The Economic Journal, **118** (*January*), 37–59. © The Author(s). Journal compilation © Royal Economic Society 2008. Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

INDIVIDUAL DIFFERENCES IN EWA LEARNING WITH PARTIAL PAYOFF INFORMATION

Teck H. Ho, Xin Wang and Colin F. Camerer

We extend experience-weighted attraction (EWA) learning to games in which only the set of possible foregone payoffs from unchosen strategies are known, and estimate parameters *separately* for each player to study heterogeneity. We assume players estimate unknown foregone payoffs from a strategy, by substituting the last payoff actually received from that strategy, by clairvoyantly guessing the actual foregone payoff, or by averaging the set of possible foregone payoffs conditional on the actual outcomes. All three assumptions improve predictive accuracy of EWA. Individual parameter estimates suggest that players cluster into two separate subgroups (which differ from traditional reinforcement and belief learning).

Central to economic analysis are the twin concepts of *equilibrium* and *learning*. In game theory, attention has turned recently to the study of learning (partly due to an interest in which types of equilibria might be reached by various kinds of learning, e.g. Mailath, 1998). Learning should be of general interest in economics because strategies and markets may be adapting much of the time or non-equilibrium behaviour emerges, due to imperfect information, rationality limits of agents, trading asynchronies, and supply and demand shocks. Understanding more about how learning works can be helpful in predicting time paths of behaviour in the economy, and designing institutional rules which speed learning. In game theory, understanding initial conditions and how learning occurs might also supply us with tools to predict which of many equilibria will result when there are multiple equilibria (Crawford, 1995).

The models of learning in simple games described in this article are not meant to be applied directly to complex markets and macroeconomic processes. However, the hope is that by honing models sharply on experimental data (where we can observe the game structure and the players moves clearly), we can create robust models that could eventually be applied to learning in naturally-occurring situations, e.g., hyperinflations, as in Marcet and Nicolini (2003).

There are two general empirical approaches to understanding learning in games (Ho, forthcoming; Camerer, 2003, chapter 6):

Population models and individual models.

1 *Population models* make predictions about how the aggregate behaviour in a population will change as a result of aggregate experience. For example, in replicator dynamics, a population's propensity to play a certain strategy will depend on its 'fitness' (payoff) relative to the mixture of strategies played previously.[1] Models like this are obviously useful but submerge differences in individual learning paths.

[1] Another important class of models involve imitation (Schlag, 1999); still another is learning among various abstract decision rules (Stahl and Haruvy, 2004).

2 *Individual learning models* allow each person to choose differently, depending on the experiences they have. Our 'experience-weighted attraction' (EWA) model, for example, assumes that people learn by decaying experience-weighted lagged attractions, updating them according to received payoffs or weighted foregone payoffs, and normalising those attractions. Attractions are then mapped into choice probabilities using a logit rule. This general approach includes the key features of reinforcement and belief learning (including Cournot and fictitious play), and predicts behaviour well in many different games; see Camerer *et al.* (2002) for a comprehensive list.

In this article, we extend the applicability of EWA in two ways: by estimating learning rules at the individual level and modelling cases where the foregone payoff from unchosen strategies is not perfectly known (e.g., most extensive-form games).

First, we allow different players to have different learning parameters. In many previous empirical applications, players are assumed to have a common learning rule, exceptions include Cheung and Friedman (1997), Stahl (2000) and Broseta (2000).

Allowing heterogeneous parameter values is an important step for four possible reasons.

(*i*) While it seems very likely that detectable heterogeneity exists, it is conceivable that allowing heterogeneity does not improve fit much. If not, then we have some assurance that 'representative agent' modelling with common parameter values is an adequate approximation.

(*ii*) If players are heterogeneous, it is likely that players fall into distinct clusters, perhaps corresponding to familiar learning rules like fictitious play or reinforcement learning, or to some other kinds of clusters not yet identified.[2]

(*iii*) If players are heterogeneous, then it is possible that a single parameter estimated from a homogeneous representative-agent model will misspecify the mean of the distribution of parameters across individuals.[3] We can test for such a bias by comparing the mean of individual estimates with the single representative-agent estimate.

(*iv*) If players learn in different ways, the interactions among them can produce interesting effects. For example, suppose some players learn according to an adaptive rule and others are 'sophisticated' and know how the first group learn (e.g., Stahl, 1999). Then in repeated games, the sophisticated players have an incentive to 'strategically teach' the learners in a way that benefits the sophisticates (Chong *et al.*, 2006). Understanding how this teaching works requires an understanding of heterogeneity in learning.

[2] Camerer and Ho (1998) allowed two separate configurations of parameters (or 'segments') to see whether the superior fit of EWA was due to its ability to mimic a population mixture of reinforcement and belief learners but they found that this was clearly not so. The current study serves as another test of this possibility, with more reliable estimation of parameters for all players.

[3] Wilcox (2006) shows precisely such a bias using Monte Carlo simulation, which is strongest in a game with a mixed-strategy equilibrium but weaker in a stag-hunt coordination game. The strongest bias is that when the response sensitivity λ values are dispersed, then when a single vector of parameters is estimated for all subjects the recovered value of δ is severely downward-biased compared to its true value. He suggests random effects estimation of a distribution of λ values to reduce the bias.

Second, most theories of learning in games assume that players know the foregone payoffs to strategies they did not choose. Theories differ in the extent to which unchosen strategies are reinforced by foregone payoffs. For example, fictitious play belief learning theories are equivalent to generalised reinforcement theories in which unchosen strategies are reinforced according to their foregone payoffs as strongly as chosen strategies are. But then, as Vriend (1997) noted, how does learning occur when players are not sure what foregone payoffs are? This is a crucial question for applying these theories to naturally occurring situations in which the modeller may not know the foregone payoffs, or to extensive-form games in which players who choose one branch of a tree do not know what would have resulted if they chose another path. In this article we compare three ways to add learning about unknown foregone payoffs ('payoff learning') to describe learning in low-information environments.[4]

The basic results can be easily stated. We estimated individual-level EWA parameters for 60 subjects who played a normal-form centipede game (with extensive-form feedback) 100 times (Nagel and Tang, 1998). Parameters do differ systematically across individuals. While parameter estimates do not cluster naturally around the values predicted by belief or reinforcement models, they do cluster in a similar way in two different player roles, into learning in which attractions *cumulate* past payoffs, and learning in which attractions are *averages* of past payoffs.

Three payoff learning models are used to describe how subjects estimate foregone payoffs, then use these estimates to reinforce strategies whose foregone payoffs are not known precisely. All three are substantial improvements over the default assumption that these strategies are not reinforced at all. The best model is the one in which 'clairvoyant' subjects update unchosen strategies with perfect guesses of their foregone payoffs.

1. EWA Learning with Partial Payoff Information

1.1. *The Basic EWA Model*

Experience-weighted attraction learning was introduced to hybridise elements of reinforcement and belief-based approaches to learning and includes familiar variants of both as special cases. This Section will highlight only the most important features of the model. Further details are available in Camerer and Ho (1999) and Camerer *et al.* (2002).

In EWA learning, strategies have attraction levels which are updated according to either the payoffs the strategies actually provided, or some fraction of the payoffs unchosen strategies *would have* provided. These attractions are decayed or depreciated each period, and also normalised by a factor which captures the (decayed) amount of experience players have accumulated. Attractions to strategies are then mapped into the probabilities of choosing those strategies using a response function which guarantees that more attractive strategies are played more often.

[4] Ho and Weigelt (1996) studied learning in extensive-form coordination games and Anderson and Camerer (2000) studied learning in extensive-form signalling games but both did not consider the full range of models of foregone payoff estimation considered here.

EWA was originally designed to study n-person normal form games. The players are indexed by i $(i = 1, 2, \ldots, n)$, and each one has a strategy space $S_i = \{s_i^1, s_i^2, \ldots, s_i^{m_i-1}, s_i^{m_i}\}$, where s_i denotes a pure strategy of player i. The strategy space for the game is the Cartesian products of the S_i, $S = S_1 \times S_2 \times \ldots \times S_n$. Let $s = (s_1, s_2, \ldots, s_n)$ denote a strategy combination consisting of n strategies, one for each player. Let $s_{-i} = (s_1, \ldots, s_{i-1}, s_{i+1}, \ldots, s_n)$ denote the strategies of everyone but player i. The game description is completed with specification of a payoff function $\pi_i(s_i, s_{-i}) \in \Re$, which is the payoff i receives for playing s_i when everyone else is playing the strategy specified in the strategy combination s_{-i}. Finally, let $s_i(t)$ denote i's actual strategy choice in period t, and $s_{-i}(t)$ the vector chosen by all other players. Thus, player i's payoff in period t is given by $\pi_i[s_i(t), s_{-i}(t)]$.

1.2. *Updating Rules*

The EWA model updates two variables after each round. The first variable is the experience weight $N(t)$, which is like a count of 'observation-equivalents' of past experience and is used to weight lagged attractions when they are updated. The second variable is $A_i^j(t)$, i's attraction for strategy j *after* period t has taken place. The variables $N(t)$ and $A_i^j(t)$ begin with initial 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.[5]

Updating after a period of play is governed by two rules. First, experience weights are updated according to

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

where ρ is a discount factor that depreciates the lagged experience weight. 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 payoff actually received, from chosen strategy $s_i(t)$, by an additional $1 - \delta$ (so it receives 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 for i's jth strategy can be written $\{\delta + (1-\delta)I[s_i^j, s_i(t)]\}\pi_i[s_i^j, s_{-i}(t)]$. The rule for updating attraction sets $A_i^j(t)$ to be a depreciated, experience-weighted lagged attraction, plus an increment for the received or foregone payoff, normalised by the new experience weight. That is,

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

The factor ϕ is a discount factor that depreciates previous attractions. Let $\kappa = (\phi - \rho)/\phi$. Then the parameter κ adjusts whether the experience weight depreciates more rapidly than the attractions. Notice that the steady-state value of $N(t)$ is $1/(1-\rho)$ (and does not depend on $N(0)$). In the estimation we impose the restriction

[5] Boulding *et al.* (1999) and Biyalogorsky *et al.* (2006) show that managers tend to have a large initial experience count $N(0)$.

$N(0) \leq 1/(1-\rho)$ which guarantees that the experience weight rises over time, so the relative weight on new payoffs falls and learning slows down.

Finally, attractions must be mapped into the probabilities of choosing strategies in some way. Obviously we would like $P_i^j(t)$ to be monotonically increasing in $A_i^j(t)$ and decreasing in $A_i^k(t)$ (where $k \neq j$). Three forms have been used in previous research: A logit or exponential form, a power form, and a normal (probit) form. The various probability functions each have advantages and disadvantages. We prefer the logit form

$$P_i^j(t+1) = \frac{\mathrm{e}^{\lambda A_i^j(t)}}{\sum_{k=1}^{m_i} \mathrm{e}^{\lambda A_i^k(t)}} \tag{3}$$

because it allows negative attractions and fits a little better in a direct comparison with the power form (Camerer and Ho, 1998). The parameter λ measures sensitivity of players to differences among attractions. When λ is small, probabilities are not very sensitive to differences in attractions (when $\lambda = 0$ all strategies are equally likely to be chosen). As λ increases, it converges to a best-response function in which the strategy with the highest attraction is always chosen.

Bracht and Ichimura (2001) investigate the econometric identification of the EWA model and show that it is identified if the payoff matrix is regular (i.e., no two strategies receive the same payoff) and $\lambda \neq 0$, $|\rho N(0)| < \infty$ and $N(0) \neq 1 + \rho N(0)$. Consequently, we impose $\lambda > 0$, $0 \leq \rho < 1$, and $0 \leq N(0) < 1/(1-\rho)$ in our estimation.[6] In some other recent research, we have also found it useful to replace the free parameters for initial attractions, $A_i^j(0)$, with expected payoffs generated by a cognitive hierarchy model designed to explain choices in one-shot games and supply initial conditions for learning (Camerer *et al.*, 2002; Chong *et al.*, 2006).[7]

1.3. *Special Cases*

One special case of EWA is choice reinforcement models in which strategies have levels of reinforcement or propensity which are depreciated and incremented by received payoffs. In the model of Harley (1981) and Roth and Erev (1995), for example

$$R_i^j(t) = \begin{cases} \phi R_i^j(t-1) + \pi_i[s_i^j, s_{-i}(t)] & \text{if } s_i^j = s_i(t), \\ \phi R_i^j(t-1) & \text{if } s_i^j \neq s_i(t). \end{cases} \tag{4}$$

Using the indicator function, the two equations can be reduced to one:

$$R_i^j(t) = \phi R_i^j(t-1) + I[s_i^j, s_i(t)]\pi_i[s_i^j, s_{-i}(t)]. \tag{5}$$

[6] Salmon (2001) evaluates the identification properties of reinforcement, belief-based, and the EWA models by simulation analysis. He uses each of these models to generate simulated data in simple matrix games and investigate whether standard estimation methods can accurately recover the model. He shows that all models have difficulties in recovering the true model but the EWA model can identify its true parameters (particularly δ) more accurately than reinforcement and belief-based models.

[7] Another approach to reducing parameters is to replacing fixed parameters with 'self-tuning' functions of experience (Ho *et al.*, 2007). This model fits almost as well as one with more free parameters and seems capable of explaining cross-game differences in parameter values.

This updating formula is a special case of the EWA rule, when $\delta = 0$, $N(0) = 1$, and $\kappa = 1$. The adequacy of this simple reinforcement model can be tested empirically by setting the parameters to their restricted values and seeing how much fit is compromised (adjusting, of course, for degrees of freedom).

In another kind of reinforcement, attractions are *averages* of previous attractions, and reinforcements, rather than cumulations (Sarin and Vahid, 2004; Mookerjhee and Sopher, 1994, 1997; Erev and Roth, 1998). For example

$$R_i^j(t) = \phi R_i^j(t-1) + (1-\phi)I[s_i^j, s_i(t)]\pi_i[s_i^j, s_{-i}(t)]. \tag{6}$$

A little algebra shows that this updating formula is a special case of the EWA rule, when $\delta = 0$, $N(0) = 1/(1-\phi)$, and $\kappa = 0$.

In belief-based models, adaptive players base their responses on beliefs formed by observing their opponents' past plays. While there are many ways of forming beliefs, we consider a fairly general 'weighted fictitious play' model, which includes fictitious play (Brown, 1951; Fudenberg and Levine, 1998) and Cournot best-response (Cournot, 1960) as special cases.

In weighted fictitious play, players begin with prior beliefs about what the other players will do, which are expressed as ratios of counts to the total experience. Denote total experience by $N(t) = \sum_{k=1}^{m_{-i}} N_{-i}^k(t)$.[8] Express the probability that others will play strategy k as $B_{-i}^k(t) = N_{-i}^k(t)/N(t)$, with $N_{-i}^k(t) \geq 0$ and $N(t) > 0$.

Beliefs are updated by depreciating the previous counts by ϕ, and adding one for the strategy combination actually chosen by the other players. That is,

$$B_{-i}^k(t) = \frac{\phi N_{-i}^k(t-1) + I[s_{-i}^k, s_{-i}(t)]}{\sum_{h=1}^{m_{-i}} \{\phi N_{-i}^h(t-1) + I[s_{-i}^h, s_{-i}(t)]\}}. \tag{7}$$

This form of belief updating weights the belief from one period ago ϕ times as much as the most recent observation, so ϕ can be interpreted as how quickly previous experience is discarded.[9] When $\phi = 0$ players weight only the most recent observation (Cournot dynamics); when $\phi = 1$ all previous observations count equally (fictitious play).

Given these beliefs, we can compute expected payoffs in each period t,

$$\mathrm{E}_i^j(t) = \sum_{k=1}^{m_{-i}} B_{-i}^k(t)\pi(s_i^j, s_{-i}^k). \tag{8}$$

The crucial step is to express period t expected payoffs as a function of period $t-1$ expected payoffs. This yields:

$$\mathrm{E}_i^j(t) = \frac{\phi N(t-1)\mathrm{E}_i^j(t-1) + \pi[s_i^j, s_{-i}(t)]}{\phi N(t-1) + 1}. \tag{9}$$

[8] Note that $N(t)$ is not subscripted because the count of frequencies is assumed, in our estimation, to be the same for all players. Obviously this restriction can be relaxed in future research.

[9] Some people interpret this parameter as an index of 'forgetting' but this interpretation is misleading because people may recall the previous experience perfectly (or have it available in 'external memory' on computer software) but they will deliberately discount old experience if they think new information is more useful in forecasting what others will do.

Expressing expected payoffs as a function of lagged expected payoffs, the belief terms disappear into thin air. This is because the beliefs are only used to compute expected payoffs, and when beliefs are formed according to weighted fictitious play, the expected payoffs which result can also be generated by generalised reinforcement according to previous payoffs. More precisely, if the initial attractions in the EWA model are expected payoffs given some initial beliefs (i.e., $A_i^j(0) = E_i^j(0)$), $\kappa = 0$ (or $\phi = \rho$), and foregone payoffs are weighted as strongly as received payoffs ($\delta = 1$), then EWA attractions are *exactly* the same as expected payoffs.

This demonstrates a close kinship between reinforcement and belief approaches. Belief learning is nothing more than generalised attraction learning in which strategies are reinforced equally strongly by actual payoffs and foregone payoffs, attractions are weighted averages of past attractions and reinforcements, and initial attractions spring from prior beliefs.[10]

1.4. *Interpreting EWA*

The EWA parameters can be given the following psychological interpretations.

1 The parameter δ measures the relative weight given to foregone payoffs, compared to actual payoffs, in updating attractions. It can be interpreted as a kind of counterfactual reasoning, 'imagination' of foregone payoffs, or responsiveness to foregone payoffs (when δ is larger players move more strongly toward *ex post* best responses).[11] We call it 'consideration' of foregone payoffs.

2 The parameter ϕ is naturally interpreted as depreciation of past attractions, $A(t)$. In a game-theoretic context, ϕ will be affected by the degree to which players realise other players are adapting, so that old observations on what others did become less and less useful. Then ϕ can be interpreted as an index of (perceived) change.

3 The parameter κ determines the growth rate of attractions, which in turn affects how sharply players converge. When $\kappa = 1$ then $N(t) = 1$ (for $t > 0$) and the denominator in the attraction updating equation disappears. Thus, attractions cumulate past payoffs as quickly as possible. When $\kappa = 0$, attractions are weighted averages of lagged attractions and past payoffs, where the weights are $\phi N(0)$ and 1.

   In the logit model, whether attractions cumulate payoffs, or average them, is important because only the *difference* among the attractions matters for their relative probabilities of being chosen. If attractions can grow and grow, as they can when $\kappa = 1$, then the differences in strategy attractions can be very large. This implies that, for a fixed response sensitivity, λ, the probabilities can be spread farther apart; convergence to playing a single strategy almost all the time can be sharper. If attractions cannot grow outside of the payoff bounds, when $\kappa = 0$ then

[10] Hopkins (2002) compares the convergence properties of reinforcement and fictitious play and finds that they are quite similar in nature and that they will in many cases have the same asymptotic behaviour.

[11] The parameter δ may also be related to psychological phenomena like regret. These interpretations also invite thinking about the EWA model as a two-process model that splices basic reinforcement, perhaps encoded in dopaminergic activity in the midbrain and striatum, and a more frontal process of imagined reinforcement. In principle these processes could be isolated using tools like eyetracking and brain imaging.

convergence cannot produce choice probabilities which are so extreme. Thus, we think of κ as an index of the degree of *commitment* to one choice or another (it could also be thought of as a convergence index, or confidence).

4 The term $A_i^j(0)$ represents the initial attraction, which might be derived from some analysis of the game, from selection principles or decision rules, from surface similarity between strategies in the game being played and strategies which were successful in similar games etc. Belief models impose strong restrictions on $A_i^j(0)$ by requiring initial attractions to be derived from prior beliefs.[12] Additionally, they require attraction updating with $\delta = 1$ and $\kappa = 0$. EWA allows one to separate these two processes: players could have arbitrary initial attractions but begin to update attractions in a belief-learning way after they gain experience.

5 The initial-attraction weight $N(0)$ is in the EWA model to allow players in belief-based models to have an initial prior which has a strength (measured in units of actual experience). In EWA, $N(0)$ is therefore naturally interpreted as the strength of initial attractions, relative to incremental changes in attractions due to actual experience and payoffs. If $N(0)$ is small then the effect of the initial attractions wears off very quickly (compared to the effect of actual experience). If $N(0)$ is large then the effect of the initial attractions persists.[13]

In previous research, the EWA model has been estimated on several samples of experimental data, and estimates have been used to predict *out-of-sample*. Forecasting out-of-sample completely removes any inherent advantage of EWA over restricted special cases because it has more parameters. Indeed, if EWA fits well in-sample purely by overfitting, the overfitting will be clearly revealed by the fact that predictive accuracy is much worse predicting out-of-sample than fitting in-sample.

Compared to the belief and reinforcement special cases, EWA fits better in weak-link coordination games – e.g. Camerer and Ho (1998), where out-of-sample accuracy was not measured – and predicts better out-of-sample in median-action coordination games and dominance solvable 'p-beauty contests' (Camerer and Ho, 1999), call markets (Hsia, 1998),'unprofitable games' (Morgan and Sefton, 2002), partially-dominance-solvable R&D games (Rapoport and Almadoss, 2000), and in unpublished estimates we made in other 'continental divide' coordination games (Van Huyck *et al.*, 1997). EWA only predicted worse than belief learning in some constant-sum games (Camerer and Ho, 1999), and has never predicted worse than reinforcement learning.

To help illustrate how EWA hybridises features of other theories, Figure 2 shows a three-dimensional parameter space – a cube – in which the axes are the parameters δ, ϕ, and κ. Traditional belief and reinforcement theories assume that learning parameters are located on specific edges of the cube. For example, cumulative reinforcement

[12] This requires, for example, that weakly dominated strategies will always have (weakly) lower initial attractions than dominant strategies. EWA allows more flexibility. For example, players might choose randomly at first, choose what they chose previously in a different game, or set a strategy's initial attraction equal to its minimum payoff (the minimax rule) or maximum payoff (the maximax rule). All these decision rules generate initial attractions which are not generally allowed by belief models but are permitted in EWA because $A_i^j(0)$ are flexible.

[13] This enables one to test equilibrium theories as a special kind of (non)-learning theory with $N(0)$ very large and initial attractions equal to equilibrium payoffs.

theories require low consideration ($\delta = 0$) and high commitment ($\kappa = 1$). (Note that the combination of low consideration and high commitment may be the worst possible combination, since such players can get quickly locked in to strategies which are far from best responses.) Belief models are represented by points on the edge where consideration is high ($\delta = 1$) but commitment is low ($\kappa = 0$). This constrains the ability of belief models to produce sharp convergence, in coordination games for example (Camerer and Ho, 1998, 1999). Cournot best-response and fictitious play learning are vertices at the ends of the belief-model edge.[14]

It is worth noting that fictitious play was originally proposed by Brown (1951) and Robinson (1951) as a computational procedure for finding Nash equilibria, rather than a theory of trial-by-trial learning. Cournot learning was proposed about 160 years ago before other ideas were suggested. Models of reinforcement learning were developed later, and independently, to explain behaviour of animals who presumably lacked higher-order cognition to imagine or estimate foregone payoffs. They were introduced into economics by John Cross in the 1970s and Brian Arthur in the 1980s to provide a simple way to model bounded rationality. Looking at Figure 2, however, one is hard pressed to think of an empirical rationale why players' parameter values would necessarily cluster on those edges or vertices which correspond to fictitious play or reinforcement learning (as opposed to other areas, or the interior of the cube). In fact, we shall see below that there is no prominent clustering in the regions corresponding to familiar belief and reinforcement models, but there is substantial clustering near the faces where commitment is either low ($\kappa = 0$) or high ($\kappa = 1$).

1.5. *EWA Extensions to Partial Payoff Information*

In this paper, partial foregone payoff information arises because we study a reduced normal-form centipede game but with extensive-form feedback (see Table 1 and Figure 1). In this game, an Odd player has the opportunity to take the majority of a growing 'pie' at odd numbered decision nodes $\{1, 3, 5, 7, 9, 11, 13\}$; the Even player has the opportunity to take at nodes $\{2, 4, 6, 8, 10, 12, 14\}$. Each player chooses when to take by choosing a number. The lower of the two numbers determines when the pie stops growing and how much each player gets. The player who chooses the lower number always gets more. Players receive feedback about their payoffs and not the other's strategy. Consequently, the player who chooses to take earlier cannot infer the other player's strategy from observing the payoffs because the game is non-generic in the sense that multiple outcomes lead to the same payoffs (see Table 1).

Our approach to explaining learning in environments with partial payoff information is to assume that players form some guess about what the foregone payoff might be, then plug it into the attraction updating equation. This adds no free parameters to the model.

First define the estimate of the foregone payoff as $\hat{\pi}_i(s_i^j, t)$ (and $\hat{\pi}$ is just the known foregone payoff when it is known). Note that $\hat{\pi}_i(s_i^j, t)$ does *not* generally depend on $s_{-i}(t)$ because, by definition, if the other players' strategy was observed then the foregone

[14] Note that EWA learning model has not been adapted to encompass imitative learning rules such as those studied by Schlag (1999). One way to allow this to allow other payoffs to enter the updating of attractions.

Table 1

Payoffs in Centipede Games, Nagel and Tang (1998)

Odd player number choices	Even player number choices						
	2	4	6	8	10	12	14
1	4	4	4	4	4	4	4
	1	1	1	1	1	1	1
3	2	8	8	8	8	8	8
	5	2	2	2	2	2	2
5	2	3	16	16	16	16	16
	5	11	4	4	4	4	4
7	2	3	6	32	32	32	32
	5	11	22	8	8	8	8
9	2	3	6	11	64	64	64
	5	11	22	45	16	16	16
11	2	3	6	11	22	128	128
	5	11	22	45	90	32	32
13	2	3	6	11	22	44	256
	5	11	22	45	90	180	64

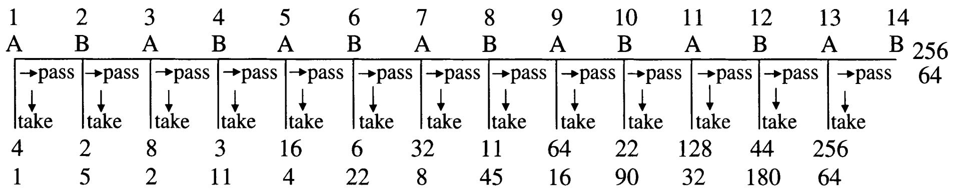

Fig. 1. *The Extensive Form of Centipede Game,* Nagel and Tang (1998)

payoff would be known. When the foregone payoff is known, updating is done as in standard EWA. When the foregone payoff is not known, updating is done according to

$$N_i^j(t) = \rho N_i^j(t-1) + 1, \quad t \geq 1 \tag{10}$$

and

$$A_i^j(t) = \frac{\phi N_i^j(t-1)A_i^j(t-1) + \{\delta + (1-\delta)I[s_i^j, s_i(t)]\}\hat{\pi}_i(s_i^j, t)}{N_i^j(t)}. \tag{11}$$

Three separate specifications of $\hat{\pi}(s_i^j, t)$ are tested: last actual payoff updating, payoff clairvoyance and the average payoff in the set of possible foregone payoffs conditional on the actual outcome. When players update according to the last actual payoff, they recall the last payoff they actually received from a strategy and use that as an estimate of the foregone payoff. Formally,

$$\hat{\pi}_i(s_i^j, t) = \begin{cases} \pi_i[s_i^j, s_{-i}(t)] & \text{if } s_i^j = s_i(t), \\ \hat{\pi}_i(s_i^j, t-1) & \text{otherwise.} \end{cases} \tag{12}$$

To complete the specification, the estimates $\hat{\pi}_i(s_i^j, 0)$ are initialised as the average of all the possible elements of the set of foregone payoffs.

Let us illustrate how this payoff learning rule works with the Centipede game given in Table 1 and Figure 1. Suppose player A chooses 7 and player B chooses 8 or higher.

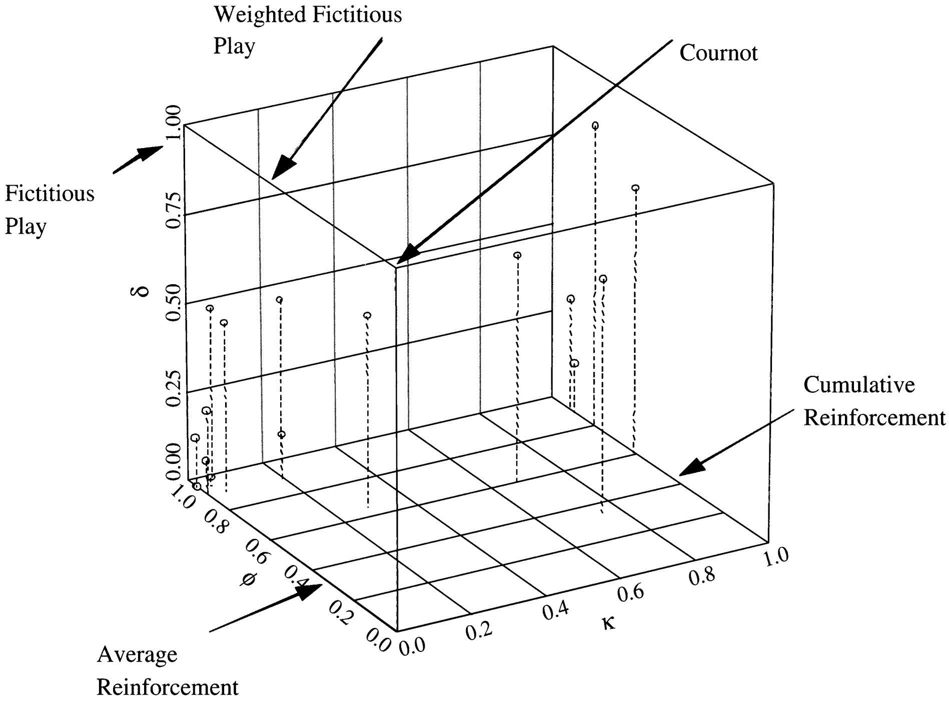

Fig. 2. *EWA's Model Parametric Space*

Since player A 'took first' she receives a payoff of 32, and she knows that if she chose 9 instead, she would receive either 11, if player B chose 8, or 64 if player B chose 10, 12, or 14. In this case we would initialise $\hat{\pi}_i(9,0) = (11+64)/2$. Notice that we average only the unique elements of the payoff set, not each payoff associated with every strategy pair. That is, even though 64 would result if player A chose 8 and B chose 10, 12, or 14, we only use the payoff 64 once, not three times, in computing the initial $\hat{\pi}_i(9,0)$.

Updating using the last actual payoff is cognitively economical because it requires players to remember only the last payoff they received. Furthermore, it enables them to adjust rapidly when other players' behaviour is changing, by immediately discounting all previous received payoffs and focusing on only the most recent one.

If one thinks of the last actual payoff as an implicit forecast of what payoff is likely to have been the 'true' foregone one, then it may be a poor forecast when the last actual payoff was received many periods ago, or if subjects have hunches about which foregone payoff they would have got which are more accurate than distant history. Therefore, we consider an opposite assumption as well – 'payoff clairvoyance'. Under payoff clairvoyance, $\hat{\pi}_i(s_i^j, t) = \pi_i[s_i^j, s_{-i}(t)]$. That is, players accurately guess exactly what the foregone payoff would have been even though they were not told about this information.

Finally, an intermediate payoff learning rule may be is to use the average payoff of the set of possible foregone payoffs conditional on the actual outcome to estimate the foregone payoff in each period. It is the same as the way we initialise the last actual payoff rule but apply the same rule in every period. Like before, we average only the unique elements in the payoff set.

The last-actual-payoff scheme recalls only observed history and does not try to improve upon it (as a forecast); consequently, it can also be applied when players do not even know the set of possible foregone payoffs. The payoff-clairvoyance scheme uses knowledge which the subject is not told (but could conceivably figure out). The average payoff rule lies between these two extreme. We report estimates and fit measures for the three models.

2. Data

Nagel and Tang (1998) (NT) studied learning in the reduced normal-form of an extensive-form centipede game. Table 1 shows the payoffs to the players from taking at each node. (Points are worth 0.005 deutschemarks.) They conducted five sessions with 12 subjects in each, playing 100 rounds in a random-matching fixed-role protocol. A crucial design feature is that while the players choose normal-form strategies, they are given extensive-form feedback. That is, each pair of subjects is only told the *lower* number chosen in each round, corresponding to the time at which the pie is taken and the game stops. The player choosing the lower number *does not* know the higher number. For example, if Odd chooses 5, takes first, and earns 16, she is not sure whether she would have earned 6 by taking later, at node 7 (if Even's number was 6) or whether she would have earned 32 (if Even had taken at 8 or higher), because she only knows that Even's choice was higher than 5. This ambiguity about foregone payoffs is an important challenge for implementing learning models.

Table 2 shows the overall frequencies of choices (pooled across the five sessions, which are similar). Most players choose numbers from 7 to 11.

If a subject's number was the lower one (i.e., they chose 'take'), there is a strong tendency to choose the same number, or a *higher* number, on the next round. This can be seen in the transition matrix Table 3, which shows the relative frequency of choices in round $t + 1$ as a function of the choice in round t, for players who 'take' in round t (choosing the lower number). For example, the top row shows that when players choose 2 and take, they choose 2 in the next round 28% of the time, but 8% choose 4 and 32% choice 6, which is the median choice (and is italicised). For choices below 7, the median choice in the next period is always higher. The overall tendency for players who chose 'take' to choose numbers which increase, decrease, or are unchanged are

Table 2

Relative Frequencies (%) Choices in Centipede Games, Nagel and Tang (1998)

Odd numbers	%	Even numbers	%
1	0.5	2	0.9
3	1.6	4	1.7
5	5.4	6	11.3
7	26.1	8	33.1
9	33.1	10	31.1
11	22.5	12	14.3
13	10.8	14	7.7

Table 3

Transitions after Lower-Number (Take) Choices, Nagel and Tang (1998)

	Choices in period $t + 1$ after 'Take'							
choice in t	2	4	6	8	10	12	14	Total no.
2	0.28	0.08	*0.32*	0.08	0.12	0.04	0.08	25
4	0.11	0.11	*0.40*	0.15	0.15	0.06	0.02	47
6		0.05	0.32	*0.41*	0.14	0.06	0.01	296
8		0.01	0.05	*0.56*	0.36	0.02	0.01	594
10			0.01	0.12	*0.73*	0.14	0.01	353
12			0.03	0.05	0.07	*0.83*	0.02	59
	1	3	5	7	9	11	13	Total no.
1	0.07	0.29	*0.21*	0.07	0.21	0.07	0.07	14
3	0.04	0.09	*0.44*	0.13	0.18	0.09	0.02	45
5	0.01	0.06	0.20	*0.47*	0.15	0.08	0.03	156
7		0.01	0.04	*0.60*	0.28	0.07		617
9			0.01	0.08	*0.62*	0.26	0.03	545
11					0.17	*0.60*	0.23	173
13						0.09	*0.91*	46

shown in Figure 3*a*. Note that most 'takers' then choose numbers which increase, but this tendency shrinks over time.

Table 4 shows the opposite pattern for players who choose the higher number and 'pass' – they tend to choose lower numbers. In addition, as the experiment progressed this pattern of transitions became weaker (more subjects do not change at all), as Figure 3*a* shows.

NT consider several models. Four are benchmarks which assume no learning: Nash equilibrium (players pick 1 and 2), quantal response equilibrium (McKelvey and Palfrey, 1995), random play and an individual observed-frequency model which uses each player's observed frequencies of choices over all 100 rounds. NT test choice-reinforcement of the Harley-Roth-Erev RPS type and implement a variant of weighted fictitious play which assumes players know population history information. The equilibrium and weighted fictitious play predictions do not fit the data well. This is not surprising because both theories predict either low numbers at the start, or steady movement toward lower numbers over time, which is obviously not present in the data. QRE and random guessing do not predict too badly, but the individual-frequency benchmark is the best of all. The RPS (reinforcement) models do almost as well as the best benchmark.

3. Estimation Methodology

The method of maximum likelihood was used to estimate model parameters. To ensure model identification as described in Section 1.2, we impose the necessary restrictions on the parameters $N(0)$, ρ, δ and λ in our estimation procedure.[15] We used

[15] Specifically, we apply an appropriate transformation to ensure each of the parameters will always fall within the restricted range. For example, we impose $\lambda = \exp(q_1)$ to guarantee that $\lambda > 0$, Although q_1 is without restriction, the parameter λ will always be positive. Similarly, we apply a logistic transformation, i.e. $\rho = 1/[1 + \exp(q_2)]$ and $\delta = 1/[1 + \exp(q_3)]$ to restrict ρ and δ to be between 0 and 1. Finally, $N(0) = [1/(1 - \rho)]/[1 + \exp(q_4)]$ so that $N(0)$ is between 0 and $1/(1 - \rho)$.

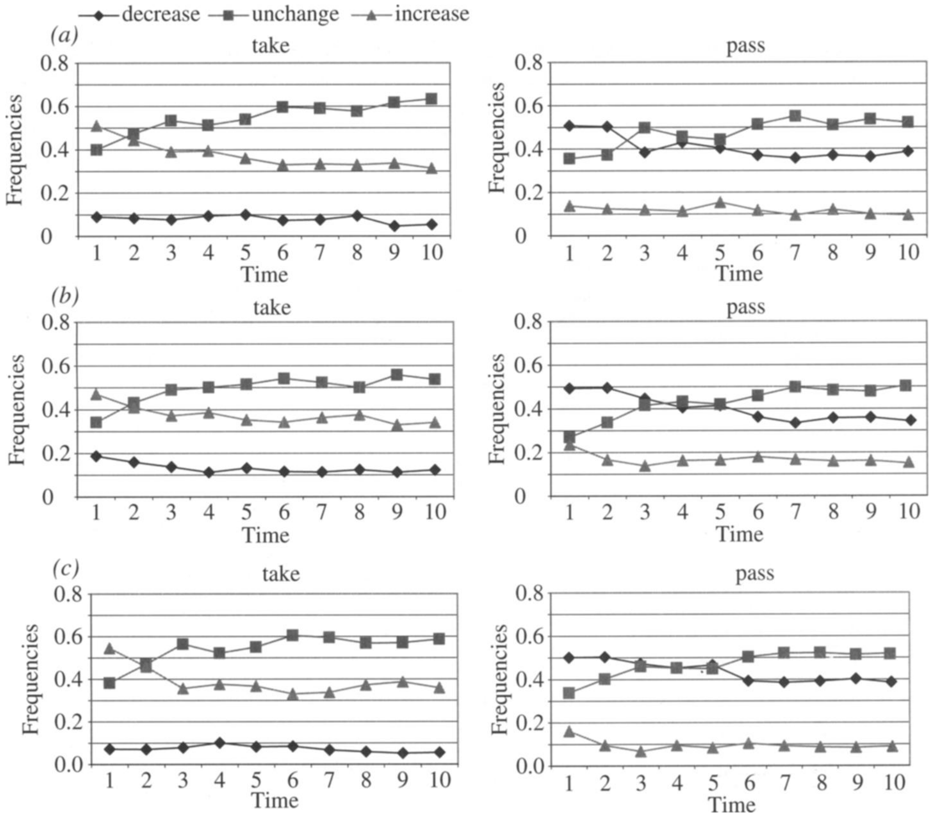

Fig. 3. *Transition Behaviour. (a) Actual Data; (b) EWA-Payoff Clairvoyance (Representative Agent Model); (c) EWA-Payoff Clairvoyance (Individual Model)*

the first 70% of the data to calibrate the models and the last 30% of the data to predict out-of-sample. Again, the out-of-sample forecasting completely removes any advantage more complicated models have over simpler ones which are special cases.

We first estimated a homogeneous single-representative agent model for reinforcement, belief, and three variants of EWA payoff learning. We then estimated the EWA models at the individual level for all 60 subjects. In the centipede game, each subject has seven strategies, numbers 1, 3, ...,13 for Odd subjects and 2, 4, ...,14 for even subjects. Since the game is asymmetric, the models for Odd and Even players were estimated separately. The log of the likelihood function for the single-representative agent EWA model is

$$LL[\delta, \phi, \kappa, \lambda, N(0)] = \sum_{i=1}^{30}\sum_{t=2}^{70}\log[P_i^{S_i(t)}(t)] \tag{13}$$

and for the individual level model for player i is:

$$LL[\delta_i, \phi_i, \kappa_i, \lambda_i, N_i(0)] = \sum_{t=2}^{70}\log[P_i^{S_i(t)}(t)] \tag{14}$$

where the probabilities $P_i^{S_i(t)}(t)$ are given by (3).

Table 4

Transitions after Higher-Number (Pass) Choices, Nagel and Tang (1998)

	Choices in period $t + 1$ after 'Pass'							
choice in t	2	4	6	8	10	12	14	Total no.
2								0
4		0.50			0.50			2
6	0.08	0.23	0.15	*0.33*	0.18	0.03		39
8	0.01	0.04	0.29	*0.49*	0.15	0.04	0.01	388
10	0.01	0.01	0.08	*0.40*	0.40	0.06	0.03	572
12		0.01	0.03	0.10	0.21	*0.54*	0.11	364
14				0.06	0.10	0.19	*0.65*	231
	1	3	5	7	9	11	13	Total no.
3	1.00							1
5		*0.60*	0.20		0.20			5
7	0.01	0.06	0.25	*0.48*	0.10	0.06	0.04	156
9		0.01	0.04	0.33	*0.48*	0.11	0.02	446
11	0.01		0.02	0.10	0.31	*0.43*	0.12	490
13			0.01	0.05	0.10	0.34	*0.50*	276

There is one substantial change from methods we previously used in Camerer and Ho (1999). We estimated initial attractions (common to all players) from the first period of actual data, rather than allowing them to be free parameters which are estimated as part of the overall maximisation of likelihood.[16] We switched to this method because estimating initial attractions for each of the large number of strategies chewed up too many degrees of freedom.

To search for regularity in the distributions of individual-level parameter estimates, we conducted a cluster analysis on the three most important parameters, δ, ϕ, and κ. We specified a number of clusters and searched iteratively for cluster means in the three-dimensional parameter space which maximises the ratio of the distance between the cluster means and the average within-cluster deviation from the mean. We report results from two-cluster specifications, since they have special relevance for evaluating

[16] Others have used this method too, e.g., Roth and Erev (1995). Formally, define the first-period frequency of strategy j in the population as f^j. Then initial attractions are recovered from the equations

$$\frac{e^{\lambda A^j(0)}}{\sum_k e^{\lambda A^k(0)}} = f^j, j = 1, \ldots, m. \qquad (15)$$

(This is equivalent to choosing initial attractions to maximise the likelihood of the first-period data, separately from the rest of the data, for a value of λ derived from the overall likelihood-maximisation.) Some algebra shows that the initial attractions can be solved for, as a function of λ, by

$$A^j(0) - \frac{1}{m}\sum_j A^j(0) = \frac{1}{\lambda} ln(\tilde{f}^j), j = 1, \ldots, m \qquad (16)$$

where $\tilde{f}^j = f^j/(\Pi_k f^k)^{1/m}$ is a measure of relative frequency of strategy j. We fix the strategy j with the lowest frequency to have $A^j(0) = 0$ (which is necessary for identification) and solve for the other attractions as a function of λ and the frequencies $\tilde{f}^j$.

Estimation of the belief-based model (a special case of EWA) is a little trickier. Attractions are equal to expected payoffs given initial beliefs; therefore, we searched for initial beliefs which optimised the likelihood of observing the first-period data. For identification, λ was set equal to one when likelihood-maximising beliefs were found, then the derived attractions which resulted were rescaled by $1/\lambda$.

whether parameters cluster around the predictions of belief and reinforcement theories. Searching for a third cluster generally improved the fit very little.[17]

4. Results

We discuss the results in three parts: Basic estimation and model fits; individual-level estimates and uncovered clusters; and comparison of three payoff-learning extensions.

4.1. *Basic Estimation and Model Fits*

Table 5 reports the log-likelihood of the various models, both in-sample and out-of-sample. The belief-based model is clearly worst by all measures. This is no surprise because the centipede game is dominance-solvable. Any belief learning should move players in the direction of lower numbers but the numbers they choose rise slightly over time. The EWA-Payoff Clairvoyance is better than the other EWA variants. Reinforcement is worse than any of the EWA variants, by about 50 points of log-likelihood out-of-sample. (It can also be strongly rejected in-sample using standard χ^2 tests.) This finding challenges (Nagel and Tang, 1998), who concluded that reinforcement captured the data well, because they did not consider the EWA learning models.

Another way to judge the model fit is to see how well the EWA model estimates capture the basic patterns in the data. There are two basic patterns:

(*i*) players who choose the lower number (and 'take earlier', in centipede jargon) tend to increase their number more often than they decrease it, and this tendency decreases over time; and
(*ii*) players who choose the higher number ('taking later'), tend to decrease their numbers.

Figure 3*a* shows these patterns in the data and Figures 3*b*–*c* show how well the EWA model describes and predicts these patterns. The EWA predictions are generally quite accurate. Note that if EWA were overfitting in the first 70 periods, accuracy would degrade badly in the last 30 periods (when parameter estimates are fixed and out-of-sample prediction begins); but it generally does not.

4.2. *Payoff Learning Models*

Tables 5–6 show measures of fit and parameter estimates from the three different payoff learning models. The three models make different conjectures on the way subjects estimate the foregone payoffs. All three payoff learning models perform better than reinforcement (which implicitly assumes that the estimated foregone payoff is zero, or gives it zero weight). This illustrates that EWA can improve statistically on reinforcement, even in the domain in which reinforcement would seem to have the biggest advantage over other models – i.e., when foregone payoffs are not known. By simply adding a payoff-learning assumption to EWA, the extended model outpredicts reinforcement. Building on our idea, the same value of adding payoff learning to EWA

[17] Specifically, a three-segment model always leads to a tiny segment that contains either 1 or 2 subjects.

Table 5

Log Likelihoods and the Parameter Estimates of the Various Representative-Agent Adaptive Learning Models

		LL		Parameter Estimates (Standard Error)				
Model	Number of parameters	In Sample	Out of Sample	φ	δ	κ	N0	λ
Odd Players								
Reinforcement	2	−2713.2	−1074.5	0.92 (0.0002)	0.00	1.00	1.00	0.01 (0.0000)
Belief	3	−3474.2	−1553.1	1.00 (0.0009)	1.00	0.00	100 (0.0000)	0.57 (0.0008)
EWA, Recent Actual Payoff	5	−2667.6	−1069.8	0.91 (0.0002)	0.14 (0.0003)	1.00 (0.0000)	1.00 (0.0000)	0.01 (0.0000)
EWA, Payoff Clairvoyance	5	−2596.6	−1016.8	0.91 (0.0002)	0.32 (0.0001)	1.00 (0.0000)	1.00 (0.0000)	0.01 (0.0000)
EWA, Average Payoff	5	−2669.3	−1064.9	0.91 (0.0002)	0.15 (0.0002)	1.00 (0.0000)	1.00 (0.0000)	0.01 (0.0000)
Even Players								
Reinforcement	2	−2831.8	−991.7	0.92 (0.0002)	0.00	1.00	1.00	0.01 (0.0000)
Belief	3	−3668.9	−1556.0	0.87 (0.0014)	1.00	0.00	0.16 (0.0004)	0.04 (0.0000)
EWA, Recent Actual Payoff	5	−2811.9	−983.0	0.91 (0.0002)	0.15 (0.0001)	1.00 (0.0000)	1.00 (0.0000)	0.01 (0.0000)
EWA, Payoff Clairvoyance	5	−2791.4	−953.2	0.90 (0.0002)	0.24 (0.0004)	1.00 (0.0006)	7.91 (0.0000)	0.13 (0.0000)
EWA, Average Payoff	5	−2802.1	−1039.2	0.90 (0.0006)	0.17 (0.0005)	0.99 (0.0015)	1.01 (0.0000)	0.01 (0.0000)

Table 6

A Comparison between the Representative-Agent and Individual-level Parameter Estimates of the Various EWA Models

	LL		Mean Parameter Estimates				
Model	In Sample	Out of Sample	φ	δ	κ	N0	λ
Odd Players							
EWA, Recent Actual Payoff							
Representative-Agent	−2667.6	−1069.8	0.91	0.14	1.00	100	0.01
Individual-level	−2371.2	−1050.6	0.86	0.25	0.48	1.65	0.19
EWA, Payoff Clairvoyance							
Representative-Agent,	−2596.6	−1016.8	0.91	0.32	1.00	1.00	0.01
Individual-level	−2301.2	−1052.0	0.92	0.44	0.38	1.84	0.13
EWA, Average Payoff							
Representative-Agent	−2669.3	−1064.9	0.91	0.15	1.00	1.00	0.01
Individual-level	−2334.6	−1017.2	0.89	0.26	0.25	2.75	0.22
Even Players							
EWA, Recent Actual Payoff							
Representative-Agent	−2811.9	−983.0	0.91	0.15	1.00	1.00	0.01
Individual-level	−2442.5	−912.7	0.89	0.32	0.33	2.80	0.17
EWA, Payoff Clairvoyance							
Representative-Agent	−2791.4	−953.2	0.90	0.24	1.00	7.91	0.13
Individual-level	−2421.7	−927.6	0.90	0.47	0.34	3.94	0.17
EWA, Average Payoff							
Representative-Agent	−2802.1	−1039.2	0.90	0.17	0.99	1.01	0.01
Individual-level	−2432.4	−960.6	0.84	0.35	0.39	4.59	0.15

is shown by Anderson (1998) in bandit problems, Chen and Khoroshilov (2003) in a study of joint cost allocation, and Ho and Chong (2003) in consumer product choice at supermarkets.

The three payoff learning assumptions embody low and high degrees of player knowledge. The assumption that players recall only the last actual payoff – which may have been received many periods ago – means they ignore deeper intuitions about which of the possible payoffs might be the correct foregone one in the very last period. Conversely, the payoff clairvoyance assumption assumes the players somehow figure out exactly which foregone payoff they would have got. The average payoff assumption seems more sensible and infers the foregone payoff based on the observed actual outcome in each period. Surprisingly, the payoff clairvoyance assumption predicts better. The right interpretation is surely not that subjects are truly clairvoyant, always guessing the true foregone payoff perfectly but simply that their implicit foregone payoff estimate is closer to the truth than the last actual payoff or the average payoff is. For example, consider a player *B* who chooses 6 and has the lower of the two numbers. If she had chosen strategy 8 instead, she does not know whether the foregone payoff would have been 8 (if the other *A* subject chose 7), or 45 (if the *A* subject chose 9, 11, or 13). The payoff clairvoyance assumption says she knows precisely whether it would have been 8 or 45 (i.e., whether subject *A* chose 7, or chose 9 or more). While this requires knowledge she does not have, it only has to be a better guess than the last actual payoff she got from choosing strategy 8 and the average payoff for the clairvoyance model to provide the best fit.

4.3. *Individual Differences*

The fact that Nagel and Tang's game lasted 100 trials enabled us to estimate individual-level parameters with some reliability (while imposing common initial attractions). Figures 4*a–b* show scatter plot 'parameter patches' of the 30 estimates from the payoff-clairvoyance EWA model in a three-parameter $\delta - \phi - \kappa$ space. Each point represents a triple of estimates for a specific player; a vertical projection to the bottom face of the cube helps the eye locate the point in space and measure its $\phi - \kappa$ values. Figure 4*a* shows Odd players and Figure 4*b* shows Even players.

Table 5 shows the mean of the parameter estimates, along with standard deviations across subjects, for the EWA models. Results for Odd and Even players are reported separately, because the game is not symmetric. The separate reporting also serves as a kind of robustness check, since there is no reason to expect their learning parameters to be systematically different; and in fact, the parameters are quite similar for the two groups of subjects.

The EWA parameter means of the population are quite similar across the three payoff-learning specifications and player groups (see Table 6). The consideration parameter δ ranges from 0.25 to 0.47, the change parameter ϕ varies only a little, from 0.84 to 0.92, and the commitment parameter κ from 0.25 to 0.48. The standard deviations of these means can be quite large, which indicates the presence of substantial heterogeneity.

Individuals do not particularly fall into clusters corresponding to any of the familiar special cases (compare Figure 2 and Figures 4*a–b*). For example, only a couple of the subjects are near the cumulative reinforcement line $\delta = 0$, $\kappa = 1$ (the 'bottom back wall'). However, quite a few subjects are clustered near the fictitious play upper left corner where $\delta = 1$, $\phi = 1$ and $\kappa = 0$.

The cluster analyses from the EWA models do reveal two separate clusters which are easily interpreted. The means and within-cluster standard deviations of parameter values are given in Table 7. The subjects can be sorted into two clusters, of roughly equal size. Both clusters tend to have δ around 0.40 and ϕ around 0.80–0.90; however, in one cluster κ is very close to zero and in the other cluster κ is close to one. Graphically, subjects tend to cluster on the front wall representing low ($\kappa = 0$) commitment, and the back wall representing high ($\kappa = 1$) commitment.

In most of our earlier work (and most other studies), all players are assumed to have the same learning parameters (i.e., a representative agent approach). Econometrically, it is possible that a parameter estimated with that approach will give a biased estimate of the population mean of the same parameter estimated across individuals, when there is heterogeneity. We can test for this danger directly by comparing the mean of parameter estimates in Table 6 with estimates from a single-agent analysis assuming homogeneity. The estimates are generally close together, but there are some slight biases which are worth noting. The estimates from the representative agent approach show that ϕ tends to be very close to the population mean. However, δ tends to be under-estimated by the representative-agent model, relative to the average of individual-agent estimates. This gap explains why some early work on reinforcement models using representative-agent modelling (which assumes $\delta = 0$), leads to surprisingly good fits. Furthermore, the parameter κ from the single-agent

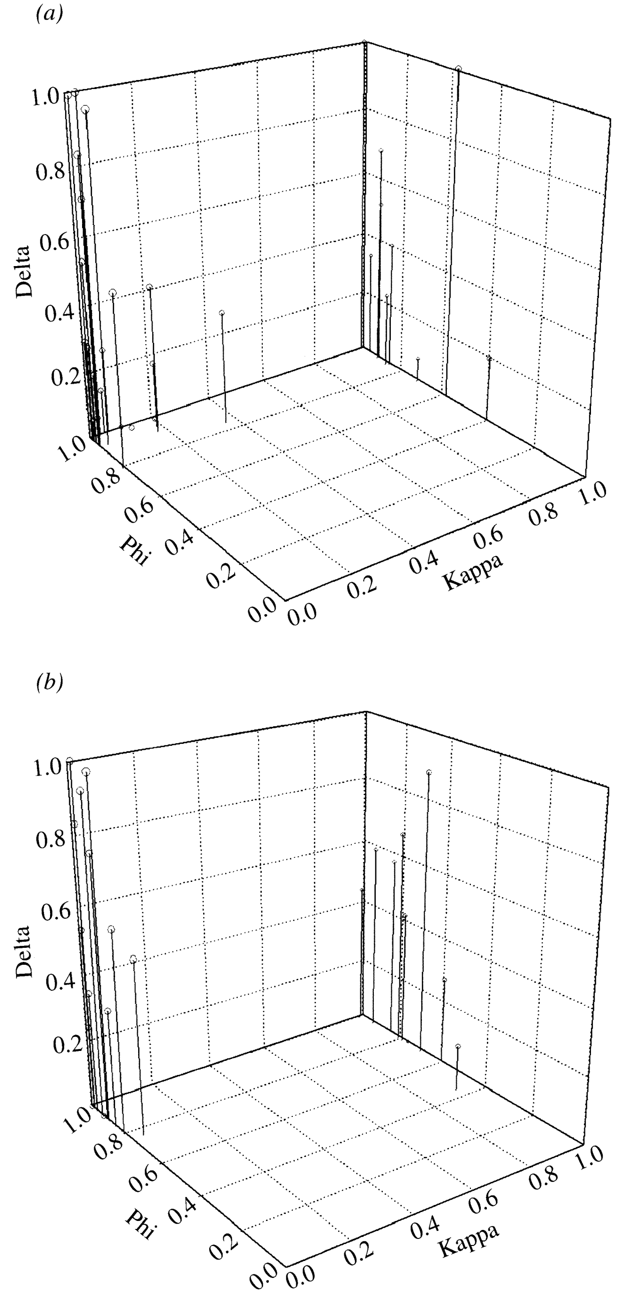

Fig. 4. *Individual-level Payoff Clairvoyance EWA Model Parameter Patches. (a) Odd Subjects; (b) Even Subjects*

model tends to take on the extreme value of 0 or 1, when the sample means are around 0.40. Since there is substantial heterogeneity among subjects – the clusters show that subjects tend to have high κs near 1, or low values near 0 – as if the single-

Table 7

A Cluster Analysis Using Individual-level Estimates

Mean Parameter Estimates (Std. Dev.)							
Odd Players				Even Players			
Number of subjects	φ	δ	κ	Number of subjects	φ	δ	κ
20	0.96	0.40	0.07	21	0.96	0.48	0.02
	(0.07)	(0.35)	(0.10)		(0.08)	(0.36)	(0.03)
10	0.82	0.51	0.99	9	0.76	0.44	0.98
	(0.20)	(0.33)	(0.01)		(0.17)	(0.27)	(0.02)

agent model uses a kind of 'majority rule' and chooses one extreme value or the other, rather than choosing the sample mean. Future research can investigate why this pattern of results occurs.

5. Conclusions

In this article, we extend our experience-weighted attraction (EWA) learning model to games in which players know the *set* of possible foregone payoffs from unchosen strategies, but do not precisely which payoff they would have gotten. This extension is crucial for applying the model to naturally-occurring situations in which the modeller (and even the players) do not know much about the foregone payoffs.

To model how players respond to unknown foregone payoffs, we allowed players to learn about them by substituting the last payoffs received when those strategies were actually played, by averaging the set of possible foregone payoffs conditional on the actual outcomes, or by clairvoyantly guessing the actual foregone payoffs. Our results show that these EWA variants fit and predict somewhat better than reinforcement and belief learning. The clairvoyant-guessing model fits slightly better than the other two variants.

We also estimated parameters separately for each individual player. The individual estimates showed there is substantial heterogeneity but individuals could not be sharply clustered into either reinforcement or belief-based models (though many did have fictitious play learning parameters). They could, however, be clustered into two distinct subgroups, corresponding to averaging and cumulating of attraction. Compared to the means of individual level estimates, the parameter estimates from the representative-agent model have a tendency to modestly underestimate δ and take on extreme values for κ.

Future research should apply these payoff-learning specifications, and others, to environments in which foregone payoffs are unknown (Anderson, 1998; Chen, 1999). If we can find a payoff-learning specification which fits reasonably well across different games, then EWA with payoff learning can be used on naturally-occurring data sets – see Ho and Chong (2003) for a recent application – taking the study of learning outside the laboratory and providing new challenges.

University of California, Berkeley
Brandeis University
California Institute of Technology

Submitted: 12 March 2005
Accepted: 16 December 2006

References

Anderson, C. (1998). 'Learning in bandit problems', Caltech Working Paper.

Anderson, C. and Camerer, C.F. (2000). 'Experience-weighted attraction learning in sender-receiver signaling games', *Economic Theory*, vol. 16 (3), pp. 689–718.

Biyalogorsky, E., Boulding, W. and Staelin, R. (2006). 'Stuck in the past: why managers persist with new product failures', *Journal of Marketing*, vol. 70 (2), pp. 108–21.

Boulding, W., Kalra, A. and Staelin, R. (1999). 'Quality double whammy', *Marketing Science*, vol. 18 (4), pp.463–84.

Bracht, J. and Ichimura, H. (2001). 'Identification of a general learning model on experimental game data', Hebrew University of Jerusalem Working Paper.

Broseta, B. (2000). 'Adaptive learning and equilibrium selection in experimental coordination games: an ARCH(1) approach', *Games and Economic Behavior*, vol. 32 (1), pp. 25–50.

Brown, G. (1951). 'Iterative solution of games by fictitious play', in (T.C. Koopmans, ed.), *Activity Analysis of Production and Allocation*, New York: John Wiley & Sons.

Camerer, C.F. (2003). *Behavioral Game Theory*. Princeton: Princeton University Press.

Camerer, C.F. and Ho, T-H. (1998). 'Experience-weighted learning in coordination games: Probability rules, heterogeneity, and time variation', *Journal of Mathematical Psychology*, vol. 42 (2), pp. 305–26.

Camerer, C.F. and Ho, T-H. (1999). 'Experience-weighted attraction learning in normal-form games', *Econometrica*, vol. 67 (4), pp. 827–74.

Camerer, C.F., Ho, T-H. and Chong, J-K. (2002). 'Sophisticated learning and strategic teaching', *Journal of Economic Theory*, vol. 104 (1), pp. 137–88.

Chen, Y. (1999). 'Joint cost allocation in asynchronously updated systems', University of Michigan Working Paper.

Chen, Y. and Khoroshilov, Y. (2003). 'Learning under limited information', *Games and Economic Behavior*, vol. 44 (1), pp. 1–25.

Cheung, Y-W. and Friedman, D. (1997). 'Individual learning in normal form games: some laboratory results', *Games and Economic Behavior*, vol. 19 (1), pp. 46–76.

Chong, J-K., Camerer, C. F. and Ho, T-H. (2006). 'A learning-based model of repeated games with incomplete information', *Games and Economic Behavior*, vol. 55 (2), pp. 340–71.

Cournot, A. (1960). *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.

Crawford, V. (1995). 'Adaptive dynamics in coordination games', *Econometrica*, vol. 63 (1), pp. 103–43.

Erev, I. and Roth, A. (1998). 'Modelling predicting how people play games: reinforcement learning in experimental games with unique, mixed-strategy equilibria', *American Economic Review*, vol. 88 (4), pp. 848–81.

Fudenberg, D. and Levine, D. (1998). *The Theory of Learning in Games*, Cambridge, MA: The MIT Press.

Harley, C.B. (1981). 'Learning the evolutionarily stable strategy', *Journal of Theoretical Biology*, vol. 89 (4), pp. 611–33.

Ho, T-H. (forthcoming). 'Individual learning in games', in (L. Blume, and S. Durlauf, eds.), *The New Palgrave Dictionary of Economics: Design of Experiments and Behavioral Economics*, Basingstoke: Palgrave.

Ho, T-H. and Chong, J-K. (2003). 'A parsimonious model of SKU choice', *Journal of Marketing Research*, vol. 40 (August), pp. 351–65.

Ho, T-H. and Weigelt, K. (1996). 'Task complexity, equilibrium selection, and learning: an experimental study', *Management Science*, vol. 42 (5), pp. 659–79.

Ho, T-H., Camerer, C.F. and Chong, J-K. (2007). 'Self-tuning experience-weighted attraction learning in games', *Journal of Economic Theory*, vol. 133(1), pp. 177–98.

Hopkins, E. (2002). 'Two competing models of how people learn in games', *Econometrica*, vol. 70 (6), pp. 2141–66.

Hsia, D. (1998). 'Learning in call markets', University of Southern California Working Paper.

McAllister, P.H. (1991). 'Adaptive approaches to stochastic programming', *Annals of Operations Research*, vol. 30 (June), pp. 45–62.

McKelvey, R.D. and Palfrey, T.R. (1995). 'Quantal response equilibria for normal form games', *Games and Economic Behavior*, vol. 10 (1), pp. 6–38.

Mailath, G. (1998). 'Do people play Nash equilibrium? Lessons from evolutionary game theory', *Journal of Economic Literature*, vol. 36 (3), pp. 1347–74.

Marcet, A. and Nicolini, J. P. (2003). 'Recurrent hyperinflations and learning', *American Economic Review*, vol. 93 (5), pp. 1476–98.

Mookerjee, D. and Sopher, B. (1994). 'Learning behavior in an experimental matching pennies game', *Games and Economic Behavior*, vol. 7 (1), pp. 62–91.

Mookerjee, D. and Sopher, B. (1997). 'Learning and decision costs in experimental constant-sum games', *Games and Economic Behavior*, vol. 19 (1), pp. 97–132.

Morgan, J. and Sefton, M. (2002). 'An experimental investigation of experiments on unprofitable games', *Games and Economic Behavior*, vol. 40 (1), pp. 123–46.

Nagel, R. and Tang, F. (1998). 'Experimental results on the centipede game in normal form: an investigation on learning', *Journal of Mathematical Psychology*, vol. 42, pp. 356–84.

Rapoport, A. and Amaldoss, W. (2000). 'Mixed strategies and iterative elimination of strongly dominated strategies: an experimental investigation of states of knowledge', *Journal of Economic Behavior and Organization*, vol. 42 (4), pp. 483–521.

Robinson, J. (1951). 'An iterative method of solving a game', *Annals of Mathematics*, vol. 54 (2), pp. 296–301.

Roth, A. (1995). 'Introduction', in (J.H. Kagel and A. Roth, eds.), *The Handbook of Experimental Economics*, Princeton: Princeton University Press.

Roth, A. and Erev, I. (1995). 'Learning in extensive-form games: experimental data and simple dynamic models in the intermediate term', *Games and Economic Behavior*, vol. 8 (1), pp. 164–212.

Salmon, T. (2001). 'An evaluation of econometric models of adaptive learning', *Econometrica*, vol. 69 (6), pp. 1597–628.

Sarin, R. and Vahid, F. (2004). 'Strategy similarity and coordination', Economic Journal, vol. 114 (497), pp. 506–27.

Schlag, K. (1999). 'Which one should I imitate?', *Journal of Mathematical Economics*, vol. 31 (4), pp. 493–522.

Selten, R. (forthcoming). 'Bounded rationality and learning', in (E. Kalai ed.), *Collected Volume of Nancy Schwartz Lectures*, pp. 1–13, Cambridge: Cambridge University Press.

Selten, R. and Stoecker, R. (1986). 'End behavior in sequences of finite prisoner's dilemma supergames: a learning theory approach', *Journal of Economic Behavior and Organization*, vol. 7 (1), pp. 47–70.

Stahl, D. (1999). 'Sophisticated learning and learning sophistication', University of Texas Working Paper.

Stahl, D. (2000). 'Rule learning in symmetric normal-form games: theory and evidence', *Games and Economic Behavior*, vol. 32 (1), pp. 105–38.

Stahl, D. and Haruvy, E. (2004). 'Rule learning across dissimilar normal-form games', University of Texas Working Paper.

Van Huyck, J., Cook, J. and Battalio, R. (1997). 'Adaptive behavior and coordination failure', *Journal of Economic Behavior and Organization*, vol. 32 (4), pp. 483–503.

Vriend, N. (1997). 'Will reasoning improve learning?', *Economics Letters*, vol. 55 (1), pp. 9–18.

Wilcox, N. (2006). 'Theories of learning in games and heterogeneity bias', *Econometrica*, vol. 74 (5), pp. 1271–92.