).

The goal of our paper is to extend Camerer and Ho (1997) (CH) by modeling agent heterogeneity and time variation of parameters, and comparing the logit and power probability functions. The econometric estimation in CH adopted a representative agent approach which assumed that all agents are the same. Obviously this is unrealistic, but allowing each agent’s parameter values to differ adds too

many free parameters for the data we estimate (cf. Cheung & Friedman, 1997). An intermediate step assumes that agents are in one of two (or more) “latent” classes. Agents in each class have the same parameter values, but parameters differ across the two classes. We can then see whether allowing two classes fit the data substantially better than allowing only a single class. In addition, we can test whether the EWA model, allowing two different classes, fits better than a population mixture model in which one class are reinforcement learners and another class are belief-based learners.

Our earlier estimation (and all others) also assumed that learning parameters are constant over periods of the experiment. Parameters might vary over time if there is a kind of “learning about learning” or shift from one learning rule to another during a game.

Our earlier estimation used the logit (exponential) probability response function to transform attraction of strategies into choice probability. We compare logit and power probability response functions, since both have been used in the literature and are not often compared.

In the next section, we describe EWA formally and shows it contains the belief-based and choice reinforcement as special cases. We discuss the advantages and disadvantages of the power and logit probability response functions. The third section provides interpretations of the model parameters and discusses why studying heterogeneity and time variation in learning parameters could help to explain dynamic behaviors better. The fourth section reports parameter estimates from two coordination games. The last section concludes.

## 2. THE EXPERIENCE-WEIGHTED ATTRACTION LEARNING MODEL

Like the reinforcement and belief-based approaches, the experience-weighted attraction (EWA) model defines an intermediate construct which measures the attraction of strategies. The probability of choosing each strategy is an increasing function of the relative attraction of a strategy (in a precise way made clear below). In all our work we assume that the strategies which are reinforced are stage-game strategies. For many reasons it is sensible, in further work, to consider other kinds of strategies (including repeated-game strategies or decision rules; see Stahl, 1996). For example, Erev and Roth (in press) suggest that it is possible to model belief learning as reinforcement of a belief-based decision rule. Our framework can also be easily adapted to model learning over strategies which are more complicated than simple stage-game strategies.

We start with notation. We study  $n$ -person normal-form games. Players are indexed by  $i$  ( $i = 1, \dots, n$ ), and the strategy space of player  $i$ ,  $S_i$  consists of  $m_i$  discrete choices, that is,  $S_i = \{s_i^1, s_i^2, \dots, s_i^{m_i-1}, s_i^{m_i}\}$ .  $S = S_1 \times \dots \times S_n$  is the Cartesian product of the individual strategy spaces and is the strategy space of the game.  $s_i \in S_i$  denotes a strategy of player  $i$ , and is therefore an element of  $S_i$ .  $s = (s_1, \dots, s_n) \in S$  is a strategy combination, and it consists of  $n$  strategies, one for each player.  $s_{-i} = (s_1, \dots, s_{i-1}, s_{i+1}, \dots, s_n)$  is a strategy combination of all players except  $i$ .  $S_{-i}$  has a cardinality of  $m_{-i} = \prod_{j=1, j \neq i}^n m_j$ .  $\pi_i(s_i, s_{-i})$  is the payoff function of player  $i$  and is scalar valued. Denote the actual strategy chosen by player  $i$  in period

$t$  by  $s_i(t)$ , and the strategy (vector) chosen by all other players by  $s_{-i}(t)$ . Denote player  $i$ 's payoff in a period  $t$  by  $\pi_i(s_i(t), s_{-i}(t))$ .

Learning models require a specification of initial attractions, how attractions are updated by experience, and how choice probabilities depend on attractions.

### 2.1. The Updating Rules

The core of the EWA model is two variables which are updated after each round. The first variable is  $N(t)$ , which we interpret as the number of "observation-equivalents" of past experience. (This number will not generally equal the number of previous observations, due to depreciation, and other forces.) The second variable is  $A_i^j(t)$ , the attraction of a strategy *after* period  $t$  has taken place.

The variables  $N(t)$  and  $A_i^j(t)$  begin with some prior values,  $N(0)$  and  $A_i^j(0)$ . These prior values can be thought of as reflecting pregame experience, either due to learning transferred from different games or due to introspection. (Then  $N(0)$  can be interpreted as the number of periods of actual experience which is equivalent in attraction impact to the pregame thinking.)

Updating is governed by two rules. First,

$$N(t) = \rho \cdot N(t-1) + 1, \quad t \geq 1. \quad (2.1)$$

The parameter  $\rho$  can be thought of as a depreciation rate or retrospective discount factor that measures the fractional impact of previous experience, compared to one new period.

The second rule updates the level of attraction. A key component of the updating is the payoff that a strategy either yielded, or would have yielded, in a period. The model weights hypothetical payoffs that unchosen strategies would have earned by a parameter  $\delta$ , and weights payoff actually received, from chosen strategy  $s_i(t)$ , by an additional  $1 - \delta$  (so they receive a total weight of 1). Using an indicator function  $I(x, y)$  which equals 1 if  $x = y$  and 0 if  $x \neq y$ , the weighted payoff can be written  $[\delta + (1 - \delta) \cdot I(s_i^j, s_i(t))] \cdot \pi_i(s_i^j, s_{-i}(t))$ .

The rule for updating attraction sets  $A_i^j(t)$  to be a weighted average of the (weighted) payoff from period  $t$  and the previous attraction  $A_i^j(t-1)$ , according to:

$$A_i^j(t) = \frac{\phi \cdot N(t-1) \cdot A_i^j(t-1) + [\delta + (1 - \delta) \cdot I(s_i^j, s_i(t))] \cdot \pi_i(s_i^j, s_{-i}(t))}{N(t)}. \quad (2.2)$$

The factor  $\phi$  is a discount factor that depreciates previous attraction.

### 2.2. Choice Reinforcement

In choice reinforcement models, reinforcement levels are increased by payoffs of chosen strategies. (In some approaches, choice probabilities are affected directly but we ignore those approaches here.) Denote the initial reinforcement level of strategy  $j$  of player  $i$   $R_i^j(0)$ . Reinforcements are updated according to

$$R_i^j(t) = \phi \cdot R_i^j(t-1) + I(s_i^j, s_i(t)) \cdot \pi_i(s_i^j, s_{-i}(t)). \quad (2.3)$$

This basic model allows previous reinforcements to “depreciate” or “decay” by a factor  $\phi$  (similar to a retrospective discount factor), and updates chosen strategies according to their payoffs. It captures the property that strategies with positive payoffs increase in relative reinforcement. It is easy to see that the reinforcement updating formula is a special case of the EWA rule (2.2) when  $R_i^j(0) = A_i^j(0)$ ,  $\delta = 0$ ,  $N(0) = 1$ , and  $\rho = 0$ .

### 2.3. Belief-Based Models

Adaptive players are those who base their responses on beliefs formed by observing the history of others. While there are many ways of forming beliefs, we consider a fairly large class of models, which include familiar ones like fictitious play (Brown, 1951) and Cournot best-response (Cournot, 1960) as special cases.

We consider models in which prior beliefs of opponents’ strategy combinations are expressed as a ratio of hypothetical counts of observations of strategy combination  $s_{-i}^k$ , denoted by  $N_{-i}^k(0)$ , which sum to  $N(t) = \sum_{k=1}^m N_{-i}^k(t)$ . ( $N(t)$  is not subscripted because the count of frequencies is assumed, in our estimation, to be the same for all players.) These observations can then be naturally integrated with actual observations as experience accumulates. Furthermore, we assume that past experience is depreciated or discounted by a factor  $\rho$ . Note that Then, the initial prior  $B_{-i}^k(0) = N_{-i}^k(0)/N(0)$ . Beliefs are updated by depreciating the previous counts by  $\rho$ , and adding one for the strategy combination actually chosen by the other players. That is,

$$B_{-i}^k(t) = \frac{\rho \cdot N_{-i}^k(t-1) + I(s_{-i}^k, s_{-i}(t))}{\sum_{h=1}^m [\rho \cdot N_{-i}^h(t-1) + I(s_{-i}^h, s_{-i}(t))]} \quad (2.4)$$

Expressing beliefs  $B_{-i}^k(t)$  in terms of previous-period beliefs  $B_{-i}^k(t-1)$  gives

$$B_{-i}^k(t) = \frac{\rho \cdot N(t-1) \cdot B_{-i}^k(t-1) + I(s_{-i}^k, s_{-i}(t))}{\rho \cdot N(t-1) + 1} \quad (2.5)$$

Expected payoffs in period  $t$  are computed by  $E_i^j(t) = \sum_{k=1}^m \pi_i(s_i^j, s_{-i}^k) \cdot B_{-i}^k(t)$ . The key step is that expected payoffs in period  $t$  can be expressed as a function of period  $t-1$  expected payoffs, according to

$$E_i^j(t) = \frac{\rho \cdot N(t-1) \cdot E_i^j(t-1) + \pi(s_i^j, s_{-i}(t))}{\rho \cdot N(t-1) + 1} \quad (2.6)$$

Suppose initial attractions  $A_i^j(0)$  are equal to expected payoffs, given initial beliefs which arise from the “experience-equivalent” strategy counts  $N_{-i}^k(0)$ . Then comparing equation (2.6) with (2.2), it is evident that substituting  $\delta = 1$  and  $\rho = \phi$  into the attraction updating equation (2.2) gives EWA attractions which are equal to updated expected payoffs using weighted fictitious play.

Weighted fictitious play includes Cournot dynamics in which only the most recent choice by one's opponent matters ( $\phi = 0$ ) and fictitious play, in which players take an equally-weighted average of all previous observations ( $\phi = 1$ ).

#### 2.4. Choice Probabilities

So far we have said nothing about how the probability of player  $i$  choosing strategy  $j$  at time  $t$ ,  $P_i^j(t)$ , depends on the attraction  $A_i^j(t)$ . Obviously we would like  $P_i^j(t)$  to be monotonically increasing in  $A_i^j(t)$  and decreasing in  $A_i^k(t)$  (where  $k \neq j$ ).

In this paper, we use both logit and power probability response functions:

$$P_i^j(t+1) = \frac{e^{\lambda \cdot A_i^j(t)}}{\sum_{k=1}^{m_i} e^{\lambda \cdot A_i^k(t)}} \quad (2.7)$$

$$P_i^j(t+1) = \frac{(A_i^j(t))^\lambda}{\sum_{k=1}^{m_i} (A_i^k(t))^\lambda}. \quad (2.8)$$

The parameter  $\lambda$  measures sensitivity of players to attractions. Sensitivity could vary due to psychophysical concerns or whether subjects are highly motivated or not. In their work on quantal response equilibrium (QRE), McKelvey and Palfrey (1995, 1996) extend standard equilibrium notions to include the complication that players respond with "error" (and realize others respond with error too).

Each probability function has advantages and disadvantages. The logit form is popular for studying discrete choice, choice among consumer products, etc. (e.g., Anderson, de Palma, Thisse, 1992). The logit form has been used to study learning in games by Mookerjee and Sopher (1994, 1997), Ho and Weigelt (1996), and Fudenberg and Levine (in press). It is also used in "quantal response equilibrium" models by McKelvey and Palfrey (1995, 1996). The logit form is invariant to adding a constant to all attractions. (As a result, one must normalize  $A_i^j(0)$  for one value of  $j$  when doing estimation.) In the logit form, negative values of  $A_i^j(0)$  are permissible, which means one can avoid the difficult question of how to update attractions when payoffs are negative.

The power form has been used (among others) by Tang (1996) and the special case  $\lambda = 1$  has been used by Erev and Roth (1997). The power form (with any  $\lambda$ ) is invariant to multiplying all attractions by a constant. Because of this invariance, the parameters  $\rho$  and  $N(0)$  make no difference when the power form is used.<sup>1</sup> Thus, invariance to multiplication in the power form means that the distinction between the attraction levels  $A_i^j(0)$  and the "weight" on initial attractions,  $N(0)$ , is

<sup>1</sup> The reason is that EWA attractions at time  $t$  depend only a function of lagged payoffs, and the product  $A_i^j(0) \cdot N(0)$  (the scaling term  $\rho \cdot N(t-1) + 1$  cancels out and thus does not affect choice probabilities). While initial choice probabilities depend on  $A_i^j(0)$  only, these probabilities are the same as those that depend on  $A_i^j(0) \cdot N(0)$  (for  $N(0) > 0$ ). As a result, multiplying the initial attractions by an arbitrary constant makes no difference (econometrically,  $N(0)$  is not identifiable). Alternatively, one can fix one of the values of  $A_i^j(0)$  and let  $N(0)$  be determined by the data.

meaningless. In the power form, negative values of  $A_i^j(0)$  are not permitted. Ultimately, it is an empirical question whether the exponential or power forms fit better. In the estimation we below we fit both power and exponential forms and compare them.

### 3. INTERPRETING THE EWA MODEL

Before turning to the estimation results, it is instructive to ask how the EWA parameters can be interpreted. In our view, sensible interpretations of the parameters are important in judging how good a model is. Parameters with natural interpretations can be measured in other ways and theorized about (by a broader range of social scientists) more fruitfully.

#### 3.1. EWA Parameters

1. *Imagination*  $\delta$ : The parameter  $\delta$  can be interpreted as a combination of information about foregone outcomes, and the ability to imagine them (cf. Van Huyck, Battalio, and Rankin, 1996, esp. p. 16). When  $\delta = 1$ , the incremental weight to the actual payoff,  $1 - \delta$ , is zero; hypothetical or imaginary payoffs have equal cognitive weight. In EWA, foregone payoffs have weight  $\delta$ , which expresses the principle of simulated effect (if  $\delta$  is positive). But actual payoffs have an additional weight  $1 - \delta$ , which incorporates the principle underlying the law of effect.

A good way to understand EWA, and compare it to other approaches, is to ask how players change strategies if they do change them at all. The key function of  $\delta$  is to point players in the direction of better strategies. Holding aside depreciation of previous attractions, the tendency of a player to switch to strategies  $s_i^j$  will depend on  $\delta \cdot \pi(s_i^j, s_{-i}(t)) - \pi(s_i(t), s_{-i}(t))$ . This quasi-error can be written as  $\delta \cdot (\pi(s_i^j, s_{-i}(t)) - \pi(s_i(t), s_{-i}(t))) - (1 - \delta) \cdot \pi(s_i(t), s_{-i}(t))$ . This is a kind of error-correction model in which players switch to strategies if a fraction  $\delta$  of the ex post error ( $\pi(s_i^j, s_{-i}(t)) - \pi(s_i(t), s_{-i}(t))$ ) is greater than a fraction  $1 - \delta$  of the payoff to the currently-chosen strategy. Intuitively, players will switch to an unchosen strategy only if there is large enough added payoff to doing so (or, equivalently, a reduction in ex post error), above and beyond a fraction of the currently-received payoff.<sup>2</sup> This form also shows how responsiveness to ex post error depends directly on  $\delta$ — when  $\delta$  is small, players are more responsive to chosen-strategy payoff than to error; when  $\delta$  is large, they are more responsive to error.

Some versions of choice reinforcement allow strategy switches to be somewhat predictable, by “spilling over” reinforcement from a chosen strategy to its neighbors (e.g., Roth & Erev, 1995). This “generalization” predicts that players will switch to nearby strategies (or similar ones, e.g., Sarin & Vahid, 1997), but not in a way that

<sup>2</sup> This analysis makes clear two-strategy games will generally provide the weakest way to distinguish theories which make different predictions about how players switch strategies, since there is only strategy to switch toward.

depends on prospective payoffs of those strategies. In the coordination game data, local generalization fits much worse than the foregone payoff weighting by  $\delta$ .<sup>3</sup>

In the belief-based case  $\delta = 1$  so players simply switch to the strategies with the highest foregone payoffs. This is very similar to “learning direction” theory (Selten & Buchta, 1994), which says that when strategies are ordered, players will move in the direction of the ex post best response.

2. *Depreciation rates  $\phi$  and  $\rho$* : The parameter  $\phi$  can be naturally interpreted as depreciation of past attractions,  $A(t)$ . The parameter  $\rho$  depreciates the experience measure  $N(t)$ . It captures something like decay in the strength of prior beliefs, which is generally different than decay of early attraction (captured by  $\phi$ ). In a game-theoretic context,  $\rho$  and  $\phi$  might be related to the degree to which players realize other players are adapting, so that old observations on what others did become less and less useful. Then  $\rho$  and  $\phi$  are like indices of perceived non-stationarity.

One way to interpret  $\rho$  and  $\phi$  is by considering the numerator and denominator of the main EWA updating equation (2.2) separately, and thinking about how reinforcement and belief-based models use these two terms differently. The top term is  $\phi \cdot N(t-1) \cdot A_i^j(t-1) + [\delta + (1-\delta) \cdot I(s_i^j, s_i(t))] \cdot \pi_i(s_i^j, s_{-i}(t))$ . This term is like a running total of (depreciated) attraction, updated by each period’s payoffs. The bottom term is  $\rho \cdot N(t-1) + 1$ . This term is like a running total of (depreciated) periods of experience-equivalence. Reinforcement models essentially keep track of the running total in the numerator, and do not adjust for the number of periods of experience-equivalence (since  $\rho = 0$ , the denominator is always one). Belief-based models also keep track of the attraction total but divide the total number of periods of experience-equivalence. By depreciating the two totals at the same rate ( $\rho = \phi$ ), the belief-based models keeps the “per-period” attractions (expected payoffs) in a range bounded by the game’s payoffs.

An analogy might help illustrate our point. Instead of determining attractions of strategies, think about evaluating a person (for example, an athlete, or a senior colleague you might hire) based on a stream of lifetime performances. The reinforcement model evaluates people based on (depreciated) lifetime performance. The belief-based models evaluate people based on “average” (depreciated) performance. Both statistics are probably useful in evaluation—in hiring a colleague or an athlete, you would want to know lifetime performance *and* some kind of

<sup>3</sup> Generalization or experimentation which spills a portion  $\varepsilon/2$  of the reinforcement from a chosen strategy’s payoff over to its two neighboring strategies, when strategies are naturally ordered (or  $\varepsilon$  if there is only a single neighbor). To see whether generalization of this sort fits better than the EWA specification of weighting all foregone payoffs by  $\delta$ , we compared the two specifications, including all other EWA parameters, allowing time-variation, and using the logit form. The log likelihoods for median action games, for  $\delta$  vs generalization, are: 320.1 vs 364.7 (two segments) and 337.3 and 366.0 (one segment). For weak-link games the results are 775.2 vs 785.8 and 806.6 vs 819.2. The generalization specification has the same number of free parameters and always fits much worse than the  $\delta$  specification used in EWA. The estimates of  $\varepsilon$  are .63 and .64 (two segments) and .61 (one segment) for median-action games, and .99 and .97 (two segments) and .88 (one segment) for weak-link games. These figures are higher than the value of .05 used by Roth and Erev (1995).

performance averaged across experience. One way to mix the two is to normalize depreciated cumulative performance by depreciated experience, but depreciate the amount of experience more rapidly. For example, if two people perform equally well on average, a person with 10 years of experience is rated somewhere between equally as good and twice as good as the person with five years of experience. When  $\phi > \rho$ , EWA models players who use something in between “lifetime” performance and “average” performance to evaluate strategies.

3. *Initial attraction and weight  $A_i^j(0), N(0)$* : The term  $A_i^j(0)$  represents the initial attraction, which might be derived from some analysis of the game (e.g., equilibrium analysis, selection principles, or decision rules like insufficient reason and maximin). Indeed, any theory of first-period play can be tested empirically as a restriction on initial attractions. Attractions also could be influenced by similarity between current strategies and strategies which were successful in similar games, etc.

The term  $N(0)$  represents the strength of initial attractions, and can be interpreted as the unit-for-unit relative weight of prior “experience” (or introspection) compared to actual payoff experience.

### 3.2. Time Variation

Players may modify the ways they learn over time. Since the EWA model contains different learning models as special cases, by allowing the learning parameters to vary, we can also capture smooth rule changes. A more direct approach is to assume that players consider a large class of rules and shift weight to those which perform well (as in Tang, 1996, “method learning” and Stahl, 1996). The rule-learning approach has the advantage of endogeneizing the way in which parameters change in response to experience, which we do not do. Our approach is simply a first look at whether allowing parameters to vary is empirically useful. If not, this suggests that a fixed-parameter specification is a useful approximation. If so, we can explore more detailed specifications of why parameters change, as Stahl does.

There are several plausible reasons why parameters may vary over time. The value of  $\delta$  could increase over time if players learn to simulate foregone payoffs better, or gradually pay more attention to foregone payoffs (as if they gradually switch from reinforcement to belief learning). Alternatively, one can interpret the rise in  $\delta$  as players learning to penalize ex post mistakes more heavily. Suppose the decay parameters  $\phi$  and  $\rho$  capture the players’ perceptions of the degree of nonstationarity in the environment—for example, these decay rates lower when players realize that other players are changing more rapidly. Then as convergence occurs and environments become more stationary,  $\phi$  and  $\rho$  will increase. If players get more sensitive to differences in attractions over time, then  $\lambda$  will rise across periods.

With these psychological motivations for time-variation in mind, we allow the four learning parameters to vary according to a compact exponential form. Denoting each parameter by  $X$ , we estimate the forms  $X_t = X_0 \cdot e^{g_X \cdot t}$  (with  $X \in \{\delta, \phi, \rho, \lambda\}$ ), where  $g_X$  is the rate of parameter change.

### 3.3. *Heterogeneity*

For analytical tractability, standard economic analysis often adopts a “representative agent approach” in which all economic agents are treated identically. Of course, people are different. For example, prior research on games and decisions has shown that players are identifiably different in their levels of sophistication or decision rules (Holt, 1991; Stahl, 1993; Ho, Camerer, & Weigelt, in press).

We think the right approach to allowing heterogeneity is to allow various “latent classes” or segments of players. An extreme form of this approach is to allow each individual to have separate parameter values, and estimate each person’s parameters from their data. For the games we study we simply do not have enough data per person to do this. Cheung and Friedman (1997) have done so for weighted fictitious play. They reject the hypothesis that all players are the same. However, their results suggest that reliably estimating individual differences is not easy.<sup>4</sup> Notice that the individual-specific parameter approach is only the best approach if players are not merely different, but also unique (i.e., each player is different from all others).

Since players are probably not all the same, and not all unique, the best approach is probably an intermediate one in which players are assumed fall into one of several discretely-distributed segments or subpopulations (e.g., Crawford, 1995). Each member of a segment has the same parameter values, but each segment’s values differ from the others.

This segmentation approach has three advantages. First, it is more parsimonious than the first approach and easy to implement computationally. Second, the idea that people fall into a discrete number of segments is appealing and widely used in various applications. For example, in personality psychology, people are often classified into a small number of classes based on bundles of correlated traits, or classified by cognitive “styles”. Clinical psychologists describe “syndromes” characterized by appearance of a cluster of symptoms. Research in marketing has shown that economic agents often differ in their preferences and in the ways they select products (Kamakura & Russell, 1987), and can be grouped into market segments. Third, it is natural to ask whether EWA fits substantially better than a population mixture of reinforcement learners and belief-based learners. A two-segment analysis can answer this question (as we do below), by comparing the fit of an EWA model with two segments to a model in which one of those segments has reinforcement parameters and the other segment has belief parameters.

## 4. PARAMETER ESTIMATION FROM EXPERIMENTAL DATA

We estimated the values of model parameters from two coordination games called “weak-link” and median-action games. We picked coordination games because the results clearly highlight the differences between choice reinforcement

<sup>4</sup> For example, one-sixth of their subjects have values corresponding to  $\phi$  in EWA (their parameter  $\lambda$ ) greater than one, and a third have values which are negative. The fact that half the subjects have estimated values which are implausibly high or low suggests that estimating individual differences reliably is difficult.

( $\delta = 0$ ), the belief models ( $\delta = 1$ ), and the more general EWA approach. Furthermore, players will be heterogeneous if they use different selection principles or begin with high attractions on different equilibria.

#### 4.1. The Likelihood Function

In both games subjects have seven strategies, numbers 1–7. Thus, the log of the likelihood function for the one-segment EWA model without time-variation is

$$LL(A^1(0), \dots, A^7(0), \delta, \phi, \lambda, N(0), \rho) = \sum_{i=1}^n \sum_{t=1}^T \log(P_i^{S_i(t)}(t)). \quad (4.1)$$

The probabilities  $P_i^{S_i(t)}(t)$  are given by equation (2.7) and (2.8).<sup>5</sup> Initial attractions are estimated and subsequent attractions are updated according to equation (2.2). The initial experience-weight  $N(0)$  is estimated, and subsequent values  $N(t)$  for  $t > 1$  are updated according to equation (2.1).

In the logit probability response function, one of the initial attractions must be fixed for identifiability—we set  $A^5 = 1$ —because the probabilities  $P_i^j(t)$  are invariant to adding a constant to all attractions. Similarly, in the power probability response function  $N(0)$  must be fixed (and  $\rho$  disappears) because the probabilities are invariant to multiplying initial attractions by a constant. This baseline model has 11 free parameters for the logit form and 10 for the power form (which drops  $\rho$ ). The two-segment models double the number of free parameters plus adds one to measure the fraction of players in each segment.

An alternative method which has occasionally been used to evaluate model fit is to simulate choice paths, given some model parameters, and compare averaged simulated paths with the data (McAllister, 1992; Erev & Roth, 1997). The accuracy of the model can be measured by squaring the deviation between the actual and simulated choice proportions in each period, and averaging those squared deviations across all the choices and periods, giving a mean squared deviation (MSD). There is no a priori reason to expect that model parameters which are chosen to minimize MSD in this way will systematically deviate from maximum-likelihood estimates.<sup>6</sup>

For the purpose of comparing the evaluation of models using MSD of simulated paths with MLE evaluation, we simulated paths for the five models described below which used the exponential probability form. Each of 1000 simulations took the MLE parameter estimates as the basis for simulation. Each period, MLE parameters imply a predicted choice probability, which are used to randomly simulate behavior of each of several artificial players in a group. The behavior of a group of players determines an order statistic (minimum or median) which determines each simulated player's payoff. These simulated payoffs are used to update their attractions, and hence determine choice probabilities, iteratively. Then total

<sup>5</sup> An implicit assumption in the exponential (logit) choice model is that the disturbances which are added to attractions have a double-exponential distribution (Yellott, 1977).

<sup>6</sup> However, optimizing parameter fit by comparing averaged simulated paths to data, and choosing the parameter vector with the lowest MSD, is a computationally inefficient method compared to gradient algorithms used in MLE packages.

choice proportions are averaged across the 1000 simulations. For the two-segment results, each player's segment membership was determined randomly and fixed throughout 1000 simulations. To control for sampling variation across determination of the segments, the random drawing of segment membership was simulated 20 times, for a total of 20,000 simulations.

#### 4.2. Weak-Link Coordination Games

Weak-link games are  $n$ -person versions of stag hunt. In the weak-link games we study, players choose a strategy from some ordered set, and their payoff depends positively on the lowest strategy picked by any player, and negatively on the difference between their strategy and the lowest one. This game was first studied experimentally by Van Huyck, Battalio and Beil (1990). We use the data collected by Knez and Camerer (1996) and Camerer, Knez, and Weber (1996) on this game. The weak-link game captures social situations in which a group's output is extremely sensitive to its "weakest link"—friends must wait for the slowest arrival before they all get seated in a restaurant, a chemical recipe or meal is ruined by one bad ingredient, a dyke bursts if it has a single leak, etc.

Figure 1 shows payoffs in the weak-link games. Players pick a number from 1 to 7. Player  $i$ 's payoff from choosing  $x_i$  (in dollars) is  $0.60 + 0.10 \cdot \min(x_1, x_2, \dots, x_n) - 0.10 \cdot (x_i - \min(x_1, x_2, \dots, x_n))$ .

		min { $X_i$ }						
		7	6	5	4	3	2	1
$X_i$	7	1.30	1.10	0.90	0.70	0.50	0.30	0.10
	6	-	1.20	1.00	0.80	0.60	0.40	0.20
	5	-	-	1.10	0.90	0.70	0.50	0.30
	4	-	-	-	1.00	0.80	0.60	0.40
	3	-	-	-	-	0.90	0.70	0.50
	2	-	-	-	-	-	0.80	0.60
	1	-	-	-	-	-	-	0.70

FIGURE 1

We pool data from three subject pools—UCLA, University of Chicago, and Caltech undergraduates—playing in groups of three. We assume that players care only about the minimum of others' numbers (since only that statistic, and their own choice, are relevant for their payoffs). Put differently, they are assumed to treat the other two players as a composite whose minimum is the composite's strategy choice. In the Chicago experiments subjects were told only the minimum choice in the entire group (including their own). In the Caltech and UCLA experiments subjects were told the choices of both other players in the group (so they could compute the minimum of others' choices exactly).

The data have a wide dispersion in first-period choices, with a large percentage of choices of 7 (the payoff-dominant equilibrium). Over time, there is some trend away from larger numbers 4–5 and toward smaller numbers (particularly 1–2), so there is evidence of learning which models should be able to capture.

Also, players frequently switched their strategies to strategies they had never picked before, which were best-responses to the observed minima. For example, consider a player who chooses 7 and observes a minimum of 5. The player earns a payoff of \$.90 but would have earned \$1.10 if she had chosen 5. We frequently see players' best-responding to the previous minima, switching from 7 to 5. These switches are hard to explain unless unchosen strategies are reinforced according to foregone payoffs and  $\delta$  is substantially different from zero.

#### 4.3. Median-Action Games

In another order-statistic coordination game that is closely related to the weak-link game, the group payoff depends the median of all players' actions instead of the minimum. Players earn a payoff which increases in the median, and decreases in the (squared) deviation from the median. These median-action games were first studied experimentally by Van Huyck, Battalio, and Beil (VHBB, 1991).

The median-action games capture social situations in which conformity pressures induce people to behave like others do, but everyone prefers the group to choose a high median. Figure 2 shows the game matrix (corresponding to the  $\Gamma$  treatment in VHBB).

We estimate EWA, choice reinforcement, and belief-based models using sessions 1–6 from VHBB. In their experiments groups of nine subjects each play ten periods together. We pool together treatments using nine-person groups and “dual market” (dm) treatments in which players play with a nine-person group and a twenty-seven person group simultaneously. In each round players choose an integer from 1 to 7, inclusive. At the end of each round the median is announced and players compute their payoffs. Since the groups are large, we assume that players form beliefs over the median of all players, ignoring their own influence and treating the group as a composite single player.

The results show initial choices are concentrated around 4–5, with a small spike at 7. Later choices move sharply toward the initial medians, which were always 4 or 5. There is strong path-dependence: The 10th-round median in *every* session was equal to the first-round median.

		Median $\{X_i\}$						
		7	6	5	4	3	2	1
$X_i$	7	1.30	1.15	0.90	0.55	0.10	-0.45	-1.10
	6	1.25	1.20	1.05	0.80	0.45	0.00	-0.55
	5	1.10	1.15	1.10	0.95	0.70	0.35	-0.10
	4	0.85	1.00	1.05	1.00	0.85	0.60	0.25
	3	0.50	0.75	0.90	0.95	0.90	0.75	0.50
	2	0.05	0.40	0.65	0.80	0.85	0.80	0.65
	1	-0.50	-0.05	0.30	0.55	0.70	0.75	0.70

FIGURE 2

#### 4.4. Estimation Results

Tables 1–2 report the results of estimates from weak-link and median-action games. The first 8 rows report log likelihoods for every possible combination of power vs exponential forms, one- and two-segments, and no time-variation vs time-variation. The last three rows report various two-segment models in which the segments are two segments of choice reinforcement, two segments of beliefs, or a mixture of one segment of reinforcement and one segment of beliefs.

The findings are organized as a series of answers to questions. Tables 3–5 summarize comparisons of the specifications in Tables 1–2 which provide the answers.

##### 1. Which fits better: Power or exponential?

The answer is that exponential fits better in general. The power form has one fewer free parameter because probabilities in that form only depend on ratios of attractions; then the denominator of the updating equations disappears and the experience decay rate  $\rho$ , which only appears in the denominator, does not matter. However, the power form and the logit form are not nested so a simple  $\chi^2$  statistic cannot be used to compare them.

TABLE 1

Maximum Likelihood for Various Models, Weak-Link Games ( $N=645$ )

Model number	Number of segment	Probability law	Time varying	Number of parameters	-LL
[1]	1	Power	No	10	821.2
[2]	1	Power	Yes	13	816.1
[3]	1	Logit	No	11	814.0
[4]	1	Logit	Yes	15	806.6
[5]	2	Power	No	21	819.2
[6]	2	Power	Yes	27	786.1
[7]	2	Logit	No	23	794.7
[8]	2	Logit	Yes	31	775.2
[9]					
CR + BB	2	Logit	Yes	23	796.7
[10]					
CR + CR	2	Logit	Yes	23	821.0
[11]					
BB + BB	2	Logit	Yes	23	894.8

Therefore, we compare the logit and power forms using “information criteria” which adjust goodness-of-fit for varying degrees of freedom, subtracting a “penalty” from log likelihood for each degree of freedom used. There a variety of criteria, but we use two well-known ones—the Akaike criterion, with a penalty of 1, and the Bayesian information criterion with a penalty of  $\ln(N)$  where  $N$  is the sample size. Of non-Bayesian criteria, the Akaike criterion imposes the largest penalty (others propose penalties of 0.75, 0.50, and 0.345; see Harless & Camerer, 1994). Thus,

TABLE 2

Maximum Likelihood for Various Models, Median Effort Games ( $N=540$ )

Model number	Number of segment	Probability law	Time varying	Number of parameters	-LL
[1]	1	Power	No	10	347.9
[2]	1	Power	Yes	13	347.4
[3]	1	Logit	No	11	344.0
[4]	1	Logit	Yes	15	337.3
[5]	2	Power	No	21	340.0
[6]	2	Power	Yes	27	327.9
[7]	2	Logit	No	23	325.2
[8]	2	Logit	Yes	31	320.1
[9]					
CR + BB	2	Logit	Yes	23	436.5
[10]					
CR + CR	2	Logit	Yes	23	388.3
[11]					
BB + BB	2	Logit	Yes	23	445.8

**TABLE 3**  
**Empirical Tests of Logit vs Power Probability Response Functions**

Game	Model description	Model comparison	Akaike criterion	Bayesian criterion
Weak-link				
	1 segment without time variation	[3] vs [1]	6.2	0.7
	1 segment with time variation	[4] vs [2]	7.5	-3.4
	2 segment without time variation	[7] vs [5]	22.5	11.6
	2 segment with time variation	[8] vs [6]	6.9	12.8
Median-action				
	1 segment without time variation	[3] vs [1]	2.9	-2.4
	1 segment with time variation	[4] vs [2]	8.1	-2.5
	2 segment without time variation	[7] vs [5]	12.8	2.2
	2 segment with time variation	[8] vs [6]	3.8	-17.3

compared to other criteria these information criteria favor simpler models (in this case, the power form).

The difference in log likelihoods of the logit and power forms, adjusted by each of the two criteria, are shown in Table 3. Positive numbers favor the logit form. The Akaike criterion favors the logit form in every comparison. The Bayesian criterion, which applies a bigger penalty to the logit form, is sometimes better for logit and sometimes better for power. Because the logit form wins overwhelmingly by the lower-penalty criterion, and the two forms are about equally good by the higher-penalty criterion, we conclude that in general the logit form is better. Put differently, the only circumstance under which logit is not better is when the Bayesian sample-size-dependent penalty is used, and even in that case power and logit are about equal. This is a stronger result than Tang (1996) and Chen and Tang (1996), who found the two forms fit about equally well.

While these results favor the logit form, the power and logit forms might be useful for different purposes. Logit enables one to avoid having to tackle the problem of adjusting for negative attractions. The power form saves degrees of freedom. If one wants to distinguish the reinforcement and belief cases from EWA (and other special cases), then the extra distinguishability afforded by  $N(0)$  and  $\rho$  is valuable; if one is not interested in model comparison then suppressing those factors eliminates a distraction.

## 2. Does Heterogeneity Exist?

The answer is Yes. Table 4 shows that in six of eight comparisons, the two-segment model fits significantly better (at  $p < 0.01$ ) than the one-segment model. The simulation results for the logit form generally corroborate this conclusion: The MSDs in weak-link games for one- and two-segment models are 0.0042 and 0.0085 (without time variation) and 0.0043 and 0.0026 (with time variation); the corresponding results for median-action games are 0.0049 and 0.0019, and 0.0141 and 0.0032. The two-segment MSDs are about half as large as those for one-segment models, except the anomalous case of weak-link games with time variation.

**TABLE 4**  
**Empirical Tests of Population Heterogeneity**

Game	Model description	Model Comparison	Chi-square statistics <sub>(df)</sub> ( <i>p</i> -value)	1 Segment		2 Segment	
				$\delta$ ( $g_a$ )	$\delta_1$ ( $g_{a1}$ )	$\delta_2$ ( $g_{a2}$ )	Segment 1 %
Weak-link	Power prob without time variation	[5] vs [1]	4.0 <sub>(11)</sub> (0.970)	0.607	0.573	0.689	53.2
	Power with time variation	[6] vs [2]	60.0 <sub>(14)</sub> (0.000)	0.459 (0.104)	0.884 (0.013)	0.000 (-3.125)	64.7
	Logit without time variation	[7] vs [3]	38.6 <sub>(12)</sub> (0.000)	0.660	0.625	0.698	50.9
	Logit with time variation	[8] vs [4]	62.8 <sub>(16)</sub> (0.000)	0.679 (0.011)	0.812 (-0.067)	0.650 (-0.003)	52.2
Median-action	Power prob without time variation	[5] vs [1]	15.8 <sub>(11)</sub> (0.149)	0.845	0.452	0.862	11.4
	Power with time variation	[6] vs [2]	39.0 <sub>(14)</sub> (0.000)	0.823 (0.005)	0.980 (-0.003)	0.499 (0.072)	44.9
	Logit without time variation	[7] vs [3]	37.6 <sub>(12)</sub> (0.000)	0.900	0.965	0.638	54.4
	Logit with time variation	[8] vs [4]	34.4 <sub>(16)</sub> (0.002)	0.803 (0.030)	0.970 (-0.002)	0.453 (0.064)	55.4

**TABLE 5**  
**Empirical Tests of Time Variation**

Game	Model description	Model Comparison	Chi-square statistics <sub>(df)</sub> ( <i>p</i> -value)	$\phi$ ( $g_\phi$ )	$\rho$ ( $g_\rho$ )	$\lambda$ ( $g_\lambda$ )
Weak-link	1 segment with power probability	[2] vs [1]	10.2 <sub>(3)</sub> (0.017)	2.348 (-0.497)	-	9.222 (0.032)
	1 segment with logit probability	[4] vs [3]	14.8 <sub>(4)</sub> (0.005)	2.000 (-0.772)	1.000 (-0.593)	7.798 (0.016)
	2 segment with power probability	[6] vs [5]	66.2 <sub>(6)</sub> (0.000)	1.860/2.000 (-0.508/-1.092)	-	19.982/19.993 (0.161/-0.834)
	2 segment with logit probability	[8] vs [7]	39.0 <sub>(8)</sub> (0.000)	1.622/2.000 (-0.479/-1.480)	1.000/0.407 (-0.298/0.118)	3.343/19.980 (0.543/-0.115)
Median-action	1 segment with power probability	[2] vs [1]	1.0 <sub>(3)</sub> (0.801)	0.416 (0.041)	-	12.310 (0.437)
	1 segment with logit probability	[4] vs [3]	13.4 <sub>(4)</sub> (0.009)	0.457 (0.032)	0.329 (-0.380)	5.606 (0.196)
	2 segment with power probability	[6] vs [5]	24.2 <sub>(6)</sub> (0.000)	0.497/0.625 (-0.015/-0.029)	-	13.450/10.223 (0.663/0.047)
	2 segment with logit probability	[8] vs [7]	10.2 <sub>(8)</sub> (0.251)	1.060/0.446 (0.005/0.053)	0.004/0.509 (0.000/-0.010)	14.830/10.788 (-0.010/-0.029)

Table 4 also shows the estimated values of the foregone-payoff-weight parameter  $\delta$ . The observed heterogeneity is substantial in size and often easily interpretable. Generally the two estimates of  $\delta$  in the two-segment model are substantially different (in six of eight cases the difference is larger than 0.15). In every case, the one-segment  $\delta$  estimate lies strictly between the pair of two-segment estimates. The estimated proportions of players in the two segments are usually close to 50%.

Estimates from the median-action game tell the most interesting story. One segment estimate  $\hat{\delta}$  is around 0.9 and the other is around 0.5. When time-variation is allowed in the two-segment model, the higher  $\delta$  is close to one (0.98 and 0.97) and constant over time, while the other segment  $\hat{\delta}$  is around 0.50 and roughly doubles over the ten periods, to one. These segments can therefore be interpreted as a segment of players who weight foregone payoffs as highly as actual payoffs throughout (like belief learners), and a segment of players who begin weighting foregone payoffs half as much as actual payoffs, but “learn” to weight them equally, as if switching from an EWA hybrid rule to a belief-based rule.

The two segment  $\delta$  estimates in the weak-link game are closer together. In one striking case, power probability with time-variation, the two segments correspond to a segment of reinforcement learners ( $\hat{\delta}=0.000$ ) and a segment of belief types ( $\hat{\delta}=0.884$ ).

### 3. *Is EWA Behaviorally Equivalent to a Mixture of Reinforcement and Belief Models?*

The answer is No. Log likelihoods for the model in which the two segments correspond to reinforcement and belief learning model are shown in row 9 of Tables 1–2. The  $\chi^2$  statistics comparing this restriction with the most general two-segment EWA model (using the logit form and time-variation) are  $\chi^2(8) = 232.8$  and  $\chi^2(8) = 43.0$  for median and weak link games, respectively. Both statistics are highly significant, indicating that EWA is a large improvement over a mixture of reinforcement and belief segments. Given the results above, this is no surprise because the two-segment estimates of  $\delta$  do not generally break neatly into one low value near zero and another value near one. And inspection of the updating equations makes it apparent that the EWA attractions are not a linear combination of reinforcements and expected payoffs. Simulation results corroborate this conclusion: The EWA and combination model MSDs are 0.0026 and 0.0205 (weak-link) and 0.0032 and 0.0180 (median-action). EWA has an MSD which is lower by a factor of between five and ten.

### 4. *Does Parameter Time-Variation Matter?*

The answer is Slightly. In Table 5, in six out of eight comparisons the  $\chi^2$  statistic testing the restriction that the time-variation parameters are zero can be rejected at  $p < 0.01$ , but the  $\chi^2$  statistics are much smaller than those from tests of heterogeneity. The simulation results, however, suggest that evidence for time variation is very weak because including it often raises the MSD. The MSDs in weak-link games with and without time variation are 0.0043 and 0.0042 (one segment) and 0.0026 and 0.0085 (two segment); the corresponding statistics for median-action games are 0.0141 and 0.0050, and 0.0032 and 0.0019. In three of four comparisons, including time-variation actually raises MSD. It may be that including time variation

is “overfitting” in the MLE procedure, which is revealed by then averaging simulated paths and comparing to the data.

There is not much interesting regularity in the nature of time-variation. The value of  $\rho$  tends to decline over time, as if players become more and more myopic in looking back at payoff history (perhaps because convergence to equilibrium means that looking at recent history is sufficient). The estimated payoff sensitivity  $\lambda$  always rises over time in the one-segment cases (the exponent coefficients range from 0.016 to 0.437). This can be interpreted as evidence that subjects learn to respond more sensitively to differences in attractions.

## 5. CONCLUSION

In earlier research we proposed a new model of learning in games. In the EWA model, strategies have attraction levels which determine their probability of being chosen. Attractions are updated by weighting lagged attractions by the amount of “experience-equivalence” they have and reinforcing a strategy’s attraction by the payoffs actually received, or some fraction  $\delta$  of the payoff that would have been received (given the other players’ moves). This EWA model includes two prominent classes of models, choice reinforcement and belief-based models, as special cases. It shows that belief and reinforcement learning have a common, surprising kinship—belief learning is exactly the same as generalized reinforcement learning in which all strategies are reinforced equally, and lagged attractions are experience-weighted and normalized.

Previous work established that EWA improves on reinforcement and belief learning, empirically, by combining their best features: The reinforcement approach allows flexible initial attractions (which are not constrained to arise from prior beliefs) and the belief approach pushes choices in the direction of ex post best responses (which reinforcement does not do). It is important to note that EWA does not average the two approaches, it hybridizes them or forms an optimal combination of features.

In this paper we compared probability response functions and tested for player heterogeneity and time-variation in parameter values.

First we compared probability functions which take attractions raised to a power and normalize, with a logit form that exponentiates attractions and normalizes them. The logit form uses more free parameters, and fits better than the power form except when a large penalty is applied for the extra degrees of freedom, when the two fit equally well.

Second, we test whether heterogeneity among players improves fit by allowing players to come from one of two parametric segments. Allowing two segments (rather than only one, in earlier work) does improve fit substantially. The analysis also shows that EWA is *not* behaviorally equivalent to simply taking a weighted average of choice reinforcement and belief models.

Finally, parameters were allowed to vary across periods of the experiment. This improves goodness-of-fit modestly by standard  $\chi^2$  tests and worse by simulating paths based on parameter estimates. We conclude that fixing parameters across an experiment is a reasonable approximation.

## REFERENCES

- Arthur, B. (1991). Designing economic agents that act like human agents: A behavioral approach to bounded rationality. *AER Proceedings*, **81**, 353–359.
- Borgers, T., & Sarin, R. (1996). *Naive reinforcement learning with endogenous aspirations*. Texas A&M University working paper.
- Brown, G. (1951). Iterative solution of games by fictitious play. In T. Koopmans (Ed.), *Activity analysis of production and allocation*. New York: Wiley.
- Bush, R., & Mosteller, F. (1955). *Stochastic models for learning*. New York: Wiley.
- Camerer, C. F., & Ho, T.-H. (1977). *Experience-weighted attraction learning in normal-form games*. Caltech and UCLA working paper.
- Camerer, C., Knez, M., & Weber, R. (1966). *Timing and virtual observability in ultimatum bargaining and weak-link coordination games*. Caltech working paper no. 970.
- Cheung, Y.-W., & Friedman, D. (1977). Individual learning in normal form games: Some laboratory results. *Games and Economic Behavior*, **19**, 46–76.
- Cournot, A. (1960). *Recherches sur les principes mathématiques de la théorie des richesses*. (N. Bacon, Trans.). [Researches in the mathematical principles of the theory of wealth]. London: Haffner.
- Crawford, V. P. (1995). Adaptive dynamics in coordination games. *Econometrica*, **63**, 103–143.
- Cross, J. G. (1983). *A theory of adaptive economic behavior*. London: Cambridge Univ. Press.
- Erev, I., & Roth, A. (1997). *Modeling how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria*. University of Pittsburgh working paper.
- Erev, I., & Roth, A. E. (in press). On the role of reinforcement learning in experimental games: The cognitive game-theoretic approach. In D. Budescu, I. Erev, and R. Zwick (Eds.), *Games and human behavior: Essays in honor of Amnon Rapoport*. Dordrecht/Norwell, MA: Kluwer Academic.
- Fudenberg, D., & Levine, D. (in press). *Theory of learning in games*. Cambridge, MA: MIT Press.
- Harley, C. B. (1981). Learning the evolutionarily stable strategy. *Journal of Theoretical Biology*, **89**, 611–633.
- Harless, D., & Camerer, C. F. (1994). The predictive utility of generalized utility theories. *Econometrica*, **62**, 1251–1289.
- Herrnstein, J. R. (1970). On the law of effect. *Journal of the Experimental Analysis of Behavior*, **13**, 243–266.
- Ho, T.-H., & Weigelt, K. (1996). Task complexity, equilibrium selection, and learning: An experimental study. *Management Science*, **42**, 659–679.
- Ho, T.-H., Camerer, C., & Weigelt, K. (in press). Iterated dominance and iterated best-response in experimental “*p*-beauty contests.” *American Economic Review*.
- Holt, D. (1991). *An empirical model of strategic choice with an application to coordination games*. Queen’s University working paper.
- Kamakura, W., & Russell, G. (1989). A probabilistic choice model for market segmentation and elasticity structure. *Journal of Marketing Research*, **26**, 379–90.
- Knez, M., & Camerer, C. (1996). *Increasing cooperation in social dilemmas through the precedent of efficiency in coordination games*. University of Chicago working paper.
- McAllister, P. H. (1991). Adaptive approaches to stochastic programming. *Annals of Operations Research*, **30**, 45–62.
- McKelvey, R. D., & Palfrey, T. R. (1995). Quantal response equilibria for normal form games. *Games and Economic Behavior*, **10**, 6–38.
- McKelvey, R. D., & Palfrey, T. R. (1996). *Quantal response equilibria for extensive form games*. Caltech working paper.
- Mookerjee, D., & Sopher, B. (1994). Learning behavior in an experimental matching pennies game. *Games and Economic Behavior*, **7**, 62–91.

- Mookerjee, D., & Sopher, B. (1997). Learning and decision costs in experimental constant-sum games. *Games and Economic Behavior*, **19**, 97–132.
- Roth, A., & Erev, I. (1995). Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior*, **8**, 164–212.
- Sarin, R., & Vahid, F. (1997). *Payoff Assessments without probabilities: Incorporating similarity among strategies*. Texas A&M University working paper.
- Stahl, D. (1993). The evolution of  $smart_n$  players. *Games and Economic Behavior*, **5**, 604–617.
- Stahl, D. (1997). Bounded rational rule learning in a guessing game. *Games and Economic Behavior*, **5**, 604–617.
- Tang, F.-F. (1996). *Anticipatory learning in two-person games: An experimental study*. University of Bonn working paper.
- Thorndike, E. L. (1911). *Animal intelligence*. New York: Macmillan.
- Van Huyck, J., Battalio, R., & Beil, R. (1990). Tacit cooperation games, strategic uncertainty, and coordination failure. *American Economic Review*.
- Van Huyck, J., Battalio, R., & Rankin, F. (1996). *Selection dynamics and adaptive behavior without much information*. Texas A&M working paper.
- Yellott, J. (1997). The relationship between Luce's choice axioms, Thurstone's theory of comparative judgment, and the double exponential distribution. *Journal of Mathematical Psychology*, **5**, 109–144.

Received: February 12, 1998