DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

EXPERIENCE-WEIGHTED ATTRACTION LEARNING IN SENDER-RECEIVER SIGNALING GAMES

Christopher M. Anderson

Colin F. Camerer

SOCIAL SCIENCE WORKING PAPER 1058

March 1999

Experience-Weighted Attraction Learning in Sender-Receiver Signaling Games[*]

Christopher M. Anderson   Colin F. Camerer

Abstract

Recent experiments have indicated that it is possible to systematically lead subjects to less refined equilibria in signaling games. In this paper, we seek to understand the process by which this occurs using Camerer and Ho's Experience Weighted Attraction (EWA) model of learning in games. We first adapt the model to extensive-form signaling games by specifying that senders update the chosen message for both the realized and unrealized type, but do not update the unchosen message. We test this model against the choice reinforcement and belief-based special cases of EWA; the latter is of particular interest because it formalizes the story about convergence to less refined equilibria offered by Brandts and Holt. We also test a variety of models which update unchosen messages. We find that while the Brandts-Holt story captures the the direction of switching from one strategy to another, it does not do a good job at capturing the rate at which the switching occurs. EWA does quite well at predicting the rate of switching, and is slightly bettered by the unchosen message models, which all perform equally well.

[*] This research was supported by NSF SBR 9511001. Thanks to Jordi Brandts and Charlie Holt for supplying their raw data. Helpful comments were received from audiences at the Universities of California (Berkeley), Texas (Austin) and the Fall 1998 ESA Meetings.

1 Introduction

For a noncooperative game of any complexity, it is likely that people learn how to play the game through experience, rather than figure it out by reasoning. A general theory of learning is therefore crucial for understanding equilibration theoretically, and for explaining the changes in strategic behavior observed in the lab and in the field.

Most research on learning in games has focussed on either of two approaches: Reinforcement models, in which payoffs to strategies increase the likelihood of playing those strategies in the future; and belief-based models, in which players construct a belief from the observed history of other players' play, and optimize (perhaps imperfectly) given their beliefs. Camerer and Ho (1999a,b) introduced a general model, called 'experience-weighted attraction learning' (EWA), which combines the most psychologically plausible and empirically useful elements of the reinforcement and belief learning approaches. The crucial insight is that belief learning is exactly the same as a kind of generalized reinforcement learning in which strategies which are not actually chosen are also reinforced according to the foregone payoff they would have yielded. Camerer and Ho estimate parameters of the model on four experimental data sets from normal-form games. They find that it is necessary to use elements of both approaches in order to explain observed learning.

In this paper we apply the EWA model to experimental data from sender-receiver signaling games. Signaling games are very widely used to model economic and political phenomena in which actions—perhaps apparently irrational ones leading to avoidable inefficiencies—are taken to convey asymmetric information. Applications include strategic delay in strikes, firm behavior in industrial organization theories, labor and insurance markets with hidden action and hidden information, "money-burning" models of gift-giving, and many more (see, for example, Tirole, 1988, and Gibbons, 1992).

EWA, and most other learning models, have previously been applied only to normal-form games. Extensive-form games with incomplete information, like signaling games, demand special modifications which extend EWA's scope. The key problem is that players do not always know the foregone payoff to a strategy which they did not choose (because its payoff depends on other players' types or moves, which they do not observe), and because foregone payoffs are used to update unchosen strategies, an extension is necessary when foregone payoffs are not known. However, since players know the set of possible foregone payoffs, we extend EWA by reinforcing unchosen strategies according to some mixture of the foregone payoffs in that set.

Studying learning in signaling games is especially interesting because these games often have many equilibria. There are a large number of 'refinement' concepts which are routinely used to justify why some equilibria are empirically likely and others are not. However, these refinements usually assume players are reasoning particularly logically. But if players learn equilibria rather than figure them out, it is an open question whether their learning will lead to logically refined equilibria more often than unrefined equilibria.

In fact, most experimental tests have yielded somewhat pessimistic results about the ability of refinements, even fairly simple ones, to predict which equilibria experimental subjects will converge to (e.g., Banks, Camerer and Porter, 1994). Brandts and Holt (1992, 1993) suggested players in their experiments were using a particular learning process (essentially a form of belief learning) (see also Cooper, Kagel and Garvin, 1997a,b). They designed a game in which players learning according to that process would be led to an equilibrium which violated the Cho-Kreps 'intuitive criterion'. The trick is that during equilibration, players leave empirical "footprints" at all information sets by choosing strategies which will, later, turn out to be rarely chosen. When behavior eventually crystallizes around an equilibrium, and players think about which types of players are likely to make out-of-equilibrium moves, they use their previous actual experience to form beliefs. In some games, these empirical beliefs contradict purely logical arguments about which players would choose the out-of-equilibrium move. The learning process therefore supports an equilibrium which is not supportable by standard game-theoretic logic.

Since the Brandts-Holt model is a special kind of belief learning model, and belief models are nested in EWA as a special case, by applying EWA to data from signaling experiments we can test the Brandts-Holt theory, and see whether adding additional EWA elements improves the fit.

Our paper therefore makes three contributions: We extend EWA to extensive-form games with incomplete information, in which there is imperfect information about foregone payoffs. We extend earlier experiments on signaling games, running them for 32 periods to see if sharper convergence occurs. Finally, we estimated the extended EWA model (which includes the Brandts-Holt dynamics as a special case) on the new data.

The paper is organized as follows. First we describe the games and the adaptive dynamics conjectured by Brandts and Holt. Then we show their data, to give the reader a concrete sense of what it is the learning models are trying to explain. Then the EWA model is described and the modifications necessary to fit it to the signaling data are detailed. Then we present data from new experiments and show how well EWA, its various extensions, and the belief and reinforcement special cases, explain the data.

2 Adaptive Dynamics and Equilibrium Selection

[1]

The main purpose of this paper is to apply EWA to signalling games, where it may be able explain how learning dynamics can lead player to unrefined equilibria.

This phenomenon is illustrated by two games taken from Brandts and Holt (1993), extending work by Banks, Camerer and Porter (1994) and Brandts and Holt (1992). Tables 1 and 2 show their Games 3 and 5. Nature chooses Type I or Type II (with equal probabilities) and the sender is told which half the table will be used to determine payoffs. The sender then selects M_1 or M_2, and the receiver is notified of the sender's choice, but not the type. The receiver then chooses an action, a_1, a_2 or a_3. Payoffs are determined from the cell in the table described by the type-message-action triple; the sender's payoff is on the left and the receiver's payoff is on the right.

In Game 3, there are two equilibria. Both are sequential but only one satisfies the Cho-Kreps (1987) intuitive criterion.[2]

[1] This section provides a detailed account of how empirical histories can contradict the theoretical criteria on which equilibrium refinement arguments are based, and thus players can be led systematically to less refined equilibria. Readers not interested in this phenomenon can skip this section without loss of continuity.

[2] In the intuitive equilibrium both types choose m_2 and are met with the response a_1, yielding t_1s 45 and t_2s 30. Since t_2s could conceivably earn more (45) by choosing m_1 instead, the equilibrium only sticks if defections to m_1 are met with responses of a_2. The a_2 response to m_1 can only be justified by the belief that m_1 defections are more likely to have come from t_2s (i.e., a_2 is optimal for receivers if $P(t_2|m_1) > 5/7$). This inference does satisfy the intuitive criterion because, indeed, t_2 types might benefit by defecting from m_2 to m_1 whereas t_1s would never benefit. Hence, the equilibrium in which both types choose m_2 satisfies the intuitive criterion.

In the unintuitive sequential equilibrium, both types of senders (t_1 and t_2) send message m_1. Since both types choose m_1, receivers come to realize this, Bayesian-update and form posteriors $P(t_1|m_1) = P(t_1) = .5$ and $P(t_2|m_1) = P(t_2) = .5$. Their best response is then to choose a_1, which gives expected payoff 45 $(.5 \cdot 0 + .5 \cdot 90)$. Note that in this equilibrium, t_1 senders earn 30 and t_2s earn 45. What prevents the t_1s from defecting to m_2, and earning 45 if their message is met with a response of a_1? The sequential equilibrium coheres only if receivers choose a_2 in response to message m_2. (Note that a_3 in response to m_2 is strictly dominated and should never be chosen; in the experiments, it rarely is.) But why would receivers choose a_2 in response to m_2? Receivers must believe that a_2 choices were more likely to be made by t_2s (more specifically, $P(m_2|t_2) > 2/3$ to justify a choice of a_2 by receivers).

At this point, the intuitive criterion weighs in with a theoretical opinion about whether $P(m_2|t_2) > 2/3$ is plausible. This belief is <u>not</u> plausible, the argument goes, because t_2s earn 45 in equilibrium (from choosing m_1 and getting response a_1) and could not possibly benefit from switching to m_2. On the other hand, t_1s earn 30 in equilibrium and could conceivably benefit if they choose m_2 and are met by the response a_1, earning 45. Hence, the intuitive criterion concludes, the off-path belief $P(m_2|t_2) > 2/3$ is not intuitive because it shifts belief from the prior toward t_2s, who are <u>least</u> likely to benefit from having picked m_2.

Note that the argument underlying the intuitive criterion is purely theoretical. It deduces from payoffs the implausibility of t_2s having switched from m_1 to m_2. From a learning point of view, this is only sensible if the t_2s have not actually chosen m_2 very often, or if subjects have been choosing m_1 for a long time and any history of m_2 choices is forgotten or dismissed. But learning implies a pre-equilibrium convergence process which leaves an empirical trace of previous m_2 choices. What if, during the learning process, most players who chose m_2 <u>did</u> happen to be t_2s? Then the intuitive criterion competes with an empirical construction of off-path beliefs, recalled from the equilibration phase in which out-of-equilibrium moves were common (before such moves <u>became</u> "out of equilibrium"). Faced with a choice between the logic of the intuitive criterion, and what they observed, it is hard to see why players would reject previous observation in favor of a purely deductive argument which flagrantly ignores history.

From an equilibrium analysis point of view, Game 5 is identical to Game 3. In the Game 5 sequential equilibrium, both types pick m_1 and are met with response a_1, so t_1s earn 30 and t_2s earn 45. The t_1s are prevented from defecting to m_2 only if a defection is met with response a_2, which is justified if receivers think a defection probably came from a t_2 sender ($P(t_2|m_2) > 2/3$). But in the m_1 (sequential) equilibrium, t_2s never do better defecting while t_1s might; so the presumption that defections were probably from t_2s is unintuitive.

2.1 Experimental Data and the Brandts-Holt Adjustment Dynamic

Figure 1 summarizes BH's results for games 3 and 5. In their experiments, subjects switch roles and play for 12 periods. The data shown are averages of 4-period blocks.

In Game 3, they observe significant initial type separation— in the first block m_1 is three times more likely to be chosen by a t_1 sender than a t_2 sender. BH explain this by assuming that senders start with a diffuse prior on what the likely action responses will be. With a diffuse prior, the expected payoffs for t_1s are 30 (=(45+15+30)/3) and 25 (=(30+0+45)/3) for the two messages, so t_1 tends to choose message m_1 more. The expected payoffs for t_2 are 20 and 30, so t_2s choose m_2 more often.

If receivers also start with diffuse priors on which types chose a particular message, they should assign the highest expected payoffs to action a_1 in response to m_1, and a_1 in response to m_2. However, they are more likely to choose a_2 in response to m_2. This happens quickly (in the first two periods) so it appears that receivers have anticipated the type-separation, or learned it very quickly, and use it update their beliefs that a message m_2 choice came from a t_2, which makes the action response a_2 optimal. As the game continues and t_2 players continue to receive a_2 responses to m_2, they earn payoffs of 15 and begin to switch to m_1. In the last four-period block all the t_1s pick m_1 and about 60% of the t_2s pick m_1, so equilibration goes reasonably swiftly in the direction of the intuitive sequential equilibrium in which both types pool on m_1.

That equilibrium is supported by the belief that a message m_2 would be chosen by a t_2 (who could conceivably benefit), leading the receiver to choose a_2 in response, which yields a lower payoff for the type t_2 than she receives from pooling on m_1, which keeps her from defecting. Because of the initial type separation, most of the historical choices of m_1 were from t_2s, so there were few observations which conflicted with the intuitive criterion.

Game 5, in contrast, is designed so that observations which are likely to emerge from early disequilibrium play will conflict with the intuitive criterion. This is indeed what happens. In early periods, nearly all t_1 senders choose m_1 and most t_2s choose m_2. Receivers seem to anticipate, or learn quickly, that different types choose different messages, and they tend toward actions which are best responses given the type separation, i.e., $a_2|m_1$ and $a_1|m_2$.

Recall that in the unintuitive sequential equilibrium in which both types choose m_2, defection to m_1 is prevented if receivers will think such a defection came from a t_2. Indeed, since the empirical probability of $m_1|t_2$ is high, their belief is justified by past experience (though it conflicts with the cold logic of the intuitive criterion). This highlights the need for a theory of equilibrium selection which includes a description of the convergence path and respects the way the observed convergence affects players' later beliefs. Without a story about how observations conflict with rational conjectures about beliefs, it is hard to explain this convergence to the less refined equilibrium.

In the Brandts and Holt story (1993), players start with beliefs about what others will do and revise their beliefs in the light of what they observe.[3] The belief story does seem to explain the major features of the data. However, in the empirical literature on learning it has been shown that one model seems to explain a convergence path reasonably well, but then be rejected in favor of a competing model when they are compared carefully. Since EWA learning includes belief-based learning as a special case, but mixes in three elements of reinforcement learning (flexible initial conditions, cumulation of reinforcement, and extra weight on received payoffs), it is possible that adding these features to the BH story will improve the fit.

3 EWA Learning

Experience-weighted attraction learning was introduced to hybridize 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 (1999b).

[3] Cooper, Kagel and Garvin (1997a,b) give a similar explanation for results in limit pricing experiments with an important twist: players assume others do not violate dominance.

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 normalized by a factor which captures the (decayed) amount of experience players have accumulated. Attractions to strategies are then related to the probability of choosing those strategies using a response function which guarantees that more attractive strategies are played more often.

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 $S_i = \{s_i^1, s_i^2, \ldots, s_i^{\ell_i - 1}, s_i^{\ell_i}\}$, 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 $\pi_i(s_i(t), s_{-i}(t))$.

3.1 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 $A_i^j(t)$, the attraction of a strategy 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.

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

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

The parameter ρ can be thought of as a depreciation rate or retrospective discount factor that measures the fractional impact of previous experience, compared to one new period. Notice that the steady-state value of $N(t)$ is $\frac{1}{1-\rho}$ (and does not depend on $N(0)$). In the estimation we usually impose the restriction $N(0) \leq \frac{1}{1-\rho}$ which guarantees that the experience weight rises over time, so the relative weight on new payoffs falls and learning slows down.

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 j^{th} strategy can be written $[\delta + (1-\delta) \cdot I(s_i^j, s_i(t))] \cdot \pi_i(s_i^j, s_{-i}(t))$.

The rule for updating attraction sets $A_i^j(t)$ to be a depreciated, experience-weighted lagged attraction, plus an increment for the received or foregone payoff, normalized by the new experience weight. That is,

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

The factor ϕ is a discount factor that depreciates previous attractions.

Finally, attractions must be related to 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

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

because it allows negative attractions and seems to fit 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.

3.2 The Cumulative Reinforcement Special Case of EWA

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

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

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

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

It is easy to see that this updating formula is a special case of the EWA rule, when $\delta = 0$, $N(0) = 1$, and $\rho = 0$. Some people have argued that reinforcement learning of this sort may be an adequate approximation of human learning in games, even though it is simple (and has been largely abandoned by cognitive psychologists studying humans). The adequacy of the approximation 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).

3.3 The Belief-Based Special Case of EWA

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) 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 $N(t) = \sum_{k=1}^{m_{-i}} N_{-i}^k(t)$.[4] Express the probability that others will play strategy k as $B_{-i}^k(t) = \frac{N_{-i}^k(t)}{N(t)}$, with $N_{-i}^k(t) \geq 0$ and $N(t) > 0$.

[4] 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.

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,

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

This form of belief updating weights the belief from one period ago ρ times as much as the most recent observation, so ρ can be interpreted as how quickly previous experience is discarded.[5] When $\rho = 0$ players weight only the most recent observation (Cournot dynamics); when $\rho = 1$ all previous observations count equally (fictitious play).

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

$$E^j_i(t) = \sum_{k=1}^{m_{-i}} B^k_{-i}(t)\pi(s^j_i, s^k_{-i}). \tag{7}$$

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

$$E^j_i(t) = \frac{\rho \cdot N(t-1) \cdot E^j_i(t-1) + \pi(s^j_i, s_{-i}(t))}{\rho \cdot N(t-1) + 1}. \tag{8}$$

If the initial attractions in the EWA model are expected payoffs given some initial beliefs (i.e., $A^j_i(0) = E^j_i(0)$), the attraction depreciation rate ϕ equals the experience depreciation rate ρ, 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 surprising kinship between reinforcement and belief approaches. Belief learning is a kind of generalized reinforcement learning in which strategies are reinforced equally strongly by actual payoffs and foregone payoffs, reinforcements are weighted averages, and initial reinforcements must spring from prior beliefs.

[5] 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.

3.4 Interpreting EWA

We believe one of the strengths of EWA is that its parameters have sensible psychological interpretations.

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 'imagination' of foregone payoffs, or responsiveness to foregone payoffs (when δ is larger players move more strongly toward ex post best responses).

The parameter ϕ is naturally interpreted as depreciation of past attractions, $A(t)$. The parameter ρ depreciates the experience measure $N(t)$. It captures something like decay in the strength of prior beliefs, which is generally different than decay of early attraction (captured by ϕ). In a game-theoretic context, ρ and ϕ will be affected by the degree to which players realize other players are adapting, so that old observations on what others did become less and less useful. Then ρ and ϕ can be interpreted as indices of the perceived rate of change.

The relation between ϕ and ρ determines the growth rate of attractions, which in turn affects how sharply players converge. To see this, set $\delta = 1$ for simplicity, and rewrite the updating equation as

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

If $\phi = \rho$, the first term in the numerator is one and disappears. Then attractions are evidently just a weighted average of lagged attractions and previous payoffs, where the weights are $\rho \cdot N(t-1)$ and 1. Then the attractions will be bounded by the scale of the payoffs; they will never grow too far apart.

If $\phi > \rho$, then the attractions are kind of 'inflated' weighted average where the depreciated, experience-weighted lagged attraction is multiplied by an inflation rate $\frac{\phi}{\rho}$. Then attractions can grow outside the bounds of the payoffs, which means that attractions for strategies can grow farther apart. In the extreme case where $\rho = 0$, $N(t) = 1$ for $t \geq 1$ and the attractions are simply

$$A_i^j(t) = \phi \cdot A_i^j(t-1) + \pi_i(s_i^j, s_{-i}(t)). \tag{10}$$

That is, attractions are just (depreciated) cumulative payoffs.

Whether attractions grow or not is important because in the logit model, only the difference among the attractions determines their relative probabilities of being chosen. (A constant added to all the attractions divides out in the logit form.) Therefore, if attractions can grow and grow, as they can when $\phi > \rho$, 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 $\phi = \rho$, then convergence cannot produce choice probabilities which are so extreme.

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.[6] Additionally, they require attraction updating with $\delta = 1$ and $\phi = \rho$. 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.

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. The effect of $N(0)$ is easiest to see by fixing $\delta = 1$ for simplicity and directly computing the attraction after two periods, $A_i^j(2)$, which gives

$$A_i^j(2) = \frac{\phi^2 \cdot A_i^j(0) \cdot N(0) + \phi \cdot \pi_i(s_i^j, s_{-i}(1)) + \pi_i(s_i^j, s_{-i}(2))}{\rho^2 \cdot N(0) + \rho + 1}. \qquad (11)$$

The parameter ϕ captures the declining weight placed on payoffs from more distant periods of actual experience (that is, the period 1 payoff $\pi_i(s_i^j, s_{-i}(1))$ is weighted ϕ but the 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

[6] 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.

parameter $N(0)$ captures the special weight placed on the initial attractions, compared to increments in attraction due to actual (or foregone) 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.[7]

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.[8] Compared to the belief and reinforcement special cases, EWA fits better in weak-link coordination games (Camerer and Ho, 1999a) and predicts better out of sample in median-action coordination games and dominance solvable "p-beauty contests" (Camerer and Ho, 1999b), call markets (Hsia, 1998) and "unprofitable games" (Morgan and Sefton, 1998). In some constant-sum games, EWA predicts slightly worse than belief learning (Camerer and Ho, 1999b).

3.5 Extending EWA to Signaling Games

The first question to address in extending EWA to signaling games is what constitutes a strategy. In these games, we denote types by t_i, messages by m_j, and actions by a_k. The sender and receiver earn payoffs $\pi_S(t_i, m_j, a_k)$ and $\pi_R(t_i, m_j, a_k)$, respectively. Because senders observe their own types, it is appropriate to define their strategies conditional on observed types. There are two options.

First, one could define contingency strategies which specify a message for each type. For example, $(m_1|t_1, m_2|t_2)$ is a strategy in which the sender plays m_1 if t_1 is observed, and m_2 if t_2 is observed. This approach assumes that a sender chooses a complete strategy in each period (a strategy for each type), but only "uses" the portion which is relevant for their observed type. In games in which complete strategies are elicited this modeling approach seems reasonable. However, in the experiments we study complete strategies are not elicited. The complete-strategy approach then begs the question of how to update attractions for several complete strategies which have the same "used" portion but different "unused" portions. For example, suppose

[7] 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.

[8] *Nota Bene* forecasting out of sample completely removes any inherent advantage of EWA over restricted special cases due to having more parameters; indeed, if EWA fits well by overfitting, it will predict especially poorly out of sample.

the sender is t_1 and the chosen message is m_2. How does one update both $(m_2|t_1, m_1|t_2)$ and $(m_2|t_1, m_2|t_2)$?

We take a second approach, which is to assume that players have different strategy sets at each reachable node, which are not linked to form complete strategies. (This is similar to the "agent form" game in which each node is played by a different "agent" for a single player, and all the agents have the same payoff.) In the example above, we simply update the attraction on $m_2|t_1$ in the example (and perhaps also $m_1|t_1$, which the player could have chosen but did not) and leave attractions for the unused strategies, $m_1|t_2$ and $m_2|t_2$, unchanged. Similarly, we assume receivers have strategies which are conditional on the message they observed the sender choosing, but not on the sender's type. A receiver's strategy to choose action i in response to message j will be denoted $a_i|m_j$.

Initial attractions for t_1 senders are denoted $A^{m_1t_1}(0)$ and $A^{m_2t_1}(0)$, and for t_2 senders, the initial attractions are $A^{m_1t_2}(0)$ and $A^{m_2t_2}(0)$. (In the logit form one of the attractions in each pair must be fixed for identifiability.) The initial experience counts are $N_S^{m_1}(0)$ and $N_S^{m_2}(0)$.

For receivers who observe message m_1, initial attractions are $A^{a_1m_1}(0)$, $A^{a_2m_1}(0)$, and $A^{a_3m_1}(0)$. For receivers who observe message m_2, the initial attractions are $A^{a_1m_2}(0)$, $A^{a_2m_2}(0)$, and $A^{a_3m_2}(0)$. (One of the attractions in each triple must be fixed for identifiability). The initial experience counts are $N_R^{m_1}(0)$ and $N_R^{m_2}(0)$.

3.6 The Baseline Model

This section discusses how the EWA model presented in equations 1 and 2 can be adapted to signaling games. For receivers, this is a simple problem because they can condition only on the sender's message. Thus, the receivers know what their foregone payoffs are at the end of each period: they update their attraction to their chosen strategy with their realized payoff, and to other strategies with (δ times) the foregone payoff given by the actual type and message. If the receiver had chosen a_1 in response to m_1 when the sender was a t_2, for example, she would update according to:

$$N_R^{m_1}(t+1) = \rho \cdot N_R^{m_1} + 1 \tag{12}$$

$$A^{a_1m_1}(t+1) = \frac{\phi \cdot A^{a_1m_1}(t) \cdot N_R^{m_1}(t) + \pi_R(t_2, m_1, a_1)}{\rho \cdot N_R^{m_1}(t) + 1} \tag{13}$$

$$A^{a_2m_1}(t+1) = \frac{\phi \cdot A^{a_2m_1}(t) \cdot N_R^{m_1}(t) + \delta \cdot \pi_R(t_2, m_1, a_2)}{\rho \cdot N_R^{m_1}(t) + 1} \tag{14}$$

$$A^{a_3m_1}(t+1) = \frac{\phi \cdot A^{a_3m_1}(t) \cdot N_R^{m_1}(t) + \delta \cdot \pi_R(t_2, m_1, a_3)}{\rho \cdot N_R^{m_1}(t) + 1}. \tag{15}$$

Since she does not observe m_2, $N_R^{m_2}(t+1) = N_R^{m_2}(t)$ and $A^{a_km_2}(t+1) = A^{a_km_2}(t)$ for $k = 1, 2, 3$.

The senders' chosen strategies are updated according to the realized payoffs in the same way. However, with senders, it is more difficult to define foregone payoffs to unchosen strategies. There are two complications.

First, conditioning on a sender's type, the foregone payoff to the unchosen message is not known perfectly because it depends on the receiver's unobserved response. The sender knows the set of possible payoffs, but she does not know which payoff in the set would have resulted. Of course, this is generally the case in extensive-form games with unreached information sets. Below we consider several ways of choosing a foregone payoff in the set, or some mixture of those payoffs, to update the attraction on the unchosen message. In the baseline model, however, we simply leave attractions for unchosen sender messages unreinforced (and thus do not decay their experience counts).

The second complication is that belief models implicitly require that the attraction for the choice of the chosen message by the unrealized type also be updated by that type's foregone payoff. The reason is that senders are forming beliefs about expected reactions to chosen messages. How a receiver reacts when a t_1 sender chooses a message informs the sender's belief about the receiver's reaction when a t_2 sender chooses the same message, thus affecting t_2's expected payoffs (or in EWA terms, the attractions of t_2s strategies). This is reasonable because the sender knows the receiver's strategies may be message-dependent but cannot be type-dependent, so she knows exactly what the foregone payoff would have been if the unrealized type had chosen the same message the realized type did. For example, if a t_1 sender sends message m_1 and gets response a_1, she receives payoff $\pi_S(t_1, m_1, a_1)$ and updates $A^{m_1t_1}$ accordingly. But she also knows that if she had been a t_2 and chosen m_1, she would have earned $\pi_S(t_2, m_1, a_1)$. Therefore, the sender updates according to

$$N_S^{m_1}(t+1) = \rho \cdot N_S^{m_1} + 1 \tag{16}$$

$$A^{m_1t_1}(t+1) = \frac{\phi \cdot A^{m_1t_1}(t) \cdot N_S^{m_1}(t) + \pi_S(t_1, m_1, a_1)}{\rho \cdot N_S^{m_1}(t) + 1} \tag{17}$$

$$A^{m_1 t_2}(t+1) = \frac{\phi \cdot A^{m_1 t_2}(t) \cdot N_S^{m_1}(t) + \delta \cdot \pi_S(t_2, m_1, a_1)}{\rho \cdot N_S^{m_1}(t) + 1} \tag{18}$$

and $N_S^{m_2}(t+1) = N_S^{m_2}(t)$ and $A^{m_2 t_k}(t+1) = A^{m_2 t_k}(t)$ for $k = 1, 2$.

The updating rules for the receivers are relatively straightforward. However, the notion of a foregone type, giving up something over which one never had a choice can be confusing. The models we propose in section 3.7 are more complicated still because they suggest ways senders might update attractions for unchosen messages. Because updating rules for different combinations of realized and unrealized types and chosen and unchosen messages can get confusing, we will display update rules for senders in the game table. To illustrate the baseline update rule described above, we will use the following form:

	$\underline{\text{Type I}}$	Type II
	$\underline{a_1}$ a_2 a_3	a_1 a_2 a_3
$\underline{M_1}$	$\frac{\phi A_S^{m1t1}(t) N_S^{m1}(t) + \pi_S(m_1, t_1, a_1)}{\rho N_S^{m1}(t)+1}$	$\frac{\phi A_S^{m1t2}(t) N_S^{m1}(t) + \delta \pi_S(m_1, t_2, a_1)}{\rho N_S^{m1}(t)+1}$
M_2	$A_S^{m2t1}(t)$	$A_S^{m2t2}(t)$

Tabular representation of sender's baseline update rules.

The underlined labels indicate that these rules represent an example where the realized type is t_1, the chosen message is m_1 and the chosen action is a_1. There are four cells in the table, one for each strategy-information set combination to which the sender has an attraction. In each cell is the attraction update rule for that cell given the the realized type and chosen message and response. In this case, Equation 17 is represented in the upper left cell, where the sender increases her attraction with the full weight of the realized payoff. The upper right cell demonstrates how the foregone type is used: since the receiver's choice is message and not type dependent, the sender knows that had the type been t_2, she would have realized $\pi_S(m_1, t_2, a_1)$, but because this is only hypothetical, it is weighted by the imagination parameter, δ (Equation 18). The cells in the lower row do not have an update rule, indicating that the attractions are just copied from one period to the next; there is no new information. To further simplify presenting the update rules, the denominator of the cells indicates how the experience counts are updated. For the baseline rule, the experience count of the chosen message is updated according to Equation 16, and the experience count of the unchosen message is simply copied into the next period.

3.6.1 Special Cases of EWA

The choice reinforcement and belief-based special cases of EWA discussed above apply, without much modification, to this adaptation of EWA for signalling games. The reinforcement model is still realized if $\delta = 0$, $\rho = 0$ and $N_R^{m_1} = N_R^{m_2} = N_S^{m_1} = N_S^{m_2} = 1$. However, the extension of the belief-based model is less obvious because it requires estimating initial belief counts rather than initial attractions.

In addition to setting $\delta = 1$ and $\phi = \rho$, we implement the belief model's implicit constraints on the $A(0)$s by estimating them indirectly: we estimate belief counts for each of the opponent's strategies and computing $A(0)$s by using these estimates to compute expected values. Thus, for the sender we estimate $N^{a_1 m_1}(0)$, $N^{a_2 m_1}(0)$, $N^{a_3 m_1}(0)$, which must sum to $N_S^{m_1}(0)$ and $N^{a_1 m_2}(0)$, $N^{a_2 m_2}(0)$, $N^{a_3 m_2}(0)$, which must sum to $N_S^{m_2}(0)$, and for the receiver we estimate $N^{m_1 t_1}(0)$, $N^{m_1 t_2}(0)$ which sum to $N_R^{m_1}(0)$ and $N^{m_2 t_1}(0)$, $N^{m_2 t_2}(0)$, which must sum to $N_R^{m_2}(0)$.

3.7 Unchosen Message Models

The appeal of the baseline model is that the sender is making all valid inferences: the receiver would have chosen the same action had the type been different, so the sender knows what her exact payoff would have been in the unrealized type case. However, the baseline model does not build in an answer to the sender's natural question, "Did I choose the right message, or should I have chosen the other message, given my realized type?" The alternative models presented here consider the possibility that a sender tries to force an answer to that question, using various imperfect inferences about what her payoff would have been had she chosen the other message.[9]

3.7.1 The Median Payoff

The sender knows the payoff to the unchosen message (conditional on the realized type) would have been one of three numbers. One way the sender may assign a foregone payoff is to use some statistic which is representative of

[9] Another model is that the sender assumes the receiver would have chosen the same (observed) action even if the sender had sent the other message. This neglects the sender's knowledge that the receiver's action choices could be message-dependent. It seems unlikely to fit the data better so we have not investigated it empirically.

those three numbers. The median is one such statistic. It is computationally simple, and more robust than say, the minimum or the maximum of the payoffs.

Implementing alternate message updating requires a number of adjustments to the update rule. Let $MED(\pi_S|m_2, t_2)$ denote $MED(\pi_S(m_2, t_2, a_1)$, $\pi_S(m_2, t_2, a_2)$, $\pi_S(m_2, t_2, a_3))$. Then rules used for each cell are represented in the table below.

	Type I	Type II
	$\underline{a_1}$ a_2 a_3	a_1 a_2 a_3
$\underline{M_1}$	$\frac{\phi A_S^{m1t1}(t)N_S^{m1}(t)+\pi_S(m_1,t_1,a_1)}{\rho N_S^{m1}(t)+1}$	$\frac{\phi A_S^{m1t2}(t)N_S^{m1}(t)+\delta\pi_S(m_1,t_2,a_1)}{\rho N_S^{m1}(t)+1}$
M_2	$\frac{\phi^\tau A_S^{m2t1}(t)N_S^{m2}(t)+\mu_1\tau MED(\pi_S\|m_2,t_1)}{\rho^\tau N_S^{m2}(t)+\tau}$	$\frac{\phi^\tau A_S^{m2t2}(t)N_S^{m2}(t)+\mu_2\tau MED(\pi_S\|m_2,t_2)}{\rho^\tau N_S^{m2}(t)+\tau}$

Tabular representation of sender's median model update rules.

The second row of the table gives the update rules for the unchosen message, for both the realized and unrealized type. The new parameters μ_1 and μ_2 can be interpreted similar to δ; they represent the weight, or vividness of imagination, used in updating the attractions to the unchosen message for realized and unrealized types, respectively.

The other new parameter, τ, allows for the possibility that updating an unchosen message by the median foregone payoff does not have as much psychological impact as updating chosen messages, and hence is not the same as a single period of 'real' experience. In addition to being the increment to the unchosen message experience counter, τ is also an exponent of ϕ and ρ for unchosen messages and it multiplies μ_1 and μ_2. These additional appearances of τ in the updating equation allow it to be interpreted as the fraction of a period's experience in the unchosen message gained in conjecturing about and updating the unchosen message attraction. To see this, imagine that $\tau = 1$. The unchosen message rules reduce to the chosen message rules with δ equal to μ_1 and μ_2 for the realized and unrealized types respectively. On the other hand, if $\tau = 0$, the unchosen message attractions are not discounted and no payoff is added to them, so they are unchanged.

3.7.2 Convex Combinations of Minimum and Maximum Payoffs

Another way to update foregone payoffs is to take a convex combination of the minimum and maximum possible payoffs. Because the weights used can vary from game to game, the model need not be sensitive to extremely high or extremely low outlier payoffs, yet it can be more robust to attractive payoffs than the median rule. For example, if players frequently switch to unchosen messages, their 'grass-is-greener-on-the-other-side' switching could be captured by assuming they are optimistically putting a lot of weight on the maximum foregone payoff. Alternatively, if their message switching is slow their inertia could be modeled by assuming they are pessimistically putting a lot of weight on the minimum foregone payoff.

Let $\Pi(m_2, t_1) = \alpha MIN(\pi_S|m_2, t_1) + (1-\alpha)MAX(\pi_S|m_2, t_1)$. Then the convex combination model can be written

	$\underline{\text{Type I}}$	Type II
	$\underline{a_1}$ a_2 a_3	a_1 a_2 a_3
$\underline{M_1}$	$\frac{\phi A_S^{m1t1}(t)N_S^{m1}(t)+\pi_S(m_1,t_1,a_1)}{\rho N_S^{m1}(t)+1}$	$\frac{\phi A_S^{m1t2}(t)N_S^{m1}(t)+\delta\pi_S(m_1,t_2,a_1)}{\rho N_S^{m1}(t)+1}$
M_2	$\frac{\phi^\tau A_S^{m2t1}(t)N_S^{m2}(t)+\mu_1\tau\Pi(m_2,t_1)}{\rho^\tau N_S^{m2}(t)+\tau}$	$\frac{\phi^\tau A_S^{m2t2}(t)N_S^{m2}(t)+\mu_2\tau\Pi(m_2,t_2)}{\rho^\tau N_S^{m2}(t)+\tau}$

Tabular representation of sender's convex combination model update rules.

This model is implemented just like the median model, except for the addition of the parameter α, which is the fraction of weight put on the minimum.

3.7.3 Mirror Sophistication: Internal Models of Other Players

A final model of unchosen message foregone payoff formulation assumes that players use all information available to them, including using their own behavior as a proxy for others'. Since subjects played both roles in the course of the experiment, it is not necessary for senders to use a rule of thumb to guess about the receiver's response—a sender can appeal to the attractions of her receiver alter-ego's actions to compute the probability of each action response to the unchosen message. Her expected payoff from the unchosen message will be the expected payoff from playing somebody like herself. We call this 'mirror sophistication' because players form a guess about what a player in another role will do by looking in a proverbial mirror at their own behavior when they were in that role.[10]

[10] Obviously, this rule will be sensitive to the experimental protocol and does not apply if the players do not switch rules. We regard this protocol-sensitivity as an advantage. There is a strong intuition among experimentalists that players do learn faster when they switch roles, which is supportive of such a rule. This is easily testable, by comparing experiments with different degrees of role-switching.

Let $p_S(a_1|m_1)$ denote the probability with which the sender's receiver alter-ego would choose a_1 given the message m_1. Let $\Pi(m_2, t_1) = \sum_{j=1}^{3} p_S(a_j|m_2)\pi_S(m_2, t_1, a_j)$. Then the simple sophistication model can be expressed as in the table below.

	Type I	Type II
	a_1 a_2 a_3	a_1 a_2 a_3
M_1	$\frac{\phi A_S^{m1t1}(t)N_S^{m1}(t)+\pi_S(m_1,t_1,a_1)}{\rho N_S^{m1}(t)+1}$	$\frac{\phi A_S^{m1t2}(t)N_S^{m1}(t)+\delta\pi_S(m_1,t_2,a_1)}{\rho N_S^{m1}(t)+1}$
M_2	$\frac{\phi^\tau A_S^{m2t1}(t)N_S^{m2}(t)+\mu_1\tau\Pi(m_2,t_1)}{\rho^\tau N_S^{m2}(t)+\tau}$	$\frac{\phi^\tau A_S^{m2t2}(t)N_S^{m2}(t)+\mu_2\tau\Pi(m_2,t_2)}{\rho^\tau N_S^{m2}(t)+\tau}$

Tabular representation of sender's mirror sophistication model update rules.

The parameters are interpreted exactly as in the two previous models. However, this model departs slightly from the spirit of EWA because this implementation of mirror-sophistication implies a belief-based interpretation of attractions. In standard EWA, the learner never directly asks herself what her opponent will do in order to best respond, as she does when using a belief-based model. In this model, however, the sophisticated learner does ask herself what she believes her opponent will do. Although the beliefs are still determined by EWA, this makes the mirror-sophistication model incompatible with a reinforcement interpretation. This is not particularly surprising because reinforcement learning uses only realized payoff streams, holding no direct role for how opponents adapt.

4 Experimental Results

In order to test our baseline adaptation of EWA, its choice reinforcement and belief-based special cases, and our unchosen message updating extensions, we use Brandts and Holt's games 3 and 5. However, while the 12 periods of data on 24 subjects they generated is sufficient to grasp the intuition behind the Brandts and Holt story, estimating a structural model as complex as EWA, and distinguishing it from special cases, requires more statistical power. Therefore, we replicated Brandts and Holt's (1993) games 3 and 5 with 32 subjects playing 32 periods.

For our replication, we recruited Caltech undergraduates who did not necessarily have any training in economics, although many had participated in other experiments. We used a standard signaling game software which presented the game table as in Tables 1 and 2.[11] In each period, the senders were randomly selected, informed of the type and prompted for their message. When all senders had selected a message, receivers were notified of their paired sender's choice and were asked to chose a response. They knew that each type was equally likely *ex ante*. At the end of each period the realized cell of the payoff table was highlighted and subjects wrote down their payoffs. There were four cohorts of eight subjects, and we used a counterbalanced design, so two cohorts played Game 3 first and two played Game 5 first. Subjects earned an average of about $27 in about two hours, and were paid in cash as they left the laboratory. The only protocol difference between our experiment and Brandts and Holt's is that our pairings were random, with replacement; we made no attempt to ensure subjects did not play subjects they had played previously.

Figure 2 presents the data from our experiment, averaged across sessions in 4-period blocks. The results in Game 3 replicate BH closely, and confirm that with more experience, play converges reasonably sharply to the intuitive sequential pooling equilibrium at m_1, supported by action responses a_1 and a_2 to the two messages.

The Game 5 results are a little more surprising. We thought additional periods might cause t_2s to choose m_1 less and less frequently, cementing convergence to the unintuitive equilibrium at m_2. However, additional periods do not eliminate the separation between messages. While the patterns are the same, the senders' strategies during the 13^{th} through 32^{nd} periods looked much like the 9^{th} through 12^{th} periods of the Brandts and Holt data, suggesting the convergence to the sequential equilibrium is not complete, even after many more periods of learning.[12] However, there is also no evidence of movement back toward the intuitive equilibrium at m_1.[13]

[11] The software we used also had a third message, which we instructed subjects never to use (they were compliant).

[12] We also conducted a session with 64 periods of Game 5 only, to see if longer-run convergence was different than what we observed in only 32 periods. There was no additional movement toward either equilibrium.

[13] To test that zero-aversion was preventing some subjects from switching to the unintuitve equilibrium, we ran eight subjects with a payoff table which added 15 to each payoff in Game 5 (call this Game 5'). These subjects converged to the unintuitive equilibrium in about 50 periods. However, we were concerned that the unintuitive equilibrium payoff of 105 may have been focal in that experiment, so we ran two more sessions on a payoff table which multiplied payoffs in Game 5' by 4/5. This behavior was indistinguishable from that of players in Game 5.

5 Estimation

In these models, the initial attractions, experience counts and model parameters can be estimated from our experimental data. We computed the maximum likelihood parameter estimates for each model using the constrained maximum likelihood procedure in Gauss (Aptech).[14] To simplify the estimation and make the models easily interpretable, we impose a number of restrictions on the parameter space. First, we impose bounds on the initial attractions, so that the set of possible attractions is not much larger for the EWA model than for the belief models (whose attractions are closely tied to the payoff structure).[15]

The second restriction is that message-specific experience weights should be the same for senders and receivers. That is, $N_S^{m_1}(0) = N_R^{m_1}(0)$ and $N_S^{m_2}(0) = N_R^{m_2}(0)$. While there is no *a priori* reason we think this is so, tests across a broad sample of data indicate that this is not a statistically significant restriction, and it saves two degrees of freedom.

Our final restriction is that each of the $N(0)$s must all be less than 50 and less than $\frac{1}{1-\rho}$. This prevents the model from putting so much weight on the initial attractions that there is almost no effect of the experience gained in the

[14] To ensure we found the peak of any local maximum we located, we used a two-step search process. From a given starting point, we used the Bernt, Hall, Hall and Hausman algorithm to search the parameter space. This algorithm estimates the Hessian, rather than calculates it exactly like Newton methods, and thus finds maxima quite quickly. However, it is not always precise. From the maximum found by the BHHH algorithm, we applied Gauss's version of the Newton gradient ascent algorithm. Using an exact (numerical) Hessian, this second algorithm often produced small improvements in fit. To ensure that the local maxima we found were global maxima, we tested a variety of starting points. We found the parameter space to be surprisingly well-behaved: in each model all of our starting points converged to the same maximum, suggesting that our estimates are in fact global maxima. The Gauss and C code used to estimate parameters is available from the first author.

[15] We look at the set of possible payoffs given the information available at the time of move and bound each initial attraction to be between the minimum and maximum attainable payoffs for each strategy. For instance, in Game 5, t_1 senders can earn payoffs $\{45, 0, 0\}$ from m_1, so $A^{m1t1}(0)$ must be in the interval $[0, 45]$ and m_1 receivers can earn $\{45, 30\}$ from a_1, so $A^{a1m1}(0)$ must be in $[30, 45]$. Because one of each type or message conditional strategy must be a constant in the logit form, we restrict one of the strategies in each information set to have an initial attraction equal to the minimum attainable payoff. There is no way to determine which strategy should have its attraction set to its minimum, so we estimated all possible combinations of these restrictions, and report only the one that yielded the best fit.

play of the game. The restriction $N(0) \leq \frac{1}{1-\rho}$ forces the experience weight to increase, which means that new payoff information is getting less and less weight compared to lagged attractions; subjects do not have less perceived experience after playing the game than they brought into the game. Note that this restriction also requires $0 \leq \rho \leq 1$ (for positive N(0)).

One of the disadvantages of imposing these restrictions is that it compromises the asymptotic normality of the maximum likelihood parameter estimates. While we might already find this assumption suspect because of our modest sample size ($N = 32$), imposing binding bounds restrictions makes the confidence intervals constructed from standard errors computed from the variance-covariance matrix unreliable. Because values outside the bound are never observed, the sample variance will be biased downward, and the confidence interval constructed from the standard error may not lie entirely within the imposed interval. Therefore, we construct bootstrapped confidence intervals using the percentile method.[16]

This approach to estimation and inference requires a lot of computing time, patience and attention to detail. Our objective in taking this intensive approach is to set a high methodological standard so the learning model discourse can focus on theoretical issues, rather than methodological shortcuts which can cloud conclusions. Often in learning we are trying to tease out small but significant effects, the presence or absence of which demonstrate the supremacy of one model or another. In using the best known estimators (maximum likelihood) and strong statistical tests (AIC and BIC model calibration measures) rather than easier to compute but ad hoc statistics, we create standard for model comparison. This standard will facilitate a direct comparison of alternative models, rather than alternative methodologies.

5.1 Fitting the Baseline Model

The objective of this paper is to test several models of how people update unchosen messages and unrealized types in signalling games. The focal point

[16] This nonparametric technique requires performing maximum likelihood estimates on a large number B data sets, where each data set is the sample with each subject weighted by a Poisson(N) random number (Aptech 1995, 31). This process gives us B estimates of each parameter, and the 95% confidence interval for a parameter is given by that parameter's 2.5^{th} and 97.5^{th} order statistics. This method allows us to present the correct confidence intervals without knowing the transformation which would make the actual error distribution normal (Efron and Tibshirani 1993, 171).

of this study is the baseline EWA model described in Section 3.6. First, we test the baseline model against models which are simpler, the choice reinforcement and belief-based special cases of EWA.

Table 3 presents the parameter estimates of EWA and its choice reinforcement and belief-based special cases for Game 3. The predictions they generate are shown in Figure 3. These are done on the first 24 periods of our data; the last eight periods are a holdout sample we try to predict. The most significant feature of the Game 3 data is that there is relatively little variance in the frequency of play of different strategies. The strategies $m_1|t_1$ and $a_1|m_1$ are played with virtually constant frequency throughout the game, $a_2|m_2$ is highly variable, but has no real trend and $m_1|t_2$ shows a steady increase. The parameter estimates show this lack of variance in the large initial experience counts and the depreciation parameters close to one. $\hat{N}^{m1}(0)$, in particular, achieves its maximum value, reflecting the relative stability of m_1 play. This stability is reinforced by $\hat{\phi} > \hat{\rho}$ which means that past attractions are amplified, so attractions are not bounded by payoffs and this convergence can be quite sharp, as it is with the m_1 data.

Table 4 presents a number of goodness of fit statistics which we use to compare models. The first row presents the average per period log-likelihood (summed across subjects) for the first 24 periods. This is the number that was minimized in estimation. The second row presents the same statistic for the 25^{th} through 32^{nd} periods, the holdout sample into which we hope to predict. These statistics can be used to compare models, but there is no penalty for extra degrees of freedom. The third rows presents the Akaike information criterion (AIC) and the fourth row presents the Bayesian information criterion (BIC).[17] These statistics can be compared directly and used for model calibration; they are designed to reach a maximum value at an optimal tradeoff between improvement of fit and additional parameters, even when models are non-nested.

[17] The AIC is the total in-sample log-likelihood minus the number of model parameters, divided by the number of sample periods (24). It is widely used for model comparison, but not motivated by any optimality considerations. The BIC is the total in-sample log-likelihood minus the half the number of model parameters times the natural log of the number of observations, divided by the number of sample periods (24). Under certain regularity conditions (which are not satisfied if parameters are either estimated on or restricted to a boundary), the BIC can be interpreted as follows: if model i has a higher BIC than j, then $\exp\{-24 * (BIC_i - BIC_j)\}$ is an approximation to the posterior odds ratio, $\Pr(BIC_i)/\Pr(BIC_j)$, of a Bayesian observer with equal priors (Carlin and Louis, 1996).

The second section of Table 4 presents some alternative measures of fit. The first two rows present the in and out of sample miss rate. The miss rate is the percentage of the time the strategy that is predicted most likely to be chosen by the model is not selected by the experimental subject. The second two rows give the average per period mean squared deviation.[18]

The columns of Table 4 all represent models discussed in this paper, except for the last one. The last column is a model we propose for comparison. It is determined by taking as the model prediction the frequency of play of each strategy throughout the 24-period calibration sample; its predictions are horizontal lines determined by the data.

Comparing choice reinforcement's miss rate and MSD to those of EWA and the baseline, we can say it does not fit the data well; it barely does better than a horizontal line. It does not fit the data well because it does not use foregone payoff information. Figure 3 shows that reinforcement mistakenly predicts t_2 senders move slightly away from m_1, when in fact they move strongly toward it. The reason is that m_2 is usually met with the action response a_2, so it yields a payoff of 15 for t_2s. This payoff reinforces that choice positively and leads them to choose it again, moving them away from m_1. But in EWA and belief learning, what the t_1s learn from choosing m_1 influences the attractions for t_2 (through updating of the unrealized type attractions). Then t_2s gradually learn that message m_1 would pay 30, which is better than 15, and this indirect learning moves them toward m_1. As a result, we can see the AIC and BIC both suggest that the additional parameters of EWA are more than justified by the improvement in fit, although the BIC also reflects the flatness of the data, suggesting that EWA does not represent a significant improvement over the baseline.[19]

The belief-based special case fits much better than the choice reinforcement model, but still not as well as EWA. Again, the belief model fails to

[18] This is calculated by creating, for each subject in each period, a vector with length equal to the number of strategies. The strategy chosen by the subject in that period is assigned a 1, and all others are zero. The MSD is the sum of squared differences between the created vector and the corresponding vector of choice probabilities predicted by the model, averaged across subjects and periods (but not across strategies).

[19] Note that the conventional way to make this comparison is a χ^2 test. We do not use it because the fact that some parameters are estimated and/or restricted to be on their boundaries violates the assumptions of the Central Limit Theorem necessary to show that $2(LL_i - LL_j)$ has a χ^2 distribution. However, the bootstrapped CIs from the model parameters which significantly influence fit (δ, ρ, ϕ) suggest different conclusions are very unlikely.

adequately track the increasing frequency of play $m_1|t_2$. The problem is that the large values of $\hat{\phi} = .98$, along with large initial experience counts, means it takes a lot of experience to alter the t_2 sender's beliefs, so the belief model does not allow learning which is fast enough. This subtle point also illustrates why we wanted to apply EWA to these type of data. The original BH story about belief formation is not precise about strength of prior, fictitious play weight, and other parameter values.[20] By estimating EWA one is forced to be very precise about the details of the model. It may not be possible to find configurations of parameters which can fit the initial conditions, the basic trend, and also get the speed of convergence right. The sluggish belief learning of $m_1|t_2$ in Game 3 shows that while the belief account gets the direction right, it does poorly on convergence speed, and adding EWA flexibility improves the fit a lot.

Table 5 presents the parameter fits for Game 5. Figure 4 shows the predictions they generate. Unlike Game 3, there is a significant trend to track in all the information sets. Look first at the central parameters, δ, ϕ and ρ. The estimated $\hat{\delta} = .54$, which is consistent with previous findings in games of complete information. This suggests that senders update the foregone type about half as much as they do their realized type. The depreciation parameters are also well within the range found in complete information games, and $\hat{\phi} > \hat{\rho}$, which means that attractions are growing over time, and are not bounded by payoffs.

The initial experience counts are only .62 and 3.37. Together with depreciation parameters much less than one, low experience counts mean initial attractions are fairly quickly swamped by the experience gained in the play of the game. The initial attractions for the sender suggest the observed initial type dependence, and suggest receivers should respond to m_1 with a predominance of a_2, and to m_2 with about equal frequencies of a_1 and a_3. In this environment, where m_2 comes mostly from t_1s, a_3 has a lower expected utility for the receiver a_1, but it has the appeal of equity, which may be particularly strong in the case when subjects must switch roles from period to period.

The choice reinforcement model discounts previous attractions at about the same rate as the EWA model, but the restrictions on δ, ρ and the $N(0)$s have significant impact on the model's fit. The primary feature of the sender's

[20] In their entry games, Cooper et al. (1997a,b) do specify parameter values, as do Brandts and Holt in 1994.

data that reinforcement does not fit is the gradual decrease in the frequency of play of m_1 given t_2 (until the sharp jump in the last block). Since m_1 gets reinforced by 30 for t_2, it is getting strongly reinforced. As in Game 3, the t_1 choices of m_2 demonstrate to players that if t_2s were to switch to m_2, they might get 45; this unrealized type reinforcement helps explain why they switch. By leaving unrealized type reinforcement out, choice reinforcement cannot account for the basic trend in $m_1|t_2$. Similarly, it is too slow to adjust to the initial decrease and subsequent increase in the frequency of a_2 given m_1, so it must fit the initial periods with an essentially smooth function.

We can compare this model with EWA using Table 6. Because it adapts too slowly to account for most major features of the data, the AIC and BIC tell us that the additional parameters of EWA are more than justified by the corresponding improvement in fit.

The belief-based model does a better job of capturing the gradual decrease and sudden increase in the frequency of play of m_1 given t_2; it does almost as well as EWA. However, it predicts essentially constant play for the a_2 in response to m_1, and an essentially constant rate of increase for a_1 in response to m_2. This mirrors our findings for Game 3: the belief-based model, which formalizes the BH dynamic, gets the direction right, but adding the flexibility of EWA substantially improves tracking of convergence.

5.2 Fitting the Unchosen Message Models

The estimates show that EWA is not too complicated (and that reinforcement and belief models are too simple), because the baseline EWA model offers significant improvement over the special cases. Now we look at the more complicated alternative models of unchosen message reinforcement. We do this to discover whether or not our EWA model fails to capture any significant aspect of the sender's behavior.

Table 7 presents the parameter estimates for Game 3 for the three alternative models we consider here, and Figure 5 presents the predictions they generate. The most striking feature of the table is that, while they each offer significant improvement over EWA, they all have similar parameter estimates and offer similar fits. There is a significant feature of senders' behavior EWA is not capturing, and all of these models capture it equally well.

The alternative models have similar initial attractions to EWA, but the values of the experience counts switches, so $\hat{N}^{m2}(0) > \hat{N}^{m1}(0)$, and the operational parameters have a different relationship. Unlike in EWA, $\hat{\phi} < 1$, so the initial attractions are depreciated as the game is played. The simulated experience increment, $\hat{\tau} \approx 1.07$, indicates that a period of simulated experience is worth slightly more than a period of actual experience. In the convex combination model, the $\hat{\alpha} = 0$ means that all the weight is put on the maximum, which is 45 in most information states; the median is 30 in all information states, which means that the lower value of $\hat{\mu}_2$ for the convex combination model makes the payoff's effect on the attraction about the same in each model.

That $\hat{\phi} < 1$ suggests that the unchosen message models are more responsive to payoffs than EWA. This can be seen in the first 16 periods of the $m_1|t_2$ series, where EWA makes a very flat prediction but the data start below the EWA prediction and increase. EWA must make a flat prediction, with a high initial experience count and no discounting, because the low payoffs typically received from selecting $m_2|t_2$ (almost always 15) are inconsistent with the slow rate of switching to $m_1|t_2$. The unchosen message models, however, slow response to low payoffs from $m_2|t_2$ by increasing the attraction to $m_2|t_2$ whenever m_1 is chosen in response to t_1, which it almost always is. In this case, the attraction to $m_2|t_2$ is updated, in the median model, with a payoff of $0.68 \cdot 30 > 15$. Thus the attraction to $m_2|t_2$ is slightly increased (on average) about half the time (when the type is t_1), only slightly decreased when the subject chooses $m_2|t_2$ (when the unchosen message models are also able to slow switching by updating $m_1|t_2$ with a payoff of 0 because $\mu_1 = 0$), and significantly decreased once the subject switches to $m_1|t_2$. This pattern matches the data more closely than the EWA predictions.

The unchosen message model estimates for Game 5 are presented in Table 8, and the predictions they generate in Figure 6. As with Game 3, the behavior of each of Game 5's alternative models is similar to EWA, and very similar to one-another. This again suggests that if there is some significant pattern in the data not captured by EWA, these three models capture it in the same way. What is less clear than in Game 3, however, is that the improvement in fit from modeling the unchosen messages is worth the extra degrees of freedom. The AIC suggests the trade-off is worth it, but the BIC concludes the model is not worth the additional parameters.

Unlike Game 3, where reinforcing the unchosen message allowed the model of predict turning points in the data that the relatively flat EWA prediction did not, EWA predicts most of the fluctuation in the data in Game 5. Therefore there is no pattern in the data not captured by EWA which we can examine to determine how the unchosen message models achieve their improved fit; the unchosen message reinforcement models are just a little closer to the data at several points. The only place where this difference is substantial is the second half of the sample where EWA is not able to track the rapid decrease in the frequency of $m_1|t_2$. This is a key feature of the data because this decrease represents a move toward the sequential equilibrium.

The unchosen message models are able to achieve better fit here because they can decrease the attraction to $m_1|t_2$ every time m_2 is chosen. In particular, the attraction to $m_1|t_2$ is decreased each time the player is a t_1 sender (since t_1 senders almost always choose m_2). Thus, as experience is gained, the frequency of $m_1|t_2$ decreases, which causes further, more dramatic, decreases in the attraction to $m_1|t_2$ because $\mu_1 = 0$. EWA cannot decrease this attraction whenever m_2 is played, so it cannot decrease fast enough to track the data. As in Game 3, the unchosen message models improve on EWA in much the same way as EWA improved on the belief-based models: the additional consideration of the payoff table better captures the speed of convergence as well as the direction.

6 Discussion

Our first objective in this paper was to replicate Brandts and Holt's results. We closely replicated their results. However, the additional periods we ran demonstrated that the convergence in Game 5 is slower than expected, and even 64 periods is not enough to realize equilibrium.

Using these data, we tested our adaptation of EWA to signalling games. Our baseline model reinforcement the foregone payoff for a sender's unrealized type. This allows the sender to make all valid inferences given that receivers are most likely playing message-contingent strategies. This model performed significantly better than its choice reinforcement and belief-based special cases. The belief-based case is of particular interest because it formalizes the BH dynamic. Our results indicate that while the BH dynamic captures the direction of the frequency trends, the formal belief-based restrictions do not allow the model to converge at the same rate as the data. This is particularly pronounced in Game 5, where long run simulations (50,000 periods) using Median players suggest that play never converges to either pure strategy equilibrium.

Although EWA performs better than its special cases, it may also be that EWA itself is too simple. Looking at the results from both games, updating unchosen messages does improve upon EWA's ability to fit the data and to predict out of sample. In developing the alternative models, we expected to capture a few specific features of the subjects' learning process. One such feature is the relative size of imagined experience, represented by τ. Since the unchosen message is updated, it is necessary to update its experience count as well. Because this experience is a result of the learner's conjecture, we hypothesized it is less valuable than actual experience. This was weakly supported, as τ is less than one in Game 3 and about one in Game 5.

A second feature we hoped to capture was the imagination coefficient on the realized type-unchosen message payoff and unrealized type-unchosen message payoff. We expected them to have a multiplicative effect: μ_1 requires only one level of counterfactual reasoning, but μ_2 requires two, suggesting μ_2 would be on the order of $\mu_1 \cdot \delta$. This expectation is not realized in our estimates, however, as μ_1 is zero in both games and μ_2 is greater than zero. This result is surprising because it implies that imagination is not necessarily nested: senders will go through two counterfactuals without learning from one. It also means that senders do not try to answer the natural question of whether or not they would have done better by choosing the other message.

Finally, we hoped to gain some insight into how subjects reinforce unchosen messages. The three unchosen message models we examine produce essentially similar fits on the two games we have examined. Because of its extra parameter, we conclude the convex combination payoff model is inferior to the median and sophisticated payoff models (indeed its AIC and BIC are higher for both games). The other two models are not statistically distinguishable on the two games we have examined. Thus, while we have been able to determine that senders do update the unchosen message attractions and the unchosen message experience counts with values less than one, we have not been successful in explaining what determines the value added to the attractions of the unchosen message. Based on these results, one might use any of the three unchosen message models and expect to do adequately.

References

[1] Aptech. *Constrained Maximum Likelihood*, March 1995.

[2] Jeffrey Banks, Colin F. Camerer, and David Porter. An experimental analysis of Nash refinements in signaling games. *Games and Economic*

Behavior, 6(1):1–31, January 1994.

[3] Jordi Brandts and Charles Holt. An experimental test of equilibrium dominance in signaling games. *American Economic Review*, 82(5):1350–65, December 1992.

[4] Jordi Brandts and Charles Holt. Adjustment patterns and equilibrium selection in experimental signaling games. *International Journal of Game Theory*, 22(3):279–302, 1993.

[5] Jordi Brandts and Charles A. Holt. Naive bayesian learning and adjustment to equilibrium in signaling games. *University of Virginia Department of Economics working paper*, April 1994.

[6] G. Brown. Iterative solution of games by fictitious play. In *Activity Analysis of Production and Allocation.* John Wiley & Sons, New York, 1951.

[7] Colin Camerer and Teck Ho. EWA learning in games: Heterogeneity, time-variation, and probability form. *Journal of Mathematical Psychology*, 42:305–326, June 1998.

[8] Colin F. Camerer and Teck Hua-Ho. Experience-weighted attraction learning in games: Estimates from weak-link games. In D. Budescu, Ido Erev, and R. Zwick, editors, *Games and Human Behavior*, pages 31–52. Lawrence Erlbaum Associates, Mahwah NJ, 1999a.

[9] Colin F. Camerer and Teck Hua-Ho. Experience-weighted attraction learning in normal form games. *Econometrica*, 67(3), April 1999b.

[10] Bradley Carlin and Thomas Louis. *Bayes and empirical Bayes methods for data analysis.* Chapman and Hall, New York, 1996.

[11] In-Koo Cho and David Kreps. Signaling games and stable equilibria. *Quarterly Journal of Economics*, 102(2):179–221, 1987.

[12] David Cooper, Susan Garvin, and John Kagel. Signalling and adaptive learning in an entry limit pricing game. *RAND Journal of Economics*, 28(4):662–83, 1997a.

[13] David Cooper, Susan Garvin, and John Kagel. Adaptive learning vs. equilibrium refinements in an entry limit pricing game. *Economic Journal*, 107(442):553–75, 1997b.

[14] A. Cournot. *Researches in the Mathematical Principles of the Theory of Wealth.* Haffner, London, 1960. Translated by N. Bacon.

[15] Bradley Efron and Robert Tibshirani. *An Introduction to the Bootstrap.* Chapman & Hall, New York, 1993.

[16] Robert Gibbons. *Applied game theory for economists.* Princeton University Press, Princeton, NJ, 1992.

[17] Calvin Harley. Learning the evolutionarily stable strategy. *Journal of Theoretical Biology*, 89:611–633, 1981.

[18] David Hsia. Learning in call markets. *USC Dept Economics*, 1998.

[19] John Morgan and Martin Sefton. An experimental investigation of unprofitable games. *Princeton University Woodrow Wilson School manuscript*, September 1998.

[20] Alvin Roth and Ido Erev. Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior*, 8:164–212, 1995.

[21] Jean Tirole. *The Theory of Industrial Organization.* MIT Press, Cambridge, MA, 1988.