DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA 91125

EXPERIENCE-WEIGHTED ATTRACTION LEARNING IN NORMAL FORM GAMES

Colin Camerer

Teck-Hua Ho
The Wharton School
University of Pennsylvania

SOCIAL SCIENCE WORKING PAPER 1003

March 1997
Revised December 1997

Experience-weighted Attraction Learning in Normal Form Games

Colin Camerer Teck-Hua Ho

Abstract

We describe a general model, 'experience-weighted attraction' (EWA) learning, which includes reinforcement learning and a class of weighted fictitious play belief models as special cases. In EWA, strategies have attractions which reflect prior predispositions, are updated based on payoff experience, and determine choice probabilities according to some rule (e.g., logit). A key feature is a parameter δ which weights the strength of hypothetical reinforcement of strategies which were *not* chosen according to the payoff they would have yielded. When $\delta = 0$ choice reinforcement results. When $\delta = 1$, levels of reinforcement of strategies are proportional to expected payoffs given beliefs based on past history. Another key feature is the growth rates of attractions. The EWA model controls the growth rates by two decay parameters, ϕ and ρ, which depreciate attractions and amount of experience separately. When $\phi = \rho$, belief-based models result; when $\rho = 0$ choice reinforcement results.

Using three data sets, parameter estimates of the model were calibrated on part of the data and used to predict the rest. Estimates of δ are generally around .50, ϕ around 1, and ρ varies from 0 to ϕ. Choice reinforcement models often outperform belief-based models in the calibration phase and underperform in out-of-sample validation. Both special cases are generally rejected in favor of EWA, though sometimes belief models do better. EWA is able to combine the best features of both approaches, allowing attractions to begin and grow flexibly as choice reinforcement does, but reinforcing unchosen strategies substantially as belief-based models implicitly do.

Keywords: Learning, behavioral game theory, reinforcement learning, fictitious play.

Experience-weighted Attraction Learning in Normal Form Games*

Colin Camerer Teck-Hua Ho

1 Introduction

How does an equilibrium arise in a noncooperative game? While it is conceivable that players reason their way to an equilibrium, a more psychologically plausible view is that players adapt or evolve toward it.[1] The flurry of recent research on adaptation and evolution mostly explores theoretical questions, like which types of equilibria specific evolutionary or adaptive rules converge to. We are interested in a fundamentally empirical question: Which models describe human behavior best? In this paper we propose a general 'experience-weighted attraction' (EWA) model and estimate the model parametrically, using three sets of experimental data.

The EWA model combines elements of two seemingly different approaches, and includes them as special cases. One approach, belief-based models, start 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 tend to choose a best-response, a strategy which maximizes their expected payoffs given the beliefs they formed.

*This research was supported by NSF grants SBR-9511001, 9511137, and 9601236, and the hospitality of the Center for Advanced Study in Behavioral Sciences. We have had helpful discussions with Chris Anderson, Bruno Broseta, Vince Crawford, Ido Erev, Drew Fudenberg, Dave Grether, Elef Gkioulekas, Yuval Rottenstreich, Rakesh Sarin, John Van Huyck, and Robert Weber, and research assistance from Hongjai Rhee, Chris Anderson, and Juin Kuan Chong. Barry Sopher generously provided data for us to analyze. Many helpful comments were received from anonymous referees and seminar participants at the Society for Mathematical Psychology conference (July 1996), the Russell Sage Foundation Summer Institute in Behavioral Economics (July 1996) the Economic Science Association meetings (October 1996), the Marketing Science Conference (March 1997), Bonn Conference on Theories of Bounded Rationality (May 1997), the FUR VIII Conference in Mons, Belgium (July 1997), Gerzensee ESSET Economic Theory Conference (July 1997), and seminars at Caltech, Harvard and Washington Universities, and the Universities of Alicante, Autonoma, California (Berkeley, Los Angeles), Chicago, Pennsylvania, Pittsburgh, Pompeu Fabra, Texas (Austin) and Texas A&M.

[1] Like most good ideas in economics, the adaptive and evolutionary interpretations of equilibration have a long pedigree. Weibull (1997) pointed out that in his famous passage, Adam Smith said that the a division of labor, which emerged as a consequence of the 'propensity to truck, barter, and exchange', emerged in a 'very slow and gradual' way (1981).

A different 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 reinforcement. Players who learn by reinforcement do not generally have beliefs about what other players will do. They care only about the payoffs strategies yielded in the past, not about the history of play that created those payoffs.

The belief and reinforcement approaches have been treated as fundamentally different since the 1950s. Until recently, nobody asked whether the two might be related, or how. But like two rivers with a surprising common source, or children raised apart who turn out to be siblings, belief and reinforcement are special kinds of one learning model. We suspect that the common heritage of these approaches was not discovered earlier because the information used by each approach is so different. Belief-based models do not specially reflect past successes (reinforcements) of chosen strategies. Reinforcement models do not reflect the history of how others played. The EWA approach includes both as special cases by incorporating both kinds of information, using three modelling features.

The crucial feature is how strategies are reinforced. In the choice reinforcement approach, when player 1 picks strategy s_1^i, and player 2 picks s_2^j, player 1's strategy s_1^i is reinforced according to the payoff $\pi_1(s_1^i, s_2^j)$. Unchosen strategies s_1^k $(k \neq i)$ are not reinforced at all. In EWA, the unchosen strategies are reinforced based on a multiple δ of the payoffs $\pi_1(s_1^k, s_2^j)$ they would have earned. This makes psychological sense because research on human and animal learning shows that people learn from experiences other than those which are directly reinforcing. (This expanded notion of reinforcement therefore liberates choice reinforcement from the limits of behaviorist psychology, toward something more cognitive and descriptive of humans.)

The second feature controls the growth rates of attractions. Attractions are numbers that are monotonically related to the probability of choosing a strategy. In reinforcement models attractions can grow and grow, which implies that convergence can be sharper (in the sense that choice probabilities diverge toward one and zero). In belief learning, attractions are expected payoffs, which are always bounded by the range of matrix payoffs. The EWA model allows growth rates to vary between these two bounds by using separate decay rates, ϕ for past attractions, and ρ for the amount of experience (which normalizes attractions).

The third modelling feature is initial attraction and experience weight. In belief models initial attractions must be expected payoffs given prior beliefs. In reinforcement models initial attractions are usually unrestricted. Therefore, initial attractions are unrestricted in EWA too. The initial experience weight $N(0)$ reflects a strength of prior in belief models, or the relative weight given to lagged attractions versus payoffs when attractions are updated.

When $\delta = 0$, $\rho = 0$, and $N(0) = 1$, the EWA attractions of strategies are equal to reinforcements, as used in many models. When $\delta = 1$ and $\phi = \rho$ (and initial attractions are determined by prior beliefs), the attractions of strategies are equal to their expected payoffs given beliefs in a general class. That is, reinforcing each strategy according to

what it would have earned (or did earn) is behaviorally equivalent to forming beliefs, based on observed history, and calculating expected payoffs. The equivalence holds because looking back at what strategies earned (or would have) in the past is the same as forming beliefs based on what others did in the past, then computing forward-looking expected payoffs based on those backward-looking beliefs.

EWA tries to mix appropriate elements of reinforcement and belief learning approaches in a way which makes sense. We think this can be judged by whether the parameters have clear psychological interpretations, and whether adding them improves statistical fit (adjusting, of course, for added degrees of freedom) and predictive accuracy. To test the empirical usefulness of EWA, we derived maximum-likelihood parameter estimates from three data sets. The data sets span a wide range of games: Constant-sum games with unique mixed-strategy equilibria; coordination games with multiple Pareto-ranked equilibria; and 'p-beauty contests' with unique dominance-solvable equilibria. Some empirical studies have evaluated belief and reinforcement models, but most have not compared them directly with statistical tests. Because EWA is a generalization which reduces to belief and reinforcement learning when parameters have certain values, it is easy to compare them to EWA and to each other.

In the next section, the EWA approach is defined and we show how a general class of choice reinforcement and adaptive belief-based approaches are special cases. The third section provides interpretations of the model parameters and discusses how they relate to principles of human learning. The fourth section describes previous findings and shows how our empirical implementation goes further than earlier work. The fifth section reports parameter estimates from several data sets. The last section concludes and mentions some future research directions.

2 The Experience-weighted Attraction (EWA) Model

We start with notation. We study n-person normal-form games. Players are indexed by i ($i = 1, \ldots, n$), and the strategy space of player i, S_i consists of m_i discrete choices, that is, $S_i = \{s_i^1, s_i^2, \ldots, s_i^{m_i-1}, s_i^{m_i}\}$. $S = S_1 \times \ldots \times S_n$ 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. $s = (s_1, \ldots, s_n) \in S$ is a strategy combination, and it consists of n strategies, one for each player. $s_{-i} = (s_1, \ldots, s_{i-1}, s_{i+1}, \ldots, s_n)$ is a strategy combination of all players except i. S_{-i} has a cardinality of $m_{-i} = \Pi_{j=1,j\neq i}^{n} m_j$. The scalar-valued payoff function of player i is $\pi_i(s_i, s_{-i})$. 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 $\pi_i(s_i(t), s_{-i}(t))$.

EWA assumes each a strategy has a numerical attraction, which determines the probability of choosing that strategy (in a precise way made clear below). Learning models require a specification of initial attractions, how attractions are updated by experience, and how choice probabilities depend on attractions.

2.1 The EWA 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. The second variable is $A_i^j(t)$, player i's attraction of strategy j 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,

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

The parameter ρ is 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 δ, and weights payoffs 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 as $[\delta + (1-\delta) \cdot I(s_i^j, s_i(t)] \cdot \pi_i(s_i^j, s_{-i}(t))$.

The rule for updating attraction sets $A_i^j(t)$ to be the sum of a depreciated, experience-weighted previous attraction $A_i^j(t-1)$ plus the (weighted) payoff from period t, normalized by the updated experience weight:

$$A_i^j(t) = \frac{\phi \cdot N(t-1) \cdot A_i^j(t-1) + [\delta + (1-\delta) \cdot I(s_i^j, s_i(t))] \cdot \pi_i(s_i^j, s_{-i}(t))}{N(t)}. \tag{2}$$

The factor ϕ is a discount factor or decay rate, which depreciates previous attraction.

2.2 Choice reinforcement

In early reinforcement models (and some recent ones) choice probabilities are updated directly (e.g., Bush and Mosteller, 1955; cf. Cross, 1983). In more recent models

(Harley, 1981; Roth and Erev, 1995), strategies have levels of reinforcement or propensity which are incremented cumulatively by received payoffs (and perhaps normalized, Arthur, 1991). We use the latter form, which gives more modelling freedom[2] and avoids some clumsy technical features (imposing boundary conditions so probabilities do not grow too high or low).

The initial reinforcement level of strategy j of player i, s_i^j, is $R_i^j(0)$. These initial reinforcements can be assumed a priori (based on a theory of first-period play) or estimated from the data. Reinforcements are updated according to two principles:

$$R_i^j(t) = \begin{cases} \phi \cdot R_i^j(t-1) + \pi_i(s_i^j, s_{-i}(t)) & \text{if } s_i^j = s_i(t), \\ \phi \cdot R_i^j(t-1) & \text{if } s_i^j \neq s_i(t). \end{cases} \tag{3}$$

The two principles can be reduced to a single updating equation:

$$R_i^j(t) = \phi \cdot R_i^j(t-1) + I(s_i^j, s_i(t)) \cdot \pi_i(s_i^j, s_{-i}(t)). \tag{4}$$

It is easy to see that this updating formula is a special case of the EWA rule, when $\delta = 0$, $N(0) = 1$, and $\rho = 0$. Thus, choice reinforcement in this form is a special case of experience-weighted attraction learning.[3]

2.3 Belief-based Models

In a belief-based model, players tend to choose strategies which have high expected payoffs given beliefs formed by observing the history of what others did. While there are many ways of forming beliefs, we consider a fairly large class of weighted fictitious play models, which include familiar ones like fictitious play (Brown, 1951) and Cournot (1960)

[2] In the Cross model, strategies have utilities which are weighted averages of past utilities and current payoffs (for chosen strategies), and players maximize utility. Sarin (1995) shows that when the weight on current payoff declines over time, this model behaves similarly to the Harley version in which attractions grow. The similarity reflects the fact that both models build in a declining effect of marginal reinforcements.

[3] Some reinforcement models add other parameters. Roth and Erev (1995) add a parameter which cuts off attractions close to zero, to avoid negative attractions. Erev and Roth (1997) add three parameters which allow reinforcement to depend on payoffs minus an (updated) reference point (as in Bush and Mosteller, 1955; Cross, 1983), where the updating may be different for losses and gains. They also add a parameter which smears a portion of the chosen-strategy reinforcement to neighboring strategies, to reflect a kind of experimentation or generalization which is (locally) similar to our δ parameter. Camerer and Ho (1998) compare the local-generalization specification with δ updating in the EWA model and find that local-generalization fits much worse.

best-response as special cases (see Fudenberg and Levine, 1995; Cheung and Friedman, 1997).[4]

In the weighted fictitious play model, prior beliefs of opponents' strategy combinations are expressed as a ratio of hypothetical counts of observations of strategy combination s^k_{-i}, denoted by $N^k_{-i}(0)$. These observations can then be naturally integrated with actual observations as experience accumulates. (Carnap (1962) shows an elegant set of axioms which implies this structure, which corresponds to Bayesian updating with a Dirichlet-distributed prior.) In our view, specifying prior beliefs (and computing initial expected payoffs based on the prior) is a crucial feature of belief models, though some papers have not imposed this assumption. Without specifying a prior, there is no guarantee that the updated beliefs which result from mixing initial expected payoffs with later experience will be valid beliefs (i.e., nonnegative probabilities which sum to one).

We also allow past experience to be depreciated or discounted by a factor ρ (presumably between zero and one). Formally, the prior beliefs for player i about choices of others are specified by a vector of relative frequencies of choices of strategies s^k_{-i}, denoted $N^k_{-i}(0)$. Call the sum of those frequencies (dropping the player subscript for simplicity) $N(t) = \sum_{k=1}^{m_{-i}} N^k_{-i}(t)$. Then the initial prior $B^k_{-i}(0)$ is:

$$B^k_{-i}(0) = \frac{N^k_{-i}(0)}{N(0)}, \tag{5}$$

with $N^k_{-i}(0) \geq 0$ and $N(0) > 0$. Beliefs are updated by depreciating the previous counts by ρ, and adding one for the strategy combination actually chosen by the other players. That is,

$$B^k_{-i}(t) = \frac{\rho \cdot N^k_{-i}(t-1) + I(s^k_{-i}, s_{-i}(t))}{\sum_{h=1}^{m_{-i}} [\rho \cdot N^h_{-i}(t-1) + I(s^h_{-i}, s_{-i}(t))]}. \tag{6}$$

Expressing beliefs in terms of previous-period beliefs,

$$B^k_{-i}(t) = \frac{\rho \cdot B^k_{-i}(t-1) + \frac{I(s^k_{-i}, s_{-i}(t))}{N(t-1)}}{\rho + \frac{1}{N(t-1)}} = \frac{\rho \cdot N(t-1) \cdot B^k_{-i}(t-1) + I(s^k_{-i}, s_{-i}(t))}{\rho \cdot N(t-1) + 1}. \tag{7}$$

[4]When the description 'fictitious play' is used below, we mean traditional fictitious play in which all past observations are weighted equally. Also, Camerer and Ho (1998) estimate models in which ϕ varies across periods, which generalizes weighted fictitious play to include cases where the weight rises or falls over time. Allowing a time-varying weight does not improve fit much, so assuming a fixed ϕ seems reasonable.

This form of belief updating weights observations from one period ago ρ times as much as the most recent observation. This includes Cournot dynamics ($\rho = 0$; only the most recent observation counts) and fictitious play ($\rho = 1$; all observations count equally) as special cases. The general case $0 \leq \rho \leq 1$ is a compromise in which all observations count but more recent observations count more.

Expected payoffs in period t, $E_i^j(t)$, are taken over beliefs according to

$$E_i^j(t) = \sum_{k=1}^{m_{-i}} \pi_i(s_i^j, s_{-i}^k) \cdot B_{-i}^k(t). \tag{8}$$

The crucial step is to express period t expected payoffs as a function of period $t-1$ expected payoffs. Substituting equation (2.7) into (2.8) and rearranging yields:

$$E_i^j(t) = \frac{\rho \cdot N(t-1) \cdot E_i^j(t-1) + \pi(s_i^j, s_{-i}(t))}{\rho \cdot N(t-1) + 1}. \tag{9}$$

This equation makes the kinship between EWA and belief approaches transparent. Formally, suppose initial attractions are equal to expected payoffs given initial beliefs which arise from the 'experience-equivalent' strategy counts $N_{-i}^k(0)$, so $A_i^j(0) = E_i^j(0) = \sum_{k=1}^{m_{-i}} \pi_i(s_i^j, s_{-i}^k) \cdot B_{-i}^k(0)$. Then substituting $\delta = 1$ and $\rho = \phi$ into the attraction updating equation (2) gives attractions which are exactly the same as updated expected payoffs in (9). Hence, the weighted belief models are a special case of EWA.

The close relation between reinforcement and belief learning is surprising because the two approaches have generally been treated as fundamentally different (e.g., Selten, 1991, p 14). However, some connection between reinforcement and belief learning was recognized very recently by others (unbeknownst to us). Fudenberg and Levine (1995, pp. 1084-1085) and Cheung and Friedman (1997, p. 54-55) both pointed out that expected payoffs computed using fictitious play beliefs, and based on history, are asymptotically the same as histories of actual payoffs. But their arguments are based on long-run asymptotic equivalence between a distribution (possible payoffs) and a sample from it (actual payoffs). Neither seemed to explicitly recognize that even in the short run, there is an exact equivalence between a special kind of reinforcement learning (EWA) and weighted fictitious play.[5]

[5]For example, Cheung and Friedman (1997) make their point by "assum[ing] for the moment (very counterfactually!), that the player somehow managed to play both strategies each period". Then "dropping the counterfactual", they show that the average experienced payoffs will correspond, up to some noise, to expected payoffs. Counterfactual simulation of foregone payoffs is precisely the mental process invoked by δ in EWA. However, the 'noise' is correlated with past observations which are included explicitly in EWA, so the relation between EWA and weighted fictitious play is exact rather than approximate.

The contrast with EWA makes clear that belief models actually make three separate assumptions: Players' initial attractions are expected payoffs based on some prior; players update attractions using EWA with $\delta = 1$; and attractions are a weighted average of lagged attractions and payoffs ($\phi = \rho$). We think the most intuitively appealing assumption is the best-responsiveness to foregone payoffs embodied in $\delta = 1$, rather than the weighted-average restriction $\phi = \rho$ or the restriction on first-period play. EWA allows one to separate the three features of belief learning: Players could have attractions which begin and grow differently than belief models assume, but update those attractions in a belief-learning way. Such players are a special kind of EWA learner.

The nonlinear interplay of parameters in the EWA updating rules is why, as a model of human learning, EWA is potentially superior to simply running a regression of choices against reinforcements and expected payoffs or combining the two in a weighted average. Reinforcements and expected payoffs differ in three crucial dimensions– initial attractions and experience weight $N(0)$, the weight δ on foregone payoffs in updating attractions, and whether attractions can grow outside the bounds of possible payoffs (which depends on ϕ and ρ). EWA is not a convex combination of reinforcement and belief models because these three dimensions are controlled by separate parameters. That is, a weighted average in which expected payoffs are given weight δ and reinforcements have weight $1 - \delta$ will update attractions like EWA does, but that weighted average will not allow the wide range of initial attractions, experience rates, and growth rates available in EWA.[6]

2.4 Choice probabilities

Attractions must determine probabilities of choosing strategies in some way. $P_i^j(t)$ should be monotonically increasing in $A_i^j(t)$ and decreasing in $A_i^k(t)$ (where $k \neq j$). Three forms have been used in previous research: Exponential (logit), power, and normal (probit). In estimation reported below we use the logit function, which is commonly used in studies of choice under risk and uncertainty, brand choice, etc. (Ben-Akiva and Leman, 1985; Anderson, Palma and Thisse, 1992), and is given by

$$P_i^j(t+1) = \frac{e^{\lambda \cdot A_i^j(t)}}{\sum_{k=1}^{m_i} e^{\lambda \cdot A_i^k(t)}}. \tag{10}$$

The parameter λ measures sensitivity of players to attractions. Sensitivity could vary due to the psychophysics of perception or whether subjects are highly motivated or not. In this probability function, the exponent in the numerator is just the weighted effect of strategy j's attraction, $\lambda \cdot A_i^j(t)$, on the probability of choosing strategy j. Models in which cross-effects of attractions on other strategies' choice probabilities are allowed

[6]Indeed, Camerer and Ho (1998) show that EWA fits much better than a convex combination of belief and reinforcement learning, in two coordination games.

have been estimated (Mookerjhee and Sopher, 1997) but we do not have the degrees of freedom to do so.[7]

The logit, power, and probit probability functions each have advantages and disadvantages. The exponential form has been used to study learning in games by Mookerjhee and Sopher (1994, 1997), Ho and Weigelt (1996), and Fudenberg and Levine (in press), and in 'quantal response equilibrium' models by Chen, Friedman and Thisse (in press) and McKelvey and Palfrey (1995, 1996). Cheung and Friedman (1997) used the probit form. The exponential form is are invariant to adding a constant to all attractions.[8] As a result, 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.[9]

The power probability form is given by

$$P_i^j(t+1) = \frac{(A_i^j(t))^\lambda}{\sum_{k=1}^{m_i}(A_i^k(t))^\lambda}. \tag{11}$$

The power form is invariant to multiplying all attractions by a constant. Because of this invariance, the parameters $N(0)$ and ρ make no difference when the power form is used (i.e., they are not identified).[10]

Depending on one's purpose, being able to ignore $N(0)$ and ρ can be an advantage or disadvantage. For the purpose of distinguishing different models, it is a big disadvantage because models impose different restrictions on $N(0)$ and ρ. By using the power form,

[7] In Mookerjhee and Sopher (1997), the exponent in the probability equation numerator is the sum of weighted effects of <u>all</u> the attractions, $\sum_{k=1}^{m_i} \lambda_{jk} \cdot A_i^k(t)$, where λ_{jk} is the cross-effect of strategy k's attraction on strategy j's score. This model allows cross-effects in which one strategy's attraction can affect other strategies' choice probabilities differently. These cross-effects are hard to interpret without knowing more about similarity of strategies or some other basis for one strategy's attraction to affect others differently. Nonetheless, they have some significance as a whole in the Mookerjhee-Sopher analysis of constant-sum games. Estimating them for our median-action and p-beauty contest data uses up far too many degrees of freedom because there are too many strategies. Including cross-effects could proceed particularly efficiently if some structural considerations were used to restrict coefficients a priori (as in Sarin and Vahid's, 1997, use of strategy similiarity).

[8] As a result, one must normalize $A_i^j(0)$ to equal a constant for one value of j in order to identify parameters. There is some evidence that adding a constant to payoffs does matter (Bereby-Meyer and Erev, 1997) but there is also evidence that logit fits better than power, so we regard the choice of proper form as a matter of one's purpose and yet-unresolved empirical debate.

[9] Borgers and Sarin (1996) avoid this problem by <u>adding</u> x to all other strategies when a chosen strategy loses x.

[10] The parameter ρ disappears because it only appears in the updating equation denominator $\rho \cdot N(t-1)+1$ which is common to all attractions and thus cancels out in the power form. Then EWA attractions at time t depend only on recent payoffs and the product $A_i^j(0) \cdot N(0)$. 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).

the difference between belief-based, reinforcement, and EWA models, besides initial attractions, is only one parameter, δ, rather than three parameters. For the purposes of estimating any one model reliably, however, conserving degrees of freedom is good so the power form is better. Since our main purpose in this paper is comparing models, having the extra tools to distinguish theories is a large advantage so we use the logit form rather than the power form. This choice of probability rule is, of course, not an essential part of the EWA model.

Ultimately, it is an empirical question whether the logit, probit or power forms fit better (adjusting for degrees of freedom). Previous studies show roughly equal fits of logit and power (Tang, 1996; Chen and Tang 1996; Erev and Roth, 1997) or better fits for the logit form over the power form (Camerer and Ho, 1998).

3 Interpreting EWA parameters

We think it is crucial to ask how a learning model's parameters can be interpreted, what general behavioral principles of learning they capture, and, for EWA, how it reveals assumptions implicit in reinforcement and belief learning. Asking these questions about any learning theory avoids the danger of adding parameters just to improve statistical fit, without adding new insight or respecting what is known in other disciplines. In addition, if parameters have natural psychological interpretations they can be measured in other ways (e.g., response times and attention measures) and used in psychological modelling.

3.1 Learning principles and δ

The parameter δ measures the relative weight given to foregone payoffs, compared to actual payoffs, in updating attractions. This is the most important parameter in EWA because it shows most clearly the different ways in which EWA, reinforcement and belief models capture two basic principles of learning– the law of actual effect and the law of simulated effect.

Many decades of learning experiments, mostly with (nonhuman) animal subjects, show that successful chosen strategies are subsequently chosen more often. Behaviorist psychologists call this the 'law of effect' (Thorndike, 1911; Herrnstein, 1970). We relabel this the 'law of actual effect' because behaviorists took it for granted for years that the only effect on subsequent choices was produced by rewards for actual choices. The behaviorists eschewed 'mentalist' constructs like imagination, which allowed the possibility that foregone rewards could affect the probability of choosing new strategies, until a series of demonstrations showed that those cognitive constructs are necessary. When applied to humans playing games with a known payoff matrix, it is sensible to propose a corollary general principle, the 'law of simulated effect'. The law of simulated effect states that unchosen strategies which would have yielded high payoffs– simulated successes– are

more likely to be chosen subsequently. Many experiments on reinforcement learning are consistent with this principle.[11]

Furthermore, most research on human and machine learning assumes that the basic process driving learning is not reinforcement, per se, but the reduction of errors. Since errors are measured by the difference between what players received and what they could have received, error-reduction algorithms use both actual payoffs and foregone payoffs too, obeying both the law of actual effect and the law of simulated effect.

The empirical strengths of the law of effect and the law of simulated effect are the key to distinguishing different models of learning in games, and are calibrated by δ. Reinforcement insists that only actual effects matter (δ=0). Belief models implicitly require that actual and simulated effects are equally strong ($\delta = 1$). EWA takes the middle ground.

The parameter δ also can be seen as a way of endogenizing a reference point or aspiration level. Many studies show that the reinforcement value of a fixed payoff can vary, depending on what aspiration level the payoff is compared to. Some reinforcement models build in an aspiration level directly, and adjust it across time based on observed payoffs, which requires at least two free parameters (an initial level and an adjustment rate). In EWA, reinforcing strategies according to foregone payoffs means the probability of a chosen strategy $s_i(t)$ only increases if its payoff is larger than δ times the average foregone payoff (see our working paper for details). Thus, a larger δ creates a more extreme aspiration level. EWA therefore creates an endogeneous, adjustable reference point at no extra parametric cost.

If δ is interpreted as the weight placed on foregone payoffs, many generalizations spring to mind. The size of the weight δ could depend on the size of the foregone payoff or on its sign, to allow the possibilities that unusually large or small foregone payoffs catch a player's attention, or that players are more sensitive to losses than to gains (cf. loss-aversion in risky choices, e.g., Tversky and Kahneman, 1992). If players are more sensitive to foregone payoffs for strategies which are closer to the chosen strategy, or more similar, then δ will depend on the distance or similarity between each strategy and the chosen strategy $s_i(t)$ (cf. Sarin and Vahid, 1997).

[11] For example, anxious patients can be taught to fear a picture of a triangle (a conditioned stimulus, or CS) when it is followed by a loud annoying noise (an unconditioned stimulus, or UCS). When patients are told to simply imagine the UCS several times, their imagination increases the strength of their conditioned fear response to the triangle CS (Davey and Matchett, 1990). A related phenomenon is 'incubation', in which presentation of the CS itself increases the fear response (Eysenck, 1979). In these cases, people are not learning by direct reinforcement, but merely by imagining either the UCS's reinforcement, or the reinforcement which typically follows a CS. There is also vast evidence that children and primates (and of course, adult humans too) learn by imitating others, which illustrates another kind of learning from simulated or hypothetical effects.

3.2 Growth of Attractions, ρ and ϕ

The parameter ϕ depreciates past attractions, $A_i^j(t)$.[12] The parameter ρ depreciates the experience measure $N(t)$. It captures decay in the strength of prior beliefs, which can be different than decay of early attraction (captured by ϕ). These factors combine cognitive phenomena like forgetting with a deliberate tendency to discount old experience when the environment is changing.

One way to interpret ρ and ϕ 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 numerator 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 a running total of (depreciated) attraction, updated by each period's payoffs. The denominator is $\rho \cdot N(t-1) + 1$. This term is 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 by 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.

EWA allows attractions to grow faster than an average, but slower than a cumulative total. An analogy might help illustrate. 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 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. Then if two people perform equally well on average every year, the 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.

The depreciation rate parameters ϕ and ρ can also be understood by how they control slowdown in learning rate or sharpness of convergence. Solving recursively for steady-state attraction levels shows that those levels equal the ratio $\frac{1-\rho}{1-\phi}$ times the steady-state average payoff. Thus, when $\rho = 0$ as in reinforcement learning, attractions can end up outside the bounds of payoff levels (and they grow as large as possible, holding ϕ constant). When $\rho = \phi$, as in belief-learning, steady-state attraction levels are equal to steady-state average payoffs. The implication of these two possibilities depends on

[12] A 'primacy effect' (or 'imprinting', Cheung and Friedman, 1997), in which early observations are remembered more strongly than recent ones, can be expressed by $\phi \geq 1$.

how attractions determine probabilities. In the logit probability form, only differences in attraction levels affect choice probabilities. Therefore, given a fixed value of λ, attractions which can grow outside the bounds of payoff levels have a wider range across strategies. This allows the possibility of sharper convergence in the sense that choice probabilities can converge closer to the boundaries at zero and one.

When attractions are bounded to be close to payoff levels, convergence cannot be as sharp. In the power probability form, only ratios of attraction levels matter. Therefore, if attractions grow the relative impact of new reinforcements falls; learning slows down. Ceteris paribus, reinforcement learning requires convergence to be as sharp as possible (in the logit form) or requires learning to slow down as quickly as possible (in the power form), while belief learning requires the opposite. EWA is able to choose an intermediate value of ρ which tailors the sharpness of convergence or rate of learning to the data.

3.3 Initial attractions $A_i^j(0)$ and their strength $N(0)$

The term $A_i^j(0)$ represents the initial attraction, which might be derived from an analysis of the game, from surface similarity between strategies and strategies which were successful in similar games, etc. Belief models restrict the $A_i^j(0)$ strongly by requiring initial attractions to be derived from prior beliefs. This requires, for example, that weakly dominated strategies will always have (weakly) lower initial attractions than dominant strategies. EWA allows more flexibility.

For example, suppose players make first-period choices randomly, by choosing what was chosen previously in a different game, by setting each strategy's initial attraction equal to its minimum payoff (the maximin rule) or maximum payoff (the maximax rule)[13], or by choosing stochastically among selection principles like payoff-dominance, risk-dominance, loss-avoidance, etc. All these decision rules are plausible models of first-period play, but none of them generate initial attractions which are always expected payoffs given some prior beliefs.

We consider the scientific problem of figuring out how people choose their initial strategies as fundamentally different than explaining how they learn. Leaving initial attractions unrestricted makes them numerical placeholders which can be filled by a theory of first-period play which supplies attractions as an input to EWA. That combination would be a complete theory of behavior in games, from start to finish.

The initial-attraction weight $N(0)$ appears in the EWA model to allow players in belief-based models to have an initial prior which has a certain strength (measured in units of actual experience). In EWA, $N(0)$ is therefore naturally interpreted as the

[13]Making a strategy's initial attraction equal to its minimum payoff, for example, is implicitly putting all the belief weight on the choices by others which yield that minimum. But the choices by others which lead to minima for different strategies are likely to be different. So the implicit beliefs underlying each attraction will be different.

strength of initial attractions, relative to incremental changes in attractions due to actual experience and payoffs. Fixing $N(0) = 1$ means that, unit for unit, initial attractions $A_i^j(0)$ and chunks of reinforcement from payoffs are weighed equally when attractions are updated. This is easiest to see by fixing $\delta = 1$ for simplicity and directly computing the attraction after two periods, $A_i^j(2)$, which gives

$$A_i^j(2) = \frac{\phi^2 \cdot A_i^j(0) \cdot N(0) + \phi \cdot \pi_i(s_i^j, s_{-i}(1)) \cdot + \pi_i(s_i^j, s_{-i}(2))}{\rho^2 \cdot N(0) + \rho + 1}. \tag{12}$$

The parameter ϕ captures the declining weight placed on payoffs from more distant periods of actual experience, compared to more recent periods. (That is, the older period 1 payoff $\pi_i(s_i^j, s_{-i}(1))$ is weighted by ϕ but the recent period 2 payoff $\pi_i(s_i^j, s_{-i}(2))$ is not.) Like previous payoffs, the initial attraction is also weighted by a power of ϕ (ϕ^2, because it 'happened' two periods earlier), but is also weighted by $N(0)$. Thus, the parameter $N(0)$ captures the special weight placed on the initial attractions, compared to increments in attraction due to payoffs. $N(0)$ can therefore be thought of as a 'pre-game (introspective) experience' weight. If $N(0)$ is small the effect of the initial attractions is quickly displaced by experience. If $N(0)$ is large then the effect of the initial attractions persists.

Notice that updating the experience-weight by $N(t) = \rho \cdot N(t-1) + 1$ implies a steady-state value of $N^* = \frac{1}{1-\rho}$. In estimation, we have found it useful to restrict $N(0)$ to be less than N^*. This implies $N(t-1) \leq N(t)$; the experience weight is (weakly) rising over time. Since the relative weight on decayed attractions, compared to recent reinforcement, is always increasing, the relative weight on observed payoffs is always declining. This implies a 'law of declining effect' which is widely observed in research of learning.

The flexibility of initial attractions and experience weight allows one to fit a variety of models. Theories of equilibrium behavior are special cases in which all 'learning' occurs before the game starts. For example, a 'stubborn' game-theoretically-minded player sets $A_i^j(0)$ equal to the equilibrium payoffs of each strategy and act as if $N(0)$ is infinite (meaning that no amount of game-playing experience can outweigh the prior calculation). An adaptive game theorist assumes $A_i^j(0)$ are equilibrium payoffs but has a small $N(0)$, so she learns from experience. A player who does not begin with prior beliefs, but updates according to experience as a belief learner does, has $\phi = \rho$ and $\delta = 1$ with arbitrary $A_i^j(0)$.

Other features could conceivably be included. Players who tend to repeat previously-chosen strategies, regardless of their outcomes, reveal a 'status quo bias' or 'habit' (Majure, 1995). Similarly, imitative learning is just acquiring somebody else's habit It is not clear how to add this feature to EWA because it presumes payoff-independent reinforcement of chosen strategies.

4 Previous research

In this section we briefly summarize previous research (see Camerer, in progress, for more details).

Several papers investigate only belief learning. Cheung and Friedman (1997) (CF) estimated a weighted fictitious play model on individual-level data from four games (hawk-dove, stag hunt, 'buyer seller' and battle-of-the-sexes). They find substantial heterogeneity across subjects but stability across games in the equivalents of ϕ and λ. A general belief model (allowing idiosyncratic shocks in beliefs) was developed by Crawford (1995) to fit data from coordination games, extended by Broseta (1997) to allow ARCH error terms, and applied by Crawford and Broseta (in press) to coordination with preplay auctions. Brandts and Holt (in press) and Cooper, Kagel, and Garvin (1997) simulate fictitious play in signaling games. Boylan and El-Gamal (1992) compare fictitious play and Cournot learning in coordination and dominance-solvable games; they find overwhelming relative support for fictitious play.

Other studies concentrate only on reinforcement learning. Versions of reinforcement in which probabilities were reinforced directly, or cumulative payoffs normalized, were used by Bush and Mosteller (1955), Cross (1983) and Arthur (1991). Harley (1981) posited a reinforcement model using cumulative payoffs and simulated its behavior in several games. The Harley model was later extended by Roth and Erev (1995) to include spillover of reinforcement to neighboring strategies. Their model fits the time trends in ultimatum, public good, and responder-competition games but converges much too slowly. McAllister (1991) shows that a modified Cross model which uses foregone payoff information fits weak-link data modestly well. Sarin and Vahid (1997) show that a modified Cross model with distance-weighted spillover of reinforcement to similar strategies fits data on coordination experiments with low information fairly well.

These studies of belief and reinforcement learning find that each approach, evaluated separately, has some explanatory power. Other studies compared models.[14] Erev and Roth (1997) add an adjustable reference point to their earlier model (cf. Cross, 1983). The extended model fits slightly better than fictitious play, at the individual level, in constant-sum games played for 100 or more periods. Mookerjhee and Sopher (1994,1997) (MS) compare average-payoff reinforcement and fictitious play in constant-sum games; reinforcement does somewhat better. Ho and Weigelt (1996) compare modified versions of fictitious play and choice reinforcement (the MS 'vindication' model) in coordination games with multiple Nash equilibria. Fictitious play fits better.

Many variants of weighted fictitious play and reinforcement (and other models) were compared by Tang (1996a,b) in games with mixed-strategy equilibria. Reinforcement does better in most games. Chen and Tang (1996) fit models to data from two public

[14] In still another approach, models in which players learn to shift weight across various rules (or 'methods'), rather than across strategies, were studied by Tang and by Stahl (1996, 1997). In Tang's comparison 'method-learning' does slightly worse than reinforcement. Stahl (1997) finds that players seem to weight rules which mimic choices of others or best-respond given diffuse priors.

goods games. In one game equilibration is so fast that Nash equilibrium outpredicts the learning models. In the other game reinforcement does better.

The overall picture from previous research is somewhat blurry. Comparisons appears to favor reinforcement over belief learning in constant-sum games but specifications of the models, estimation techniques, and games vary across studies. Our approach allows one to compare models more systematically by including features which have been used differently in different studies. Two general features are notable.

First, most papers assume equal initial attractions or, for belief models, uniform priors. Some papers estimate initial attractions using first-period data (which does not generally optimize overall fit). Our procedure is more general because we estimate initial attractions and experience weight as part of an overall maximization of fit. Estimating initial experience weight $N(0)$ allows belief models to express a prior strength. This is an important feature of belief learning; omitting it may explain why belief models have sometimes fit relatively poorly (in Mookerjhee and Sopher, 1997; Tang, 1996a; Chen and Tang, 1996; Erev and Roth, 1997).

Second, some reinforcement models assume averaged-payoffs affect choices, while others assume reinforcements cumulate. This difference can be captured by allowing ρ to vary between ϕ (for averaging) and 0 (for maximum cumulation), as EWA does. In addition, some studies of belief learning did not allow weighted fictitious play, as EWA does. Including ϕ and ρ therefore allows us to determine whether previous mixed results depend on whether reinforcements are averaged or cumulated, and on whether belief models are weighted.

Our methodology for model estimation is more general than most earlier papers in four ways. First, we compare across three classes of games using the same estimation technique (only Cheung and Friedman (1997) have done this in one paper). Second, our method uses standard statistical tests to judge whether differences in fit are due to chance, or put differently, to decide whether simple models are too simple or not. (Only Stahl (1996, 1997) compared models using tests which correct for the number of free parameters.) Third, we calibrate models on the first 70% of the periods in each sample and predict the rest of the sample to validate the estimates and avoid overfitting (no previous paper has done this). Fourth, we allow heterogeneity across individuals by comparing a model with a single class of agents with a two-segment model, which has not been done before.[15]

[15]The only paper which estimates individual-level parameters on these kinds of models is Cheung and Friedman (1997). While the median parameter estimates are reasonable and similar across games when expected to be, the individual-level estimates are variable (e.g., a third of the ϕ estimates are negative and a sixth are above one). This reflects some imprecision in individual-level estimation which suggests that multiple-segment estimation, which lies between single-segment estimation and individual-level estimation may be a reasonably parsimonious compromise between the desires to allow heterogeneity and to estimate reliably.

5 Parameter Estimation from Experimental Data

5.1 Estimation Strategy

We estimated the values of model parameters from three samples of experimental data[16] and validate the models by predicting behavior out of sample. The games are: Constant-sum games with unique mixed-strategy equilibria (and one weakly dominated strategy); a 'median-action' coordination with multiple Pareto-ranked equilibria; and a dominance-solvable 'p-beauty contest' game with a unique equilibrium. We chose these games for several reasons.

First, the games have a range of different structural features (as in Cheung and Friedman (1997) and Stahl (1997)). This avoids the possible mistake of concluding that a model generally fits well because it happens to fit one class of games.

Second, the games have different spans– the constant-sum games last 40 periods and the others last 10 periods. Longer spans provide more data and more power for estimating individual differences. But a mixture of long and short spans are valuable too, because some games– like the coordination and beauty contest games reported below– converge fairly rapidly. Learning models should be able to explain why convergence is fast in those games and slow in others.

Third, most previous studies have reported results which are favorable to either reinforcement or belief learning. The games we use each present some new challenges to these models. The presence of dominated strategies in the constant-sum games is a challenge for belief models, which predict those strategies will be played relatively rarely. Rapid convergence in the coordination and dominance-solvable games is a challenge for reinforcement learning (see also Van Huyck, Battalio and Rankin, 1997).

Next we describe some general features of the estimation method. For simplicity we assume that players' strategies are the stage-game strategies, and denote player i's strategy choice in period t by $s_i(t)$. (Of course, in general strategies could be history-dependent or be decision rules.)

We use a 'latent class' approach in which there are one or two segments of players, and all players in a segment are assumed to have the same parameter values. This technique is standard in some fields (e.g., analyses of brand choice in marketing) and was also suggested by Crawford (1995).[17] The single-class estimation provides a representative-agent

[16] Our working paper includes two other samples of data, on weak-link coordination games and matching pennies (Mookerjhee and Sopher, 1994). We dropped these because the weak-link results did not have a long enough span to permit both calibration and validation; calibration is reported in Camerer and Ho, 1997). The matching pennies data did not distinguish models from each other or from Nash equilibrium.

[17] Note that even though all agents in a class have the same parameter values, after the first period they will be predicted to behave differently because their actual choices and experiences vary.

benchmark. Allowing a second class gives a clue about how important it is to allow heterogeneity. For example, our results show that in constant-sum games allowing a second class hardly improves the fit at all while fit is improved substantially in coordination games.

The two-class procedure makes sense because in these data sets there are not enough observations per subject to reliably estimate many more classes.[18] And while including more segments would be desirable, assuming all players have the same parameters does not impose a heavier penalty for some models than for others, so it is unlikely that the two-class assumption will lead us to incorrectly favor one model over another.

We estimate initial attractions $A^j(0)$ (suppressing the player subscript). Assuming equal initial attractions (or equal priors in belief models) saves degrees of freedom but fits poorly in our data. Estimating initial attractions also creates numbers which may be useful for constructing a good theory of first-period play.

Let the stage game be repeated for T rounds. Recall that the indicator function $I(s_i^j, s_i(t))$ is equal to 1 if $s_i^j = s_i(t)$ and 0 otherwise. Define the vector of initial attractions for player i to be $A_i(0) \stackrel{def}{=} (A_i^1(0), A_i^2(0), \ldots, A_i^{m_i}(0))$. Since we study symmetric games and assume all players have the same parameter values, for this paper there is a common set of initial attractions $A(0) = A_i(0) \forall i$. Define the number of subjects by N. The overall sample size, $.7 \cdot T \cdot N$, is denoted by M. Then the log-likelihood function, $LL(A(0), N(0), \phi, \rho, \delta, \lambda)$, is

$$LL(A(0), N(0), \phi, \rho, \delta, \lambda) = \sum_{t=1}^{0.7 \cdot T} \sum_{i=1}^{N} ln(\sum_{j=1}^{m_i} I(s_i^j, s_i(t)) \cdot P_i^j(t)) \quad (13)$$

$$= \sum_{t=1}^{0.7 \cdot T} \sum_{i=1}^{N} ln(\sum_{j=1}^{m_i} I(s_i^j, s_i(t)) \cdot \frac{e^{\lambda \cdot A_i^j(t-1)}}{\sum_{k=1}^{m_i} e^{\lambda \cdot A_i^k(t-1)}}) \quad (14)$$

Keep in mind that in the exponential form, attractions are only identified up to a constant, so we must fix one of the $A^j(0)$ to equal a constant. We searched over parameter values to maximize the LL function using the MAXLIK routine in GAUSS, which uses a gradient method. To avoid converging to local optima we tried a variety of starting points. We restricted ϕ, λ to be positive, $0 \leq \delta, \rho \leq 1$.

In order to make the value of $N(0)$ interpretable as a weight on initial attractions relative to reinforcing payoffs, we restricted the range of $A^j(0)$ to be less than or equal to the difference between the minimum and maximum payoffs (while also setting one

[18] Two segments are also useful because one can then compare a two-segment EWA model with a two-segment model in which one segment are reinforcement learners and the other segment are belief learners. We did this in Camerer and Ho (1998) on weak-link and median-action data and EWA fits much better than the mixture model.

of the attractions equal to zero for identifiability).[19] Since this restriction is naturally satisfied in belief models, in order to compare EWA to belief and reinforcement learning we imposed it in EWA and reinforcement as well.[20] We also restricted $0 \leq N(0) \leq \frac{1}{1-\rho}$ to guarantee that the weights $N(t)$ rise over time.

Standard errors of parameters were estimated using a jackknife procedure. In each run of the jackknife, one subject was excluded from the analysis and the model was estimated using all remaining subjects.[21] Doing this sequentially produces N vectors of estimates (where N is the number of subjects). The parameter standard errors are then the standard deviations of parameter estimates across the N runs. (Correlations between parameters can also be computed this way, and help detect identification problems.)

Since EWA is always more general than the special cases, it will necessarily fit the data better so there is some danger of overfitting. To guard against this, we both calibrate the models and validate them, by deriving MLE estimates using the first 70% of the observations in each sample, then using these estimates to predict the path of play in the remaining 30% of the sample. This procedure uses enough data to estimate parameters reliably, but also forecasts out-of-sample to ensure models are not being overfit. To evaluate model accuracy in the calibration phase, we report four criteria: Log likelihoods, Akaike and Bayesian information criteria which penalize theories according to the number of free parameters,[22] and a pseudo-R^2 which is denoted ρ^2.[23] For the validation sample we report the log likelihood and also compute a mean squared deviation (MSD), which is defined as

$$MSD = \sum_{t=.7\cdot T+1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{m_i} \frac{[P_i^j(t) - I(s_i^j, s_i(t))]^2}{.3 \cdot T \cdot N \cdot m_i} \tag{15}$$

(Note that this MSD does not average observations across individuals.) Model fits are also compared to a random choice model in which all strategies are chosen equally often in each period. We do not compare results with Nash equilibrium because it does very poorly in constant sum games and beauty-contest games (in which iteratively-dominated strategies predicted to have zero probability are often played) and does not exclude any choices in the coordination games.

[19] If the attractions are not restricted in this way, then the experience weight $N(0)$ expresses both the relative weight on initial attractions and payoffs, and a scaling factor which puts attractions and payoffs on the same scale. By restricting attractions to have the same range as payoffs, we can then interpret $N(0)$ as a relative weight.

[20] In our working paper we allowed initial attractions to have arbitrary scale, which made MLE convergence slower and identification worse. Allowing arbitrary attractions helps reinforcement a bit in constant-sum games but does not help much in median-action and beauty-contest games.

[21] For the constant-sum games, with only twenty subjects per game, every pair of row and column players were excluded, giving 100 jackknife runs.

[22] The Akaike criterion (AIC) is $LL - k$ and the Bayesian criterion (BIC) is $LL - \frac{k}{2} \cdot log(M)$ where k is the number of degrees of freedom and M is the size of the calibration sample.

[23] The measure ρ^2 is the difference between the Akaike measure and the log likelihood of a model of random choices, normalized by the random-model log likelihood.

For each game, we describe the game and basic details of how the experiments were conducted. Then we compare models and discuss parameter estimates.

Table 1 previews and summarizes the results. Within each game and measure, other than LL and ρ^2, the best fit statistic is printed in italics and marked with an asterisk. In both the calibration and validation phases, EWA fits substantially better in four of six games; in two cases the belief models fit a little better. (If EWA was overfitting, it would do relatively better in calibration than in validation, but this isn't the case.) Belief models do better than reinforcement in constant-sum games and worse in the median-action game. In the beauty contest game, the belief model does worse than reinforcement during calibration and better during validation. The two-segment models generally fit a little better during both validation and calibration, but the improvement in fit over one-segment models is small.

[Table 1 about here]

5.2 Constant-sum games with dominated actions

We fit data from four constant-sum games: two are 4x4 (G1 and G3) and the other are 6x6 (G2 and G4) from Mookherjee and Sopher (1997). Tables 2a-2b show the payoff matrices.[24] The 4x4 games essentially collapse three of the undominated actions (actions 3-5) of the 6x6 games into a single action (action 3).

[Tables 2a-b about here]

Note that these games each have a weakly dominated action (action 4 in G1 and G3 and 6 in G2 and G4). Dominated actions are useful for model discrimination because belief-based models always predict these actions will be chosen (weakly) less frequently than dominant actions, whereas the arbitrary initial attractions allowed by EWA and choice reinforcement can allow frequent choices of dominated strategies.

All these games have a unique mixed strategy equilibrium which is symmetric (even though the games are not symmetric). In games G1 and G3, in equilibrium actions 1-4 are played with probabilities $\frac{3}{8}, \frac{2}{8}, \frac{3}{8}, 0$ respectively. In games G2 and G4, equilibrium proportions are $\frac{3}{8}, \frac{2}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, 0$ for actions 1-6.

Each game was played by 10 different pairs of subjects playing with the same partner

[24] The fractional payoffs (e.g., 2/3W), denote probabilistic chances of winning W. These present a complication for reinforcement models, including EWA– do you reinforce the actual payoff (which has a one-third chance of being zero if 2/3W is the payoff) or the expected payoff? We reinforce according to the expected payoff.

40 times. At the end of each period players were told their partner's choice and their own payoff. In games G1 and G2 a win paid 5 rupees; in games G3 and G4 the payoffs were doubled to 10 rupees. (A typical student's monthly room and board cost 600 rupees.)

We derived MLE parameter estimates using the first 28 periods, and validated by predicting the last 12 periods. Because the payoff matrix is not symmetric (even though the equilibrium mixed-strategy proportions are), we estimate separate initial attractions $A_i^j(0)$ and separate initial experience-weights $N_i^j(0)$ for row and column players (though we restrict the total experience weight $N(0)$ to be the same for both types of players). Tables 3a-b show the MLE parameter estimates of the models, and χ^2 tests of the belief and reinforcement restrictions (along with p-values and degrees of freedom). We report only the one-segment results because the two-segment results do not improve much and offer no special insights.

Tables 3a-b shows that for one-segment models, belief-based models and choice reinforcement restrictions are weakly and strongly rejected by χ^2 tests, respectively, in the calibration phase. In the validation phase, the reinforcement model is worst. The belief model is better than EWA in the four-strategy games G1 and G3, and worse in the six-strategy games G2 and G4. These differences are not large, however, and seem to be due to an idiosyncracy in game G1.[25]

[Tables 3a-b about here]

Tables 3a-b report parameter estimates and jackknifed standard errors. The initial conditions $\hat{A}^j(0)$ are encouragingly similar in pairs of low- and high-stakes games (G1-G3 and G2-G4), and put low initial attraction on the dominated strategies. The initial experience weight $\hat{N}(0)$ varies between about 10-20 and is close to its steady-state value of $\frac{1}{1-\rho}$. This means that initial reinforcements do not have much effect, which is reasonable given the slow convergence in these 40-period games. The decay parameters $\hat{\phi}$ and $\hat{\rho}$ are close to one, with $\hat{\phi} > \hat{\rho}$. These numbers imply that attractions grow only slightly on average. By forcing $\rho = 0$, in contrast, the reinforcement model forces attractions to grow and 'locks in' initial behavior too quickly. Finally, $\hat{\delta}$ is between .4 and .7 and significantly different from both zero and one, except in game G1 where it is estimated to be zero.

Notice how the EWA estimates reflect a hybridization of elements of reinforcement and belief learning. First, the initial EWA attractions place much less relative weight on the dominated strategies (the highest-numbered strategies 4 or 6) than the corresponding expected payoffs in belief models. In the belief model the gap between the initial expected

[25] In game G1, EWA overfits the first 28 periods because it detects some upward trend in strategies S1 and S3, and a downward trend in S2. These trends are reversed in the last 12 periods so EWA predicts poorly there. The belief model estimates differences in initial expected payoffs but has a huge value of $\hat{N}(0) = 300$, so it doesn't predict much movement at all.

payoffs of strategy 2 (the dominant strategies) and the dominated strategies cannot be too large because the strategies are only weakly dominated. For example, in game G1 the estimated EWA attractions on row strategies 2 and 4 are 1.14 and .00, while the corresponding estimated expected payoffs are 1.42 and .95, a gap less than half as large. Thus, EWA exploits the flexibility of initial attractions from reinforcement models to squash the likelihood of playing weakly dominated strategies further down than belief models can. Second, EWA borrows the belief-model property that attractions do not grow much, since $\hat{\phi}$ and $\hat{\rho}$ are very close. Third, the estimates of δ around .5 (except G1) reflect both the law of simulated effect ($\delta > 0$) and stronger effects of actual payoffs than foregone payoffs ($\delta < 1$).

Our conclusions about the relative performance of reinforcement and belief models are different from the findings of Mookerjhee and Sopher (1997), whose analysis differed in a couple of important ways.[26] Their version of reinforcement used 'average achieved earnings' rather than (weighted) cumulative earnings. The fact that $\hat{\phi}$ was very close to $\hat{\rho}$ in the EWA estimates indicates that MS took the right tack by using average earnings rather than cumulative earnings, because the cumulative-earnings assumption predicts a sharpness of convergence which is not evident in the data. However, their version of the belief model (which uses time-averaged expected payoffs) does not begin with an initial pre-game experience count expressing prior beliefs. Our estimates of $N(0)$ range from 30 to 300, which means that the belief model does best when it starts with a strong prior and updates very little. Thus, the difference between our results and theirs is primarily due to the fact that they use averaged reinforcements rather than cumulative ones (which improves reinforcement relative to our method), and they did not allow strong prior beliefs (which handicaps the belief model relative to our method).

Finally, notice that these constant-sum games do not distinguish models empirically very well. Coordination games, in which players converge quickly, may prove to be a better domain in which to distinguish theories.

5.3 Median-action games

We study median-action order statistic coordination games in which the group payoff depends on the median of all players' actions.[27] Table 4 shows the payoff matrix. Players earn a payoff which increases in the median, and decreases in the (squared) deviation from the median. The median-action games capture social situations in which conformity

[26] Their analysis used logit estimation of strategy choices to judge whether choices depended more strongly on a player's own average past earnings (a kind of choice reinforcement) or on expected earnings based on opponent's past history (fictitious play). They also compared models based on the entire previous history, weighting all observations equally, with models based on a five-period moving average. (The entire-history models fit better.) They allowed cross-effects so that the attraction $A_i^j(t)$ can affect other strategies differently, which is more general than our approach.

[27] Camerer and Ho (1997) also report estimates from 'weak-link' coordination games in which the group payoff depends on the minimum. The parameter estimates are similar to those reported here– for example, $\hat{\delta}$ is .65 and $N(0)$ is around two.

pressures induce people to behave like others do, but everyone prefers the group to choose a high median.

These median-action games were first studied experimentally by Van Huyck, Battalio and Beil (VHBB,1991), whose data we use.

[Table 4 about here]

We estimate EWA, choice reinforcement, and belief models using sessions 1-6 from VHBB (game γ). In their experiments groups of nine subjects each play ten periods together, so the sample has 54 subjects.[28] In each round players choose an integer from 1 to 7, inclusive. At the end of each round the median is announced (but not the full distribution of choices) 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 on the median and treating the group as a composite single player.

Figure 1a shows the actual frequencies across the six sessions, pooled together. Initial choices are concentrated around 4-5, with a dip at 6 and small spikes at 3 and 7. Later choices move sharply toward the initial medians, which were always 4 or 5. A striking feature, which is masked by pooling sessions, is that the 10th-round median in every session was equal to the first-round median. In three sessions the median began at 4 and stayed there; in the other three sessions the median began at 5 and stayed there.

¿From a learning point of view, median-action games are interesting because the penalty for deviating is fairly small if the players are close to equilibrium. Yet sharp convergence occurs within a couple of periods. Learning models which assume choices are reinforced must explain why players move quickly to equilibrium despite the large reinforcement if they are close to equilibrium and the small extra gain from moving precisely to equilibrium. The EWA model can account for this swift convergence if δ is close to one, which corresponds to the best-responsiveness inherent in belief learning.

[Figure 1a about here]

Table 5 shows estimation results for the median-action games. First we focus on one-segment results. EWA fits better than the reinforcement model ($\chi^2 = 64.8$) and much better than the belief model ($\chi^2 = 258.9$). The sources of EWA's improved fit are evident from looking at the data and plots of prediction errors.

---

[28] They compared two 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. There is no apparent or statistically-significant difference between these treatments so we pool them together.

[Table 5 about here]

Figure 1a shows that in the actual data, there are two large spikes in initial choices at 4-5, smaller spikes (about 15% of the observations) at 3 and 7, and few observations at 6. The estimated EWA initial attractions basically reflect this pattern in the data. The accuracy of the reflection can be judged from Figure 1b, an EWA error plot. This figure shows the difference between (MLE) predicted frequencies of the EWA model and the actual frequencies. The largest error is that EWA underpredicts the frequency of choices of 3 by about .06; predictions of 6 and 7 are too high by .03 and .01.

Reinforcement and belief learning cannot fit the initial conditions as well as EWA, but for different reasons. Reinforcement learning underpredicts the actual initial frequencies of 3 and 7 by about .08. Players who chose strategy 7 in the first period quickly switch to lower numbers in period 2, as Figure 1a shows. (The same is true for players who chose strategy 3, but this cannot be seen in Figure 1a.) Reinforcement learning cannot predict how quickly this convergence occurs. Since the initial medians are 4-5, choices of 3 or 7 earn between \$.55 and \$.95, while ex-post best responses earn \$1.00 to \$1.10. Since the initial choices are positively reinforced, reinforcement learning cannot explain why subjects will abandon these strategies so quickly and switch in the direction of the observed median. (EWA explains convergence with a high estimate of $\hat{\delta} = .85$.) Since choice reinforcement does not adjust chosen strategies quickly enough, to maximize overall fit it assumes the initial frequencies are close to frequencies in later periods, underpredicting choices of 3 and 7 (and overpredicting 4-5).

[Figures 1a-d about here]

Figure 1c shows that the belief model underpredicts 3 and 7 also, but for a different reason. In the belief-based framework it is hard to explain why players would play 6 less than the play 5 or 7. The problem is that initial beliefs which give a high expected payoff to 4-5 (expecting a median of 4-5) also give an expected payoff to 6 which is nearly as large, and larger than the expected payoff to 7. Beliefs which give a large expected payoff to 7, because there is large belief on a median of 7, will also give a high expected payoff to 6. Thus, it is difficult to find a single set of beliefs which can explain the spikes at 4-5 and 7, without also predicting a spike at 6. As a result, Table 5 shows that the one-segment model generates initial expected payoffs which are higher for 6 (\$.78) than for 3 or 7 (\$.71 and \$.60), so it overpredicts 6 and underpredicts 3 and 7 (and also overpredicts 5).

Adding a second segment of players improves the belief-model fit dramatically. As Table 5 shows, the log likelihood improves a lot (the χ^2 statistics for the two-segment results compare one- and two-segment fits within each model). The two belief-model segments correspond naturally to a large (78%) segment with high expected payoffs for 4-5 generated by high initial beliefs in 4-5, and a smaller (22%) segment with belief only

in 7, which generates the highest expected payoff for 7. While testing the restriction that the second segment does not improve fit rejects strongly ($\chi^2 = 119.0$), the two-segment belief model still does not fit as well as the one- or two-segment EWA model.

Besides fitting initial conditions, a good learning model must explain why convergence in the first couple of periods is fast and sharp. EWA does this by estimating a large value of δ (.85) and $\hat{\phi}$ much larger than $\hat{\rho}$, which allows attractions to grow rapidly so that choice probabilities move toward zero and one swiftly. The low value of $N(0)$, .65, also allows players to learn quickly from payoff reinforcement relative to initial attractions.

The estimates show how EWA mixes and matches the best features of belief and reinforcement learning: It allows near-best response because δ is close to one as in belief models, explaining why players choosing near-equilibrium strategies move quickly toward equilibrium. But as in reinforcement, it can allow arbitrary initial attractions, which explains the relative paucity of choices of 6 in the first period, and allows attractions to grow (because $\rho = 0$) to explain the sharpness of convergence. As a result, the EWA errors (Figure 1b) are generally much smaller than those in reinforcement (Figure 1c) and belief learning (Figure 1d).

The results shown in the error plots are for one-segment models. Adding a second segment does improve fits significantly for all three models. In EWA, the main difference in segments is that the larger segment (with frequency 66%) has an estimate $\hat{\delta} = .95$, very close to the belief restriction of one, while the smaller second segment has $\hat{\delta} = .50$. This corresponds to a segment of people with belief-type equal weighting of actual and foregone payoffs, and another segment who weight actual payoffs twice as heavily. Notice that these two segments do not particularly correspond to one segment of reinforcement learners and another segment of belief learners, so EWA is not simply capturing a mixture of these two special cases.

In reinforcement, the larger segment (80%) has parameter values which are similar to those in the single segment, except the estimates of initial attractions for 3 and 7 are zero. The smaller second segment (20%) is the opposite– strategies 3 and 7 have the largest possible initial attractions and all the others are close to zero– except that $\hat{\phi} = 0$.[29] This means the two-segment structure is trying to solve the problem of explaining first-period choices of 3 and 7 which are quickly extinguished by creating a second segment of players who choose only 3 or 7 initially, then immediately decay their initial attraction. But adding this segment does not improve log likelihood much and the two-segment reinforcement model still fits worse than the one-segment EWA model.

The two-segment belief model improves fit substantially, as noted above, but it still does not capture initial attractions flexibly enough (compared to EWA). We think the problem is that the belief model, as we define it, requires initial behavior to be consistent

[29]The estimate of zero for ϕ is the full-sample MLE estimate. The jackknifed standard error of .235 means that in many jackknife samples ϕ is estimated to be positive. Indeed, the mean of the jackknife estimates is .18, but this does not substantially affect the point we make in the text.

with prior beliefs and requires beliefs to be updated using weighted fictitious play. The latter assumption boils down to $\delta = 1$ and $\phi = \rho$. In games like the median-action game, the $\delta = 1$ assumption may be reasonable but $\phi = \rho$ does not allow sharp enough convergence.[30] More importantly, forcing initial attractions to spring from expected payoffs does not flexibly explain behavior of players who decision rules. For example, a player who randomizes among different selection principles will not necessarily choose according to expected payoffs given a prior.

5.4 Dominance-solvable p-beauty contest games

In a p-beauty contest game, n players simultaneously choose numbers x_i in some interval, say [0,100]. The average of their numbers $\bar{x} = \frac{\sum_i^n x_i}{n}$ is computed, which establishes a target number, τ, equal to $p \cdot \bar{x}$. The player whose number is closest to the target wins a fixed prize $n \cdot \pi$ (and ties are broken randomly[31]).

P-beauty contest games were first studied experimentally by Nagel (1995) and extended by Ho, Camerer and Weigelt (in press) and Duffy and Nagel (in press). These games are useful for estimating the number of steps of iterated dominance players use in reasoning through games. To illustrate, suppose $p = .7$. Since the target can never be above 70, any number choice above 70 is stochastically dominated by simply picking 70. Similarly, players who obey dominance, and believe others do too, will pick numbers below 49 so choices in the interval (49,100] violate the conjunction of dominance and one step of iterated dominance. The unique Nash equilibrium is 0.

There are two behavioral regularities in beauty contest games (see Nagel, in press, for a review). First, initial choices are widely dispersed and centered somewhere between the interval midpoint and the equilibrium. This basic result has been replicated with students on three continents and with several samples of sophisticated adults, including economics Ph.D.'s and a sample of CEOs and corporate presidents (see Camerer, 1997). Second, when the game is repeated, numbers gradually converge toward the equilibrium.

Explaining beauty contest convergence is a challenge for adaptive learning models. Standard choice reinforcement are likely to converge far too slowly, because only one player wins each period and the losers get no reinforcement. Belief models with low values of ϕ, which update beliefs very quickly, may track the learning process reasonably well, but earlier work suggests Cournot dynamics do not converge fast enough either (Ho et al, in press).

The three models were estimated on a subsample of data collected by Ho *et. al* (in

[30] The fact that $\hat{\rho} = 0$ in EWA (and never varies across the jackknife runs) also suggests that adding more segments to the belief model will not improve fit substantially compared to EWA models with the same number of segments, because the belief models are always constrained to have $\rho = \phi$.

[31] Formally, $\pi(x_i, x_{-i}) = \frac{n \cdot \pi \cdot I(x_i, argmin_{x_j} |x_j - \tau|)}{\sum_i I(x_i, argmin_{x_j} |x_j - \tau|)}$ where $I(x, y)$ is the indicator function that equals one if $x = y$ and 0 otherwise.

press). Subjects were 196 undergraduate students in computer science and engineering in Singapore. Each seven-person group of players played 10 times together twice, with different values of p in the two 10-period sequences. (One sequence used $p > 1$ and is not included below.) The prize was .5 Singapore dollars per player each time, about $2.33 per group for seven-person groups. They were publicly told the target number τ and privately told their own payoff (i.e., whether they were closest or not).

We analyze a subsample of their data with $p = .7$ and .9, from groups of size 7. This subsample combines groups in a 'high experience' condition (the game is the second one subjects play, following a game with a value of $p > 1$) and the 'low experience' condition (the game is the first they play). The experience conditions were pooled to create enough data to get reliable estimates.

Several design choices were necessary to implement the model. The subjects chose integers in the interval [0,100], a total of 101 strategies. If we allow 101 possible values of $A^j(0)$ we quickly use too many degrees of freedom estimating the initial attractions. Rather than imposing too many structural requirements on the distribution of $A^j(0)$, we assumed initial attractions were equal in ten-number intervals [0,9], [10, 19], etc.[32]

To implement EWA we assumed subjects knew the winning number, $w = argmin_{x_j}[|x_j - \tau|]$, and neglect the effect of their own choice on the target number.[33] Define the distance between the winning number and the target number as $d = |\tau - w|$. All subjects reinforced numbers in the intervals $(\tau - d, \tau + d)$ by δ times the prize, and numbers in the intervals $[0, \tau - d)$ and $(\tau + d, 100]$ received no reinforcement. Winners reinforced the boundary number they chose, either $\tau - d$ or $\tau + d$, by the prize divided by the number of winners, and reinforced the other boundary number by δ times the prize divided by the number of winners. Losers reinforced both boundary numbers $\tau - d$ and $\tau + d$ by δ times the prize, divided by the number of winners plus one.

Implementing the belief model is not straightforward because subjects were told only the target number, and whether they won, so they do not have enough information to form beliefs about what other subjects will do, and use these updated beliefs to calculate expected payoffs. Reinforcing numbers in some intervals, as in the EWA updating, will not necessarily correspond to belief learning based in information about all others' numbers (which they do not know anyway). As a result, we estimate a restricted form of EWA with belief-type parameters by setting $\delta = 1$, $\phi = \rho$, estimating initial belief counts in the ten-number intervals, and taking initial expected payoffs to be normalized belief counts multiplied by the prize. Numbers in the winning interval $(\tau - d, \tau + d)$ are

[32] In our working paper we assumed the distribution of the values of $A^j(0)$ came from a beta distribution but the basic results were not much different. We also tried fitting asymmetric triangular distributions, in which $A^{100}(0) = 0$, $A^{50}(0) = c$, $A^0(0) = b$, and $A^j(0)$ was piecewise linear between 0 and 50, and 50 and 100, with slopes $(c - b)/50$ and $-c/50$, respectively, and tried normal distributions but the basic results were unchanged.

[33] Since subjects were not told the winning number (unless their number won), the fact that we must assume they do to estimate the model could be considered a handicap for the EWA and belief-based models, and a possible advantage for choice reinforcement, which does not require this assumption.

reinforced by one times the prize. This corresponds to a special kind of belief learning in which players are learning what the target number will be and best-responding given their beliefs.

Table 1 reports overall results. Generally the fit is not very impressive; ρ^2 values are only around 7%. In the calibration sample, EWA is slightly better than reinforcement, which is better than the belief model. Out of sample, the belief model and EWA model are about equally good (and reinforcement is clearly worst); the belief model is slightly better on MSD and much worse in log likelihood than EWA.

Table 6 reports results of parameter estimates.

[Table 6 about here]

The EWA model seems to be fitting the data as best it can in an odd way: It assumes there is a general tendency to pick lower numbers which grows stronger over time. This can be seen in the initial attractions, which are largest for the lowest number intervals[34], even though the first-period choices are clustered around 40-49 (i.e., attraction category $A^5(0)$). Then the model assumes these initial attractions 'inflate' over time ($\hat{\phi} = 1.33$). The model is not capturing learning from experience well because lagged attractions are weighted heavily compared to payoff reinforcement ($\hat{N(0)}$ is 16.82), and the estimate $\hat{\delta}$ is small (.23).

Choice reinforcement uses the same ingredients– high initial attractions for lower numbers, inflated by $\hat{\phi} = 1.38$– but fits substantially worse because $N(0)$ is forced to be one and there is little reinforcement from direct payoffs (since most players lose and get nothing). The belief model, in contrast, fits best by assuming initial expected payoffs are highest for choices in the interval [40,49], responding to payoff experience strongly (δ is fixed at one), and decaying attractions fairly quickly ($\hat{N(0)} = 1.67$ and $\hat{\phi} = .40$).

The two-segment analysis of EWA improves calibration substantially, compared to the one-segment model, and improves on the validation log-likelihood modestly. The two-segment reinforcement and belief models add very little to fit, especially in validation.

The two EWA segments that emerge (not reported in Table 6) are interesting. The larger segment (66%) is very much like the one-segment EWA estimate: Estimated initial attractions increase for smaller-number intervals, $\hat{\phi}$ is 1.61, $\hat{\delta}$ is zero, and the experience weight $\hat{N(0)}$ is 16.83. The smaller segment (34%) is remarkably like the one-segment belief model estimate: Initial attractions are highest for choices in the middle interval [50,59], $\hat{\phi}$ and $\hat{\rho}$ are small and very close (.50 and .43), $\hat{\delta}$ is estimated to be 1.0, and $\hat{N(0)} = 1.76$.

[34]The exception is that attractions are high for the interval [90,100]. This is to account for the occasional outlying choices of 100, which are discussed at length in Ho et al (in press).

None of these models capture the nature of learning well. The reinforcement and one-segment EWA models simply pretend that the first period is like later periods and inflate initial attractions to gradually reproduce the latter-period data. Belief models converge too slowly. The problem is that all these models are adaptive, so they only use information about previous payoffs (including previous foregone payoffs). Adaptive models of this sort cannot account for learning when players sophisticatedly realize that other players are learning as well (cf. Milgrom and Roberts, 1991). Our earlier work (Ho et al, in press) showed that a fraction of players seem to 'iteratively best-respond' in the sense that they choose numbers which are not best responses to observed history (as in weighted fictitious play), but instead choose numbers which are best responses to anticipated best-responding by others. Because the belief and reinforcement models do not have this kind of sophistication, the hybrid EWA does not either. The main lesson from fitting the beauty-contest data is that more work is left to be done, by including sophistication in some parsimonious way.

5.5 Identification of parameters and model diagnostics

The results generally show that EWA fits better than either of the special cases, both adjusting for extra parameters and predicting out of sample. A further test for model specification is to ask whether there are regular corelations among the three added parameters, δ, $N(0)$, and ρ, and other parameters. Because the EWA model is highly nonlinear, it is possible that certain parameters covary so closely that it is difficult to identify them econometrically. (By definition, a nonidentified parameter could be dropped from the model without reducing fit.) It is easy to show algebraically that the parameters are identified, in the sense that for arbitrary data sets and MLE parameter estimates, no other vector of parameter values which fit equally well. However, it is possible that parameters are nearly non-identified in some data sets.

An easy way to check the severity of nonidentifiability is to compute correlations among parameter estimates across jackknife runs. Two parameters which cannot be disentangled will be perfectly correlated across runs. Low or modest correlations across runs indicate that parameters have detectably separate influences. By inspecting the intercorrelations of the three important added parameters we can check whether each parameter contributes to predictive power.

A good overall statistic is the mean absolute correlation of the estimates of a parameter with all the other parameters with which it might be misidentified. We exclude initial attractions and compute correlations among $\phi, \delta, \rho, N(0)$, and λ.

For δ the mean absolute correlation with the other parameters is .31, .39, and .23 across the constant-sum, median-action, and beauty-contest games. None of the correlations with a specific parameter are consistent in magnitude and sign across games. This indicates that δ is well-identified. The same statistics for $N(0)$ are .19, .22, and .32. The latter number excludes the correlation between $N(0)$ and ρ in the beauty-contest

game, which is nearly one because the declining-effect constraint is binding. [35] These figures show that $N(0)$ is well-identified too (except when the constraint binds). The mean absolute correlations for ρ are .48, .30, and .32 (the latter again excludes the high correlation with N(0)). These correlations are somewhat higher than for δ and $N(0)$, especially in constant-sum games, indicating possible identification problems. The most systematic large correlation is between ρ and ϕ, which have an average correlation of .88 in the constant-sum games, (and the correlations are nearly equal in all four games). They are also correlated .50 in the median-action game and uncorrelated (-.03) in the beauty contest game. This pattern of correlations is a hint that the two depreciation parameters may be fundamentally related, in some games, in a way we hope to explore in further research.

The fact that the intercorrelations among estimates are modest and unsystematic (with noted exceptions) confirm that the parameters added in EWA contribute separately to its fit. We can also ask whether adding these parameters helps solve identification problems which arise in the belief and reinforcement special cases. For the reinforcement model, λ and ϕ are correlated -.79, -.68, and .05 in the three classes of games. The large negative correlations arise because when ϕ is lower attractions decay more rapidly, so λ must be larger to magnify small differences in attractions into large differences in choice probabilities. (The same effect does not seem to happen across runs of the beauty-contest game, where $\hat{\phi}$ is 1.38 and none of the models captures learning well.) Therefore, it is difficult to identify separate influences of the two parameters. Adding ρ and $N(0)$ in the EWA model reduces the correlations between λ and ϕ in magnitude, to .15, -.40, and -.20, eliminating any possible identification problem.

In the belief model the only apparent identification problem is between $N(0)$ and ϕ, which are correlated .20, -.86 and .99 in the three games. When ρ is included in the EWA model, these correlations become .23, .31, and .99, so the identification problem is partly eliminated.

Overall, there are modest identification problems in all three models. Problems in the reinforcement and belief models are largely alleviated by introducing ρ, $N(0)$, and δ in EWA. These new parameters are fairly well identified, except for modest-to-strong correlation between ρ and ϕ in two of three games. EWA therefore solves minor identification problems in the simpler models at the expense of creating another minor one, which could be explored in further research.

[35] When the declining-effect constraint $N(0) \leq \frac{1}{1-\rho}$ is binding $N(0)$ and ρ are not identified separately. (The same is true in the belief model.) We regard this as a shred of evidence about the way in which parameters may vary systematically across classes of games (see Cheung and Friedman, 1997). It may be that dominance-solvable games in which observed strategy choices are constantly shifting location have this general property so the restriction $N(0) = \frac{1}{1-\rho}$ can be safely imposed.

6 Discussion and conclusion

We proposed a general 'experience-weighted attraction' (EWA) learning model in which the probability of choosing a strategy is determined by its relative attraction. A strategy's attractions are updated by weighting lagged attractions by the number of periods of 'experience-equivalence' they contain, adding the payoffs actually received or a fraction of the payoffs that would have been received, then normalizing by an experience weight.

We see the paper as making two basic contributions.

First, we show that belief learning is not fundamentally different from reinforcement learning; both are special examples of one general learning rule– EWA. By showing their common basis, EWA lays bare the essential components of reinforcement and belief learning, and shows how those components can be combined to make a better model.

Comparing choice reinforcement to EWA makes it clear that reinforcement assumes players ignore foregone payoffs, and attractions cumulate as quickly as possible. Comparing weighted fictitious play to EWA makes it clear that belief models assume initial attractions are consistent with prior beliefs, foregone and actual payoffs are equally reinforcing, and attractions are weighted averages of past attractions and payoffs.

Second, by estimating the more general EWA model, along with reinforcement and belief-learning restrictions, our study combines methodological strengths of earlier studies while avoiding weaknesses. All earlier studies did one or more of the following: Concentrated on only one or two models, focussed on one class of games, ignored player heterogeneity, restricted the generality of models, derived parameter values using methods which do not guarantee best-fits, or did not report inferential statistics testing relative fit. Our paper had none of these limits because we compared three general models, on three classes of games, allowed some heterogeneity, derived parameter values optimally, and reported both test statistics (adjusting for free parameters three ways) and out-of-sample predictive accuracy.

EWA fits better than the reinforcement models in all cases, and better than beliefs in most cases, both adjusting for degrees of freedom within-sample and in out-of-sample prediction. Belief models are more accurate than reinforcement in some games, and by some measures, and less accurate in others.

Because reinforcement and belief approaches place clear restrictions on parameter values, it is useful to describe specific findings by parameter estimates.

The foregone payoff weight δ is estimated to be .42 (averaging across the four constant-sum games) .85 in median-action games, and .23 in beauty contests. The raw average of these numbers, .50, suggests that players generally weight foregone payoffs about half as much as actual payoffs. This result incorporates the intuitions underlying both reinforcement (actual payoffs are stronger) and belief learning (foregone payoffs matter).

Put differently, players seem to obey both the law of actual effect and a corollary law of simulated effect.

In the three games, the decay parameters ϕ and ρ average 1.00 and .94, .80 and 0, and 1.33 and .94. The first two games indicate that sometimes attractions seem to be approximately averages (as in belief models) and other times they seem to cumulate as rapidly as possible (as in reinforcement). The value of ϕ above one in beauty contests, as discussed above, reflects a likely misspecification because the adaptive EWA model does not incorporate sophistication and hence learns too slowly (a shortcoming the belief and reinforcement models also share).

The initial experience weight $N(0)$ averages 15.80, .65 and 16.82. The large values in constant-sum and beauty-contest games imply that players learn slowly, because they give much more weight to lagged attractions than to payoffs. The low value of .65 in median-action games means players respond more strongly to payoffs, learning faster.

EWA also exploits the flexibility of initial attractions shared by reinforcement models, compared to belief models in which initial attractions must be expected payoffs based on some prior. This flexibility is particularly helpful in the coordination games.

The results show how EWA is able to 'gene-splice' the best features of belief and reinforcement learning while avoiding weaknesses. For example, in the median-action games players begin with dispersed choices that seem to reflect different selection principles, and converge quickly. Explaining this pattern well requires initial attractions which are flexible and cumulate (as in reinforcement), rather than belief-based initial attractions which are averages, but also requires players to respond strongly to foregone payoffs (as in belief learning).

The fact that parameter values vary widely across the data sets is not too surprising; other studies have found differences in parameter estimates across games (e.g., Chen and Tang, 1996; Erev and Roth, 1997). Furthermore, the parameters capture different features of the data– speed of learning and sharpness of convergence. Since these features are different across the games we consider, parameter values should differ. Nonetheless, our understanding of learning will not be complete until there is a theory of how parameter values depend on game structure and experimental conditions (see Cheung and Friedman, 1997, for important progress). These estimates, and others', provide raw material for such theorizing.

6.1 EWA extensions

There are many directions for future research.

Theorizing about the kinds of equilibria EWA learning rules converge to would be extremely useful. Progress might be made by restricting attention to special classes of EWA players (e.g., those with δ equal to zero, or one) in specific classes of games.

An empirical direction for further research is measurement of model parameters using psychological methods. For example, if δ is interpreted as attention to foregone payoffs from unchosen alternatives, then values of δ should correlate with direct measures of attention, such as the amount of time subject spend looking at different numbers in a payoff matrix (see Camerer et al, 1993). (In general, measuring attention to information provides a direct way to test theories which assume certain kinds of information are not used.[36]) Or if $N(0)$ is the number of pregame 'trials' a player simulates which form prior beliefs, then $N(0)$ should be related to the ratio of initial response times to later-period response times.

EWA will also have to be upgraded to cope with three modelling challenges– sophistication, imperfect payoff information, and specification of strategies– before it is generally applicable.

Incorporating sophistication is important because EWA players only use information about their opponents' past choices, ignoring information about payoffs of others. Using this information in an expanded learning rule which incorporates sophistication could help explain data like those from the beauty-contest games. Iterating sophistication might also link sophisticated-EWA to equilibrium theories like quantal-response equilibrium.

Incorporating imperfect payoff information is important because any general model should be able to explain learning in low-information environments, where players do not know everything about their own payoffs, opponents' strategies, etc. EWA can obviously be applied in these settings by fixing $\delta = 0$ (which means EWA can apply to any environment choice reinforcement applies to). A more general approach would use imperfect information in some other way, rather than just giving it zero weight.

Incorporating a richer specification of strategies is important because stage-game strategies are not always the most natural candidates for the strategies which players learn about. For example, players may learn about history-dependent repeated-game strategies or a wide variety of decision rules (like minimax, Nash equilibrium, or imitation; e.g., Stahl, 1997). Once a set of richer strategies is specified, of course, EWA can still model learning about those strategies. The open question, therefore, is what rules to specify a priori, and how a model can winnow down a very large set of possible rules as quickly as humans probably do.

Adding these difficult extensions to EWA, and a theory of first-period play to supply initial attractions, might eventually create a unified way to predict how people play games in the lab and, eventually, how they play outside as well.

[36]For example, choice reinforcement predicts that players do not use information other than their own payoff history. Experiments which vary the information subjects are given have shown this prediction is wrong (Mookerjhee and Sopher, 1994; cf. Van Huyck, Battalio and Rankin, 1996). Direct measures of attention provide a more direct test: if players look at foregone payoffs frequently, then reinforcement models have some explaining to do. Similarly, all adaptive models predict that players do not use information about others' payoffs; looking at those payoffs is evidence of sophistication.

References

[1] Anderson, Simon, de Palma, Andre, and Thisse, Jacques-Francois, Discrete Choice Theory of Product Differentiation, MIT Press, 1992.

[2] Arthur, Brian, 'Designing economic agents that act like human agents: A behavioral approach to bounded rationality,' AER Proceedings, May 1991, Vol 81: 353-359.

[3] Ben-Akiva, Moshe, and Lerman, Steven, Discrete Choice Analysis: Theory and Application to Travel Demand, MIT Press, 1985.

[4] Bereby-Meyer, Yoella, and Erev, Ido. 'On learning to become a successful loser: A comparison of alternative abstractions of learning processes in the loss domain,' Technion-Israel Institute of Technology working paper, 1997.

[5] Boylan, Richard T. and Mahmoud A. El-Gamal, 'Fictitious Play: A Statistical Study of Multiple Economic Experiments,' Games and Economic Behavior, 5, 205-222, 19xx.

[6] Brown, G. 'Iterative Solution of Games by Fictitious Play,' In Activity Analysis of Production and Allocation, New York: John Wiley & Sons, 1951.

[7] Borgers, Tilman and Sarin, Rajiv, 'Naive Reinforcement Learning With Endogenous Aspirations,' Texas A&M University Working Paper, 1996.

[8] Brandts, Jordi and Charles Holt. 'Naive Bayesian learning and adjustment to equilibrium in signalling games,' Journal of Economic Behavior and Organization, in press.

[9] Broseta, B. 'Estimation of an Adaptive Learning Model in Experimental Coordination Games: An ARCH(1) Approach,' University of Arizona Department of Economics working paper, 1995.

[10] Brown, G. 'Iterative Solution of Games by Fictitious Play,' In Activity Analysis of Production and Allocation, New York: John Wiley & Sons, 1951.

[11] Bush, R. and Mosteller, F. Stochastic Models for Learning. New York, Wiley, 1955.

[12] Camerer, Colin F. 'Progress in Behavioral Game Theory'. Journal of Economic Perspectives, Vol. 11, pp. 167-188, 1997.

[13] Camerer, Colin F. Experiments on Strategic Interaction. Manuscript in progress.

[14] Camerer, Colin F., and Teck-Hua Ho, 'EWA Learning in Games: Preliminary Estimates from Weak-Link Games'. In D. Budescu, I. Erev and R. Zwick (Eds.), Games and Human Behavior: Essays in Honor of Amnon Rapoport, 1997.

[15] Camerer, Colin F., and Teck-Hua Ho, 'EWA Learning in Games: Probability form, heterogeneity, and time variation', Journal of Mathematical Psychology, in press.

[16] Camerer, Colin F.; Eric Johnson; Sankar Sen; and Talia Rymon. 'Cognition and framing in sequential bargaining for gains and losses'. In K. Binmore, A. Kirman and P. Tani, Frontiers of Game Theory. MIT Press, 1993, pp. 27-48.

[17] Carnap, Rudolf. The Logical Foundations of Probability (2nd edition) Chicago: University of Chicago Press.

[18] Chen, Yan and Feng-Feng Tang, 'Learning and Incentive Compatible Mechanisms for Public Goods Provision,' University of Michigan, 1996.

[19] Chen, H., J. W. Friedman, and J. F. Thisse. 'Boundedly Rational Nash Equilibrium: A Probabilistic Choice Approach,' Games and Economic Behavior, in press.

[20] Cooper, David; Garvin, Susan; and Kagel, John. 'Signaling and adaptive learning in an entry limit pricing game,' RAND Journal of Economics, in press.

[21] Cournot, A. *Recherches sur les principes mathematiques de la theorie des richesses.* Translated into English by N. Bacon as Researches in the Mathematical Principles of the Theory of Wealth. London: Haffner, 1960.

[22] Crawford, V. P. 'Adaptive Dynamics in Coordination Games,' Econometrica, 63, 103-143, 1995.

[23] Crawford, V.P. and Bruno Broseta, 'What price coordination? Auctioning the right to play as a form of preplay communication.' American Economic Review, in press.

[24] Cross, J. G. A Theory of Adaptive Economic Behavior, New York/London Cambridge University Press, 1983.

[25] Davey, Graham C. L. and George Matchett. Unconditioned stimulus rehearsal and the retention and enhancement of differential 'fear' conditioning: Effects of trait and state anxiety. Journal of Abnormal Psychology, 1994, 103, 708-718.

[26] Duffy, John and Nagel, Rosemarie. 'On the Robustness of Behavior in Experimental Guessing Games,' Economic Journal, forthcoming.

[27] Erev, Ido and Roth, Alvin. 'Predicting How People Play Games: Reinforcement learning in experimental games with unique, mixed-strategy equilibria,' University of Pittsburgh Working Paper 1997.

[28] Eysenck, H.J The conditioning model of neurosis. Behavioral and Brain Sciences, 2, 155-199.

[29] Friedman, D. 'Evolutionary Games in Economics,' Econometrica, 59, 637-666, 1991.

[30] Fudenberg, D. and Levine, D. Theory of Learning in Games, forthcoming.

[31] Ho, Teck-Hua and Weigelt, Keith. 'Task Complexity, Equilibrium Selection, and Learning: An Experimental Study,' Management Science, 42, 659-679.

[32] Ho, Teck-Hua, Camerer, Colin, and Weigelt, Keith, 'Iterated Dominance and Iterated Best-response in p-Beauty Contests,' American Economic Review, in press.

[33] Holt, Debra. 'An empirical model of strategic choice with an application to coordination games,' Department of Economics, Queen's University, 1993.

[34] Tversky, Amos and Kahneman, Daniel. 'Advances in prospect theory: Cumulative representation of uncertainty,' Journal of Risk and Uncertainty, 5, 297-323.

[35] McAllister, Patrick H. 'Adaptive approaches to stochastic programming,' Annals of Operations Research, 1991, 30, 45-62.

[36] McKelvey, Richard D. and Palfrey, Thomas R. 'Quantal response equilibria for normal form games,' Games and Economic Behavior, 1995, 10, 6-38.

[37] McKelvey, Richard D. and Palfrey, Thomas R. 'Quantal response equilibria for extensive form games,' Caltech working paper 1996.

[38] Mookerjee, Dilip and Sopher, Barry 'Learning Behavior in an Experimental Matching Pennies Game' Games and Economic Behavior, 7, 62-91, 1994.

[39] Mookerjee, Dilip and Sopher, Barry 'Learning and Decision Costs in Experimental Constant-sum Games,' Games and Economic Behavior, 19, 97-132, 1997.

[40] Nagel, R. 'A review of beauty contest games,' In D. Budescu, I. Erev, and R. Zwick (Eds.), Games and Human Behavior: Essays in Honor of Amnon Rapoport, Dordrecht: Kluwer, in press.

[41] Roth, Alvin and Erev, Ido. 'Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term,' Games and Economic Behavior, 8, 164-212 (1995).

[42] Sarin, Rajiv. 'Learning through reinforcement: The Cross model,' Texas A&M Department of Economics working paper, 1995.

[43] Sarin, Rajiv, and Vahid, Farshid. 'Payoff assessments without probabilities: Incorporating 'similarity' among strategies'. Texas A&M Department of Economics working paper, September 1997.

[44] Smith, Adam. An Inquiry into the Nature and Causes of the Wealth of Nations. Indianapolis: Liberty Classics, 1981.

[45] Stahl, Dale O. 'Rule learning in symmetric normal-form games: Theory and evidence,' University of Texas Department of Economics working paper, November 1997.

[46] Tang, F. 'Anticipatory Learning in Two-person Games: An Experimental Study', University of Bonn Working Paper, 1996.

[47] Van Huyck, John, Battalio, Raymond, and Beil, Richard, 'Tacit Cooperation Games, Strategic Uncertainty, and Coordination Failure,' The American Economic Review, 1990.

[48] Van Huyck, John, Battalio, Raymond, and Beil, Richard, 'Strategic Uncertainty, Equilibrium Selection, and Coordination Failure in Average Opinion Games,' Quarterly Journal of Economics, 106, 885-909, 1991.

[49] Van Huyck, John, Battalio, Raymond, and Rankin, Frederick W. 'Selection Dynamics and Adaptive Behavior Without Much Information,' Texas A & M Department of Economics working paper, September 1996.

[50] Weibull, Jörgen W. 'What have we learned from evolutionary game theory so far?' Stockholm School of Economics Research Institute of Industrial Economics working paper no. 487, 1997.

7 Appendix

Proof of Proposition 1.

The EWA model rules are:

(e-1) Choose $A_i^j(0)$ for all i, j.

(e-2) Choose $N(0) \geq 0$ for all i.

(e-3) Set $A_i^j(t)$ for $t \geq 1$ according to:

$$A_i^j(t) = \frac{\phi \cdot A_i^j(t-1) \cdot N(t-1) + [\delta + (1-\delta) \cdot I(s_i^j, s_i(t))] \cdot \pi(s_i(t), s_{-i}(t))}{N(t)} \tag{16}$$

(e-4) Set $N(t)$ for $t \geq 1$ according to:

$$N(t) = 1 + \rho \cdot N(t-1) \tag{17}$$

The belief-model rules are:

(b-1) Choose initial 'observation-equivalents' $N_{-i}^k(0) \geq 0$ for all opponents $-i$, strategies k. Define $N(0) = \sum_{k=1}^{m_{-i}} N_{-i}^k(0)$. Define prior beliefs by $B_{-i}^k(0) = \frac{N_{-i}^k(0)}{N(0)}$.

(b-2) Define $N(t) = 1 + \rho \cdot N(t-1)$ for $t \geq 1$. Update beliefs according to

$$B_{-i}^k(t) \;=\; \frac{I(s_{-i}^k, s_{-i}(t)) + \rho \cdot N_{-i}^k(t-1)}{N(t)} \tag{18}$$

or

$$B_{-i}^k(t) \;=\; \frac{\rho \cdot B_{-i}^k(t-1) + \frac{I(s_{-i}^k, s_{-i}(t))}{N(t-1)}}{\frac{N(t)}{N(t-1)}} \tag{19}$$

(b-3) Define attraction by expected payoff,

$$E_i^j(t) = \sum_{k=1}^{m_{-i}} B_{-i}^k(t) \cdot \pi(s_i^j, s_{-i}^k) \tag{20}$$

(i) Rule (e-1) is satisfied by initial attraction $A_i^j(0)$ induced by initial beliefs $B(0)$ given by (b-1) and the expected payoff rule (b-3).

(ii) Rules (e-2) and (e-4) are satisfied by (b-1) and (b-2).

(iii) The key step is showing that the belief-updating rule (b-2) and expected-payoff rule (b-3) for determining attraction are consistent with (e-3).

Assume $\rho = \phi$ and $\delta = 1$. Then the EWA rule (e-3) rule becomes

$$A_i^j(t) = \frac{\phi \cdot A_i^j(t-1) \cdot N(t-1) + \pi(s_i(t), s_{-i}(t))}{1 + \phi \cdot N(t-1)}. \tag{21}$$

Suppose attraction is expected payoff, so substituting (b-3) into the restricted form of updating rule gives

$$A_i^j(t) = \frac{\phi \cdot \sum_{k=1}^{m_{-i}} B_{-i}^k(t-1) \cdot \pi(s_i^j, s_{-i}^k) \cdot N(t-1) + \pi(s_i(t), s_{-i}(t))}{1 + \phi \cdot N(t-1)} \tag{22}$$

Note that $B_{-i}^k(t-1) \cdot N(t-1) = N_{-i}^k(t-1)$. Then the payoff terms in the numerator may be collected to write

$$A_i^j(t) = \frac{\sum_{k=1}^{m_{-i}} [\phi \cdot N_{-i}^k(t-1) + I(s_{-i}^k, s_{-i}(t))] \cdot \pi(s_i^j, s_{-i}^k)}{1 + \phi \cdot N(t-1)} \tag{23}$$

or

$$A_i^j(t) = \sum_{k=1}^{m_{-i}} \pi(s_i^j, s_{-i}^k) \cdot \frac{\phi \cdot N_{-i}^k(t-1) + I(s_{-i}^k, s_{-i}(t))}{1 + \phi \cdot N(t-1)}. \tag{24}$$

By the belief updating rule (b-2), this is simply

$$A_i^j(t) = \sum_{k=1}^{m_{-i}} \pi(s_i^j, s_{-i}^k) \cdot B_{-i}^k(t) = E_i^j(t) \tag{25}$$

Hence, the updated attractions are expected payoffs given updated beliefs. QED

Proof of Proposition 2.

The choice reinforcement rules are:

(c-1) Choose $R_i^j(0)$ for all i, j.

(c-2) Set $R_i^j(t)$ for $t \geq 1$ according to:

$$R_i^j(t) = \phi \cdot R_i^j(t-1) + I(s_i^j, s_i(t)) \cdot \pi(s_i^j, s_{-i}(t)) \tag{26}$$

The proof works by substitution and algebra.

(i) Condition (c-1) follows from (e-1).

(ii) Assume $N(0) = 1$ and $\rho = 0$. Then by (e-4), $N(t) = 1$ for all $t \geq 1$.

(iii) Assume $\delta = 0$. With (ii), condition (e-3) then becomes condition (c-2):

$$A_i^j(t) = \phi \cdot A_i^j(t-1) + I(s_i^j, s_i(t)) \cdot \pi(s_i(t), s_{-i}(t)) \tag{27}$$

Hence, all three conditions (c-1) to (c-3) follow, so under these conditions $A_i^j(t) = R_i^j(t)$. QED