Economic Theory **16**, 689–718 (2000)

Economic Theory
© Springer-Verlag 2000

Experience-weighted attraction learning in sender-receiver signaling games*

Christopher M. Anderson and Colin F. Camerer

Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125, USA (e-mail: cma@hss.caltech.edu)

Received: April 26, 1999; revised version: April 25, 2000

Summary. We apply Camerer and Ho's experience-weighted attraction (EWA) model of learning to extensive-form signaling games. Since these games often have many equilibria, logical 'refinements' have been used to predict which equilibrium will occur. Brandts and Holt conjectured that belief formation could lead to less refined equilibria, and confirmed their conjecture experimentally. Our adaptation of EWA to signaling games includes a formalization of the Brandts-Holt belief formation idea as a special case. We find that the Brandts-Holt dynamic captures the direction of switching from one strategy to another, but does not capture the *rate* at which switching occurs. EWA does better at predicting the rate of switching (and also forecasts better than reinforcement models). Extensions of EWA which update unchosen signals by different functions of the set of unobserved foregone payoffs further improve predictive accuracy.

Keywords and Phrases: Learning, Game theory experiments, Signaling games, Equilibrium refinement.

JEL Classification Numbers: C72, C92.

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

* 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) and Texas (Austin), Ohio State University, the Fall 1998 ESA Meetings, and from guest editor Charles Noussair and an anonymous referee.
Correspondence to: C.M. Anderson

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.

In this paper we apply Camerer and Ho's (1999) experience-weighted attraction (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 signaling product quality by price and advertising, strategic delay in strikes, excess capacity building by firms, choice of insurance policies, entry and pricing in monopolistic markets, incentives and personnel policies in labor markets with hidden action and hidden information, "money-burning" models of gift-giving, and many more (see, for example, Tirole, 1988, and Gibbons, 1992).

Signaling games are especially interesting because they often have many equilibria. The equilibria are theoretically distinguished by a variety 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 our discussion, we present an example of how historical circumstances may lead all firms to offer dividends, a possible unintuitive equilibrium.

Experimental tests of refinement concepts have yielded somewhat pessimistic results about the ability of refinements, even fairly simple ones, to predict to which equilibria experimental subjects will converge (e.g., Banks, Camerer and Porter, 1994). Brandts and Holt (1992, 1993) suggest players in their experiments are instead using a particular learning process (essentially a form of belief learning) (see also Cooper, Kagel and Garvin, 1997a,b). They observed that players were led to an equilibrium which violated the Cho-Kreps 'intuitive criterion'. During equilibration, players left empirical 'footprints' at all information sets by choosing strategies which later turned out to be rarely chosen. When behavior eventually crystallized around the (unintuitive) equilibrium, and players thought about which types of players were likely to make out-of-equilibrium moves, they used their previous experience to form beliefs. In this game, these empirical beliefs contradicted purely logical arguments about which players would choose the out-of-equilibrium move. The learning process supports an equilibrium which is not supportable by standard game-theoretic logic.

Although Brandts and Holt's belief-based dynamic provides the intuition for 'unrefined' play, is it the best model to characterize the relationship between the history of play and eventual convergence? Camerer and Ho (1999a,b) introduced a general model of learning in games called 'experience-weighted attraction learning' (EWA). EWA hybridizes the two most popular approaches to learning in games—reinforcement and belief formation (like Brandts and Holt)—and includes these as parametric special cases. The EWA model has been estimated on 29 experimental data sets, and outperforms the familiar special cases in 25-27 of the 29 cases, correcting for extra parameters (see Camerer, Ho, and Hsia, 2000).[1] The general conclusion is that combining features of reinforcement with belief learning in a particular way is helpful for explaining observed learning.

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. However, extensive-form games with incomplete information, like signaling games, demand special modifications which extend EWA's scope (see Vriend, 1997). The key problem is that players do not always know the foregone payoff to a signal which they did not choose (because its payoff depends on other players' reactions to the unsent signal, which is usually not known). Because foregone payoffs are used to update unchosen strategies, an extension is necessary when foregone payoffs are not known. 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.

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 estimate 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 the EWA model is described and the modifications necessary to fit it to the signaling data are detailed. Data from new experiment are then presented and we investigate how well EWA, its various extensions, and the belief and reinforcement special cases, explain the data.

2 Adaptive dynamics and equilibrium selection

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

[1] In-sample estimation is done in weak-link coordination games (Camerer and Ho, 1999a). Out-of-sample forecasting has been done for median-action coordination games and dominance solvable "p-beauty contests" (Camerer and Ho, 1999b), call markets (Hsia, 1998), "unprofitable games" (Morgan and Sefton, 1998), centipede games (Camerer, Ho and Wang, 1999), and bilateral call markets (Camerer, Ho, and Hsia, 2000). In some constant-sum games, EWA predicts slightly worse than belief learning (Camerer and Ho, 1999b).

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.

Table 1. Game 3

	Type I			Type II		
	a_1	a_2	a_3	a_1	a_2	a_3
m_1	45,30	15,0	30,15	30,30	0,45	30,15
m_2	30,90	0,15	45,15	45,0	15,30	30,15

Table 2. Game 5

	Type I			Type II		
	a_1	a_2	a_3	a_1	a_2	a_3
m_1	45,30	0,0	0,15	30,30	30,45	30,0
m_2	30,90	30,15	60,60	45,0	0,30	0,15

In both games, in the unintuitive sequential equilibrium both types of senders choose m_2. Receivers respond with $a_2|m_1$ and $a_1|m_2$.[2] In the sequential equilibrium which satisfies the Cho-Kreps (1987) intuitive criterion both types of senders choose m_1 and receivers respond with $a_1|m_1$ and $a_2|m_2$.[3]

2.1 Experimental data and the Brandts-Holt adjustment dynamic

In their Game 3 experiments, Brandts and Holt observe significant initial type separation—in the early periods, m_1 is three times more likely to be chosen by a t_1 sender than a t_2 sender. BH conjecture 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_1s tend to choose message m_1 more. The expected payoffs for t_2s are 20 and 30, so t_2s choose m_2 more often.

[2] Since both types choose m_2, receivers come to realize this, Bayesian-update and form posteriors $P(t_1|m_2) = P(t_1) = .5$ and $P(t_2|m_2) = P(t_2) = .5$. Their best response is then to choose a_1, which gives expected payoff 45 ($.5 \cdot 0 + .5 \cdot 90$). The sequential equilibrium coheres only if receivers choose a_2 in response to message m_1, which only happens when receivers believe that m_1 choices were more likely to be made by t_2s (more specifically, $P(t_2|m_1) > 2/3$ to justify a choice of a_2 by receivers). This belief does *not* satisfy the intuitive criterion because t_2s earn 45 in equilibrium (from choosing m_2 and getting response a_1) and could not possibly benefit from switching to m_1, whereas t_1s earn 30 in equilibrium and could conceivably benefit if they choose m_1.

[3] In the intuitive equilibrium both types choose m_1 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_2 instead, the equilibrium only sticks if defections to m_2 are met with responses of a_2. The a_2 response to m_2 can only be justified by the belief that m_2 defections are more likely to have come from t_2s (i.e., a_2 is optimal for receivers if $P(t_2|m_2) > 5/7$). This inference *does* satisfy the intuitive criterion because, indeed, t_2 types might benefit by defecting from m_1 to m_2 whereas t_1s would never benefit. Hence, the equilibrium in which both types choose m_1 satisfies the intuitive criterion.

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. Empirically, 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 periods 9–12 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). The receiver chooses a_2 in response, yielding a lower payoff for the type t_2 than she receives from pooling on m_1, which keeps her from defecting. The intuitive criterion requires that this be deduced from the payoff table. However, in this game, this regularity is also revealed to players through the path of play: 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_2 and most t_2s choose m_1. 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 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 dynamic (1993), players start with beliefs about what others will do (i.e., play each strategy with equal probability) and revise their beliefs in the light of what they observe. Cooper, Kagel and Garvin (1997a,b) give a similar explanation for results in limit pricing experiments.[4] This belief dynamic does explain the major features of the data. However, these earlier studies typically hypothesize particular parameter values, simulate paths for those values, and show that the simulated paths resemble the data. We improve on this procedure by estimating best-fitting parameters using maximum-likelihood. This technique yields standard errors for parameter estimates, and permits a formal

[4] Their analysis has an important twist: players assume others do not violate dominance.

hypothesis test of whether adding EWA features to belief learning improves accuracy significantly (as it has in 27 of 29 other data sets).

3 EWA learning

Experience-weighted attraction learning is a generalized reinforcement model which hybridizes elements of reinforcement and belief-based theories. Since it includes familiar models as special cases, it can be used as a statistical tool to compare theories, and ask whether adding more features improves their predictive accuracy. This section will highlight the important features of the model. See Camerer and Ho (1999b) or Camerer, Ho and Chong (2000) for more details.

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.

3.1 EWA in the normal form

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 actual payoff in period t is given by $\pi_i(s_i(t), s_{-i}(t))$.

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. The second variable is $A_i^j(t)$, the i's attraction to strategy j *after* period t has taken place. $N(t)$ and $A_i^j(t)$ begin with initial values $N(0)$ and $A_i^j(0)$, which are driven by pregame thinking due to introspection or learning transferred from similar games (e.g., Samuelson, 2000).[5]

[5] In a full model, the initial conditions would be determined by some theory of which decision rules players use (e.g., Costa-Gomes, Crawford and Broseta, 2000) or a disequilibrium theory like asymmetric response equilibrium (Weiszacker, 2000).

After a period of play, experience weights are updated according to

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

Attractions are reinforced by a weighted payoff for i's j^{th} strategy, $[\delta + (1-\delta) \cdot I(s_i^j, s_i(t))] \cdot \pi_i(s_i^j, s_{-i}(t))$ (where $I(x, y)$ is the indicator function which equals one if $x = y$ and zero otherwise). The model weights hypothetical payoffs that unchosen strategies would have earned by a parameter δ, and weights the payoff actually received, from chosen strategy $s_i(t)$, by an additional $1-\delta$ (so it receives a total weight of 1). Updated attractions $A_i^j(t)$ are a depreciated, experience-weighted lagged attraction, plus an increment for the received or foregone payoff, normalized by the new experience weight. In formal terms,

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

The factor ϕ is a discount factor that depreciates previous attractions. Along with ρ, ϕ determines the limiting values of the attractions: if $\phi > \rho$, then attractions are not bounded by payoffs, and can grow arbitrarily far apart.

Finally, attractions determine choice probabilities using the logit form[6]

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

The parameter λ measures sensitivity of players to differences among attractions: $\lambda = 0$ is equal likelihood and as λ increases, it converges to a best-response function.

3.1.1 Special cases of EWA

Simple algebra shows that parameter restrictions on $N(0), \delta$, and ρ reduce the general model to special cases of historical interest. For example, when $N(0) = 1, \rho = \delta = 0$ the model reduces to a simplified form of cumulative reinforcement in which only chosen strategies are reinforced (see Harley, 1981; Roth and Erev, 1995; Erev and Roth, 1998; Roth et al., 2000). When $\rho = \phi$ and $\delta = 1$, the model reduces to belief-based learning in which players form beliefs according to weighted fictitious play and choose strategies with high expected payoffs given those beliefs (e.g., Fudenberg and Levine, 1998, and many others).[7] When $\phi = 0$ the model corresponds to Cournot best-response dynamics, in which players

[6] Camerer and Ho (1998) show that the logit form fits slightly better than a power (or exponentiated ratio) form, and has the advantage of being usable even when attractions are negative.

[7] The key insight is that the belief updating equation can be written as a function of lagged beliefs and plugged the computation of expected payoffs. When the expected payoff computation is written as a function of its lag, the result is $E_i^j(t) = \frac{\rho \cdot N(t-1) \cdot E_i^j(t-1) + \pi(s_i^j, s_{-i}(t))}{\rho \cdot N(t-1)+1}$. The belief term magically disappears: since the only function of beliefs is to anticipate possible payoffs, and the belief updating is based on past observations, the expected payoff impact of those past observations can be mimicked by keeping track of a foregone payoff history directly. Beliefs are an unnecessary ‘middleman’.

simply choose a best response to what happened in the previous period. When $\phi = 1$ the model corresponds to fictitious play, in which beliefs about what an opponent will do are an arithmetic average of what she has done in the past. This kinship between reinforcement and belief learning is surprising because most previous researchers had thought the two were fundamentally different. Instead, the EWA framework shows that belief learning is simply a kind of generalized reinforcement learning in which unchosen strategies are reinforced as strongly as chosen ones and reinforcements are weighted averages of lagged reinforcements and payoffs.

Many researchers who have studied these models have suggested that, while simple, they may be useful approximations. This begs the question of how to judge usefulness. We adopt conventional statistical criteria which enable us to judge precisely when adding parameters helps, or when omitting parameters hurts.

3.2 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 her observed type. In games in which complete strategies are elicited this modeling approach seems reasonable. However, in our experiments 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 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.[8] In the example above, we simply update the attraction to $m_2|t_1$, and the attraction for the same message chosen by the ‘unrealized type’, $m_2|t_2$. In extensions of the model, we also allow updating of $m_1|t_1$, which the player could have chosen but did not, and $m_1|t_2$. Similarly, we assume receivers have strategies which are conditional on the message they observed the sender choosing, but not on the

[8] A referee wondered whether there is evidence that players act as if they use agent-normal-form reasoning. Note that our model allows the amount of cross-node dependence to vary parametrically, because δ, μ_1, and μ_2 express the degree of dependence of the choice at one node on future behavior at other nodes.

sender's type. A receiver's strategy to choose action k in response to message j will be denoted $a_k|m_j$.

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 $N_R^{m_2}(0)$.

3.3 The baseline model

This section discusses how the EWA model presented in Eqs. 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 their foregone payoffs 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:

$$N_R^{m_1}(t+1) = \rho \cdot N_R^{m_1} + 1 \tag{4}$$

$$A^{a_1m_1}(t+1) = \frac{\phi \cdot A^{a_1m_1}(t) \cdot N_R^{m_1}(t) + \pi_R(t_2, m_1, a_1)}{\rho \cdot N_R^{m_1}(t) + 1} \tag{5}$$

$$A^{a_2m_1}(t+1) = \frac{\phi \cdot A^{a_2m_1}(t) \cdot N_R^{m_1}(t) + \delta \cdot \pi_R(t_2, m_1, a_2)}{\rho \cdot N_R^{m_1}(t) + 1} \tag{6}$$

$$A^{a_3m_1}(t+1) = \frac{\phi \cdot A^{a_3m_1}(t) \cdot N_R^{m_1}(t) + \delta \cdot \pi_R(t_2, m_1, a_3)}{\rho \cdot N_R^{m_1}(t) + 1}. \tag{7}$$

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 sender's 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 chosen message by the *unrealized type* also be updated by that type's foregone payoff. Remember that the sender knows the receiver's strategies may be message-dependent but cannot be type-dependent. Therefore, 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. 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

$$N_S^{m_1}(t+1) = \rho \cdot N_S^{m_1} + 1 \tag{8}$$

$$A^{m_1t_1}(t+1) = \frac{\phi \cdot A^{m_1t_1}(t) \cdot N_S^{m_1}(t) + \pi_S(t_1, m_1, a_1)}{\rho \cdot N_S^{m_1}(t) + 1} \tag{9}$$

$$A^{m_1t_2}(t+1) = \frac{\phi \cdot A^{m_1t_2}(t) \cdot N_S^{m_1}(t) + \delta \cdot \pi_S(t_2, m_1, a_1)}{\rho \cdot N_S^{m_1}(t) + 1} \tag{10}$$

and $N_S^{m_2}(t+1) = N_S^{m_2}(t)$ and $A^{m_2t_k}(t+1) = A^{m_2t_k}(t)$ for $k = 1, 2$.

In the belief learning restriction of EWA, cross-type updating occurs with $\delta = 1$. Just as EWA showed that belief learning is generalized reinforcement with 'full' reinforcement of unchosen strategies, belief learning in signaling games requires full reinforcement of unrealized types. If this seems behaviorally implausible, that implausibility should count as a strike against belief learning (as a predictive model).

The updating rules for the receivers are relatively straightforward. However, the notion of updating a foregone type, learning about a situation which did not occur (but could have), can be confusing. The models we propose in Section 3.4 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:

Tabular representation of sender's baseline update rules

	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	$A_S^{m2t1}(t)$			$A_S^{m2t2}(t)$		

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, Eq. 9 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 (Eq. 10). 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 Eq. 8, and the experience count of the unchosen message is simply copied into the next period.

3.3.1 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 signaling 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 $N_R^{m_2}(0)$.

3.4 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]

[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

3.4.1 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 a median rule.[10] 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 $\Pi(m_2, t_1) = \alpha MIN(\pi_S | m_2, t_1) + (1 - \alpha) MAX(\pi_S | m_2, t_1)$, where α is the parameter of the convex combination. Then the convex combination model can be written

Tabular representation of sender's convex combination model update rules

	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	$\frac{\phi^\tau A_S^{m2t1}(t) N_S^{m2}(t) + \mu_1 \tau \Pi(m_2, t_1)}{\rho^\tau N_S^{m2}(t) + \tau}$			$\frac{\phi^\tau A_S^{m2t2}(t) N_S^{m2}(t) + \mu_2 \tau \Pi(m_2, t_2)}{\rho^\tau N_S^{m2}(t) + \tau}$		

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 are 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, suppose that $\tau = 1$. The unchosen message rules then 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

receiver's action choices could be message-dependent. It seems unlikely to fit the data better so we have not investigated it empirically.

[10] We also estimated a model where unchosen messages were updated with the median of the set of foregone payoffs. Results were virtually identical to the convex combination model presented here, and are available in Anderson and Camerer (1999).

message attractions are not discounted and no payoff is added to them, so they are unchanged.

3.4.2 Mirror sophistication: internal models of other players

A second 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'.[11] 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 in 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.[12]

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 $\Pi(m_2, t_1) = \sum_{j=1}^{3} p_S(a_j|m_2) \times \pi_S(m_2, t_1, a_j)$. Then the simple sophistication model can be expressed as in the table below.

Tabular representation of sender's mirror sophistication model update rules

	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	$\frac{\phi^\tau A_S^{m2t1}(t)N_S^{m2}(t)+\mu_1\tau\Pi(m_2,t_1)}{\rho^\tau N_S^{m2}(t)+\tau}$	$\frac{\phi^\tau A_S^{m2t2}(t)N_S^{m2}(t)+\mu_2\tau\Pi(m_2,t_2)}{\rho^\tau N_S^{m2}(t)+\tau}$

The parameters are interpreted exactly as in the previous model. 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

[11] Weber (2000) found this phenomenon in a repeated dominance-solvable beauty-contest game, when subjects made strategy choices repeatedly with *no feedback* after each choice. He found that players 'learned' even without feedback. Weber estimated a model in which one's own choice is taken as a proxy for what others might have done, and found a significant influence like mirror sophistication.

[12] Obviously, this rule will be sensitive to the experimental protocol and does not apply if the players do not switch roles. 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. (For example, Binmore, Shaked and Sutton, 1985, found that second-moving bargainers learned to make subgame perfect offers in one trial after they reversed roles and became first-movers.) This is easily testable, by comparing experiments with different degrees of role-switching, but we know of no such experiments.

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, and has no further provision for thinking about 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.[13] 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 choose 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 1 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 expected additional periods to cause t_2s to choose m_1 less and less frequently, cementing convergence to the unintuitive equilibrium at m_2. Our hunch was wrong: additional periods do not eliminate the separation between messages. The senders' strategies during the 13^{th} through 32^{nd} periods look much like the 9^{th} through 12^{th} periods of the Brandts and Holt data, suggesting the convergence to the sequential equilibrium

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

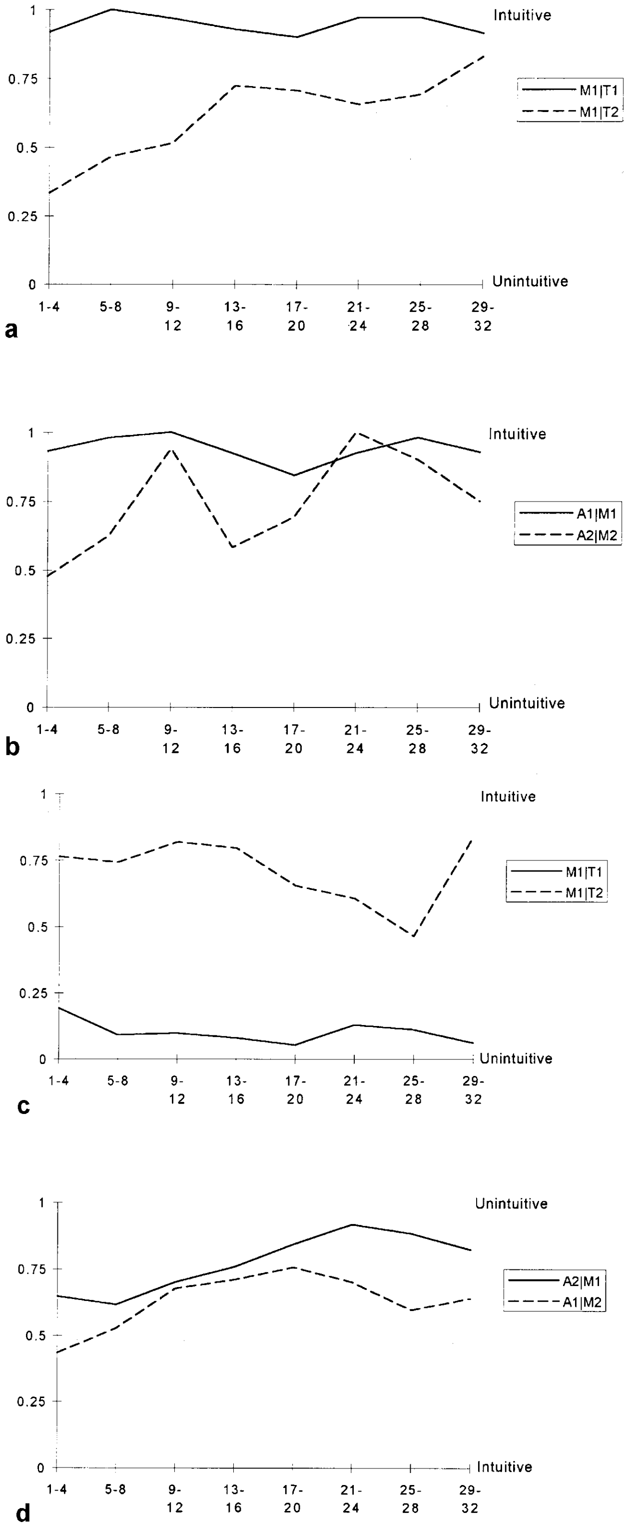

Figure 1. a Game 3 Sender. **b** Game 3 Receiver. **c** Game 5 Sender. **d** Game 5 Receiver

is not complete, even after many more periods of learning.[14] However, there is also no evidence of movement back toward the intuitive equilibrium at m_1.[15]

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).[16] 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).[17]

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.

[14] 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.

[15] To test that 'zero-aversion' was preventing some subjects from switching to the unintuitive 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, because (due to software constraint prohibiting three-digit numbers) it was indicated by a post-it note pasted on subjects' computer screens. Therefore, we ran three more sessions using a payoff table which multiplied payoffs in Game 5' by 4/5 (call this game 5*). Behavior in game 5* was indistinguishable from that in Game 5. We brought back experienced subjects from game 5* and ran them for 64 more periods, for a total of 128 periods, but did not observe further convergence.

[16] 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.

[17] 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.

Our final restriction is that each of the $N(0)$s must 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 play of the game. Since $\frac{1}{1-\rho}$ is the asymptotic bound of $N(t)$ as t gets large, 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)).

Imposing these restrictions compromises the asymptotic normality of the maximum likelihood parameter estimates, so we construct bootstrapped confidence intervals using the percentile method.[18]

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 of this study is the baseline EWA model described in Section 3.3. 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 and belief-based (BB) model for Game 3.[19] The predictions they generate are shown in Figure 2 (along with choice reinforcement). The estimates are performed 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 several 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

[18] 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 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).

[19] Reinforcement model estimates are reported in Anderson and Camerer (1999). The only interesting parameter is $\hat{\phi}$ = .70 in Game 3 and $\hat{\phi}$ = .68 in Game 5.

Table 3. Parameter estimates for Game 3 (underlined values are fixed for identification or for model restrictions and bootstrapped 95% confidence intervals are in parentheses)

	EWA	BB		
		Value	Param	Count
δ	0.69 (0.47,1.00)	<u>1.00</u>		
ϕ	1.02 (0.99,1.04)	0.97 (0.97,0.99)		
ρ	1.00 (0.98,1.00)	0.97 (0.97,0.99)		
λ	0.41 (0.34,0.54)	0.32 (0.27,0.38)		
$A_S^{m_1 t_1}(0)$	<u>15.00</u>	25.50	$N_S^{a_1 m_1}(0)$	0.00 (0.00,0.00)
$A_S^{m_2 t_1}(0)$	9.90 (9.04,10.72)	17.56	$N_S^{a_2 m_1}(0)$	10.40 (10.39,10.41)
$N_S^{m_1}(0)$	50.00 (49.91,50.00)	34.64	$N_S^{a_3 m_1}(0)$	24.24 (24.23,24.24)
$A_S^{m_1 t_2}(0)$	14.72 (13.94,15.24)	20.99	$N_S^{a_1 m_2}(0)$	0.00 (0.00,0.00)
$A_S^{m_2 t_2}(0)$	<u>15.00</u>	20.85	$N_S^{a_2 m_2}(0)$	15.90 (15.89,15.90)
$N_S^{m_2}(0)$	32.91 (32.81,32.94)	26.07	$N_S^{a_3 m_2}(0)$	10.17 (10.17,10.18)
Receiver				
$A_R^{a_1 m_1}(0)$	<u>30.00</u>	30.00	$N_R^{m_1 t_1}(0)$	17.39 (17.39,17.39)
$A_R^{a_2 m_1}(0)$	25.12 (24.61,25.74)	22.40	$N_R^{m_2 t_1}(0)$	6.88 (6.86,6.88)
$A_R^{a_3 m_1}(0)$	15.00 (15.00,15.00)	15.00	$N_R^{m_1 t_2}(0)$	17.25 (17.24,17.25)
$N_R^{m_1}(0)$	50.00 (49.91,50.00)	34.64	$N_R^{m_2 t_2}(0)$	19.19 (19.19,19.20)
$A_R^{a_1 m_2}(0)$	20.08 (19.37,20.32)	23.76		
$A_R^{a_2 m_2}(0)$	21.78 (21.16,21.87)	26.04		
$A_R^{a_3 m_2}(0)$	<u>15.00</u>	15.00		
$N_R^{m_2}(0)$	32.91 (32.81,32.94)	31.57		

Table 4. Goodness of fit statistics for Game 3 (parameters estimated to minimize In Sample LL)

	EWA	BB	CR	Convex	Soph	Freq
Calibration						
In Sample LL	-12.06	-12.55	-14.15	-11.16	-11.17	-12.74
Out Sample LL	-9.02	-10.28	-9.83	-7.56	-7.55	-10.22
AIC	-12.56	-13.05	-14.48	-11.83	-11.80	-12.99
BIC	-13.71	-14.21	-15.26	-13.38	-13.25	-13.57
Fit						
In Sample Miss	0.147	0.184	0.177	0.131	0.134	0.178
Out Sample Miss	0.090	0.090	0.125	0.062	0.062	0.102
In Sample MSD	0.095	0.101	0.108	0.087	0.087	0.106
Out Sample MSD	0.063	0.076	0.082	0.051	0.052	0.083
DOF	12	12	8	16	15	6

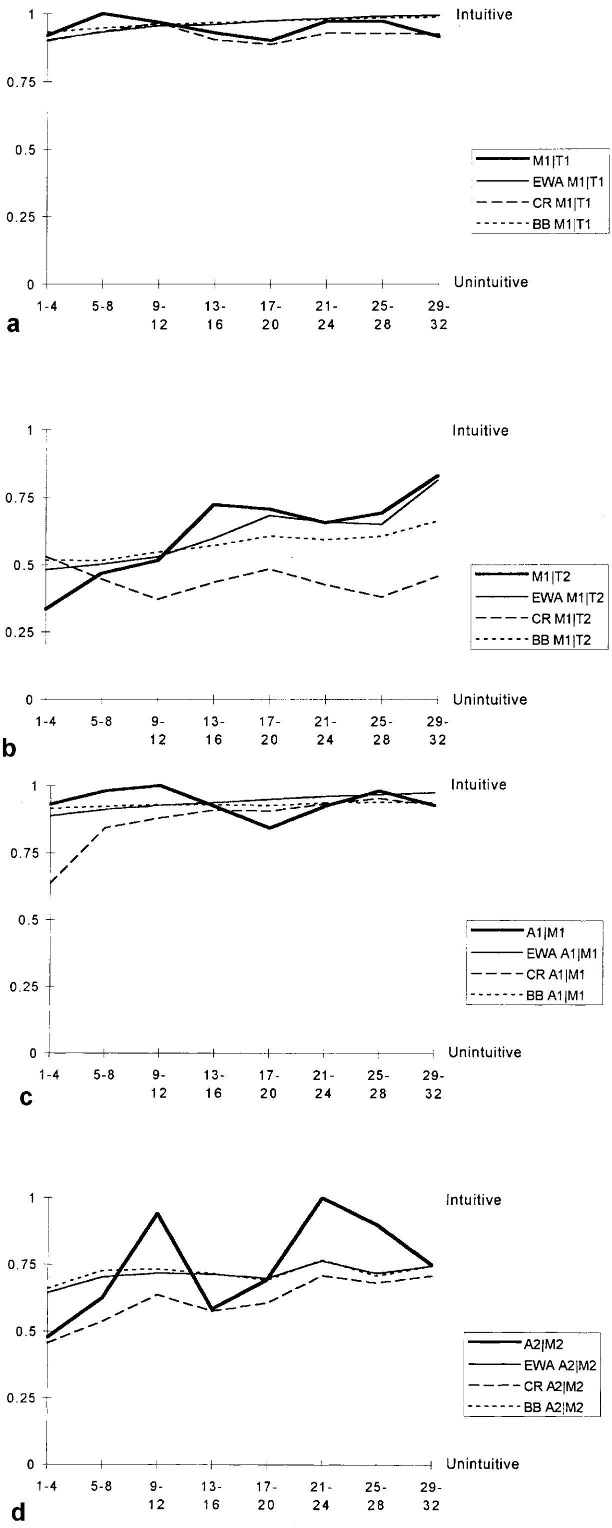

Figure 2. **a** Game 3 $m_1|t_1$. **b** Game 3 $m_1|t_2$. **c** Game 3 $a_1|m_1$. **d** Game 3 $a_2|m_2$

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 (because a more general model will not automatically fit better out-of-sample). The third row presents the Akaike information criterion (AIC) and the fourth row presents the Bayesian information criterion (BIC).[20] 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.[21]

The second section of Table 4 presents more 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 subject *does not* pick the strategy that is predicted most likely to be chosen by the model. (It is one minus the hit rate.) The second two rows give the average per-period mean squared deviation.[22]

The columns of Table 4 all represent models discussed in this paper, except for the last one. The last column (Freq) 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.

Choice reinforcement's out-of-sample miss rate and MSD are much worse than EWA and barely better than the frequency model. It does not fit the data well because it does not use foregone payoff information. Figure 2 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 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

[20] 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 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).

[21] Note that the conventional way to make nested comparisons 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.

[22] 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).

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 frequency model.

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 adequately track the increasing frequency of play $m_1|t_2$. The problem is that the large value 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.[23] 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 belief model 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 converges too slowly. Adding EWA parameters improves the fit considerably, producing a plot which hugs the data except at the start and in periods 13-16.

Table 5 presents the parameter estimates for Game 5. Figure 3 shows the predictions they generate. Table 6 presents goodness of fit statistics. 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 $\hat{\delta}$ = 0.54, 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 $\hat{\phi} = 0.68 > \hat{\rho} = 0.46$, which means that attractions are growing over time, and are not bounded by payoffs.

The initial experience counts are only 0.62 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 (which represent most of the non-a_1 responses to m_2). 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.

Reinforcement does not fit well because it misses 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

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

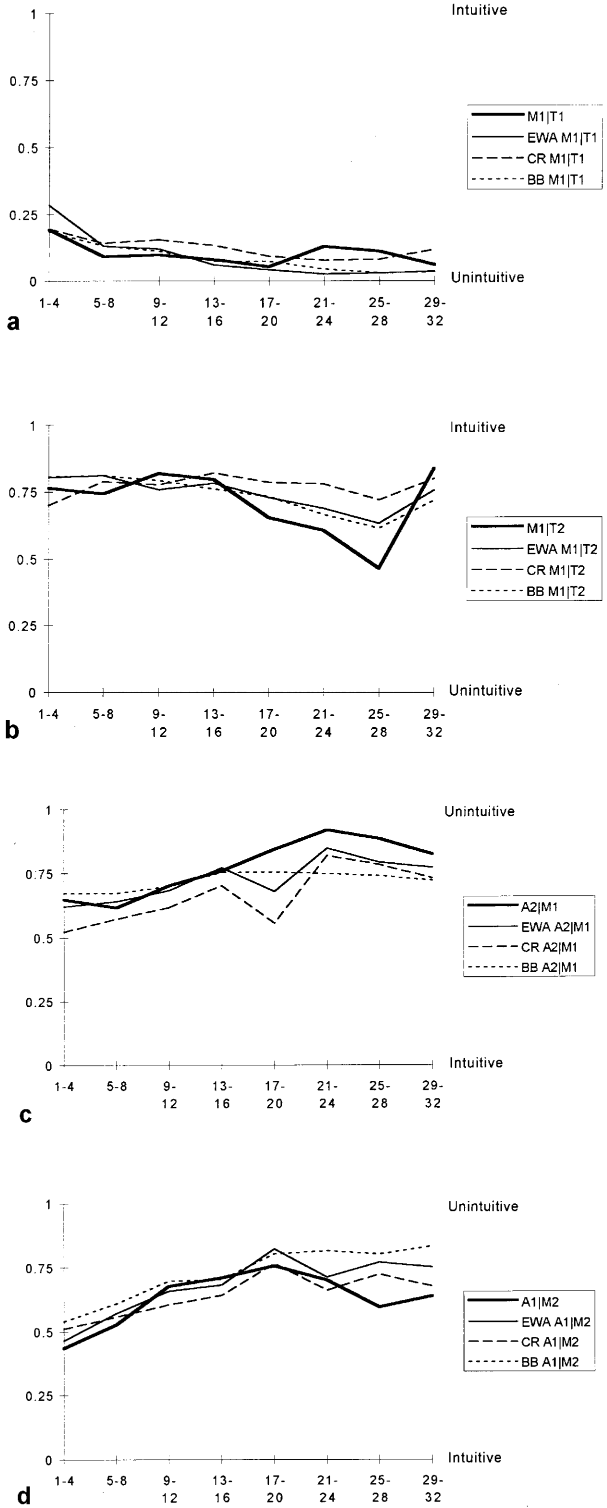

Figure 3. **a** Game 5 $m_1|t_1$. **b** Game 5 $m_1|t_2$. **c** Game 3 $a_2|m_1$. **d** Game 5 $a_1|m_2$

Table 5. Parameter estimates for Game 5 (underlined values are fixed for identification or for model restrictions and botstrapped 95% confidence intervals are in parentheses)

	EWA	BB		
		Value	Param	Count
δ	0.54 (0.45,0.63)	1.00		
ϕ	0.65 (0.59,0.71)	0.88 (0.88,0.94)		
ρ	0.46 (0.39,0.54)	0.88 (0.88,0.94)		
λ	0.09 (0.07,0.11)	0.15 (0.12,0.18)		
Sender				
$A_S^{m_1t_1}(0)$	18.25 (18.25,18.26)	34.70	$N_S^{a_1m_1}(0)$	2.18 (2.17,2.20)
$A_S^{m_2t_1}(0)$	30.00	44.62	$N_S^{a_2m_1}(0)$	0.00 (0.00,0.01)
$N_S^{m_1}(0)$	0.62 (0.59,0.66)	2.84	$N_S^{a_3m_1}(0)$	0.65 (0.62,0.67)
$A_S^{m_1t_2}(0)$	30.00	30.00	$N_S^{a_1m_2}(0)$	3.60 (3.58,3.61)
$A_S^{m_2t_2}(0)$	11.34 (11.34,11.34)	19.95	$N_S^{a_2m_2}(0)$	0.56 (0.54,0.59)
$N_S^{m_2}(0)$	3.37 (3.37,3.38)	8.12	$N_S^{a_3m_2}(0)$	3.96 (3.95,3.97)
Receiver				
$A_R^{a_1m_1}(0)$	30.00	30.00	$N_R^{m_1t_1}(0)$	0.63 (0.58,0.65)
$A_R^{a_2m_1}(0)$	37.26 (37.26,37.26)	35.04	$N_R^{m_2t_1}(0)$	3.06 (3.04,3.07)
$A_R^{a_3m_1}(0)$	0.00 (0.00,0.01)	3.32	$N_R^{m_1t_2}(0)$	2.21 (2.20,2.23)
$N_R^{m_1}(0)$	0.62 (0.59,0.66)	2.84	$N_R^{m_2t_2}(0)$	5.06 (5.05,5.08)
$A_R^{a_1m_2}(0)$	41.88 (41.88,41.88)	33.93		
$A_R^{a_2m_2}(0)$	15.00	24.35		
$A_R^{a_3m_2}(0)$	43.26 (43.26,43.26)	31.96		
$N_R^{m_2}(0)$	3.37 (3.37,3.38)	8.12		

they switch. By leaving out unrealized type reinforcement, 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. Table 6 shows that adding EWA parameters to reinforcement improve in- and out-of-sample fit by every statistic.

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 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

Table 6. Goodness of fit statistics for Game 5 (parameters estimated to minimize In Sample LL)

	EWA	BB	CR	Convex	Soph	Freq
Calibration						
In Sample LL	-13.49	-16.41	-15.26	-13.13	-13.15	-18.00
Out Sample LL	-17.44	-20.40	-19.20	-17.69	-17.29	-18.60
AIC	-13.99	-16.91	-15.60	-13.80	-13.78	-18.25
BIC	-15.15	-18.07	-16.37	-15.35	-15.23	-18.83
Fit						
In Sample Miss	0.181	0.221	0.190	0.171	0.172	0.253
Out Sample Miss	0.184	0.227	0.199	0.191	0.191	0.254
In Sample MSD	0.105	0.127	0.116	0.101	0.101	0.146
Out Sample MSD	0.113	0.138	0.127	0.116	0.113	0.145
DOF	12	12	8	16	15	6

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.

Table 7 presents the parameter estimates for Game 3 for the two alternative models.[24] While they both offer significant improvement over EWA, they have similar parameter estimates and offer similar fits. The fit statistics in Table 4 show that there is a significant feature of senders’ behavior EWA is not capturing, and both unchosen message models capture it equally well.

The unchosen message models make slight improvements over EWA in a number of places. We highlight one such improvement to convey the subtle nature of the dynamic the unchosen message models predict.

The $m_1|t_1$ time series (Figure 2a) shows that t_1s almost always choose m_1, but occasionally choose m_2 (0-10% of the time, depending on the block). Starting with period 13, the m_2 rate fluctuates from about 3-10% and increases over time. The EWA model underpredicts this rate, starting about 4% and decreasing over time, to 1% in the last few periods. The unchosen message models do much better because they predict that the $m_2|t_2$ rate is around 4%, and slightly increasing over time.

How do the unchosen message models keep $m_2|t_2$ around while EWA all but extinguishes it? Consider how EWA updates when t_1s choose m_1, which they

[24] We do not include figures for the unchosen message models because their predictions are very similar to EWA and to each other in both games. Figures are available in Anderson and Camerer (1999).

Table 7. Parameter estimates for Game 3 (underlined values are fixed for identification or for model restrictions and bootstrapped 95% confidence intervals are in parentheses)

	EWA	Convex	Soph
δ	0.69 (0.47,1.00)	0.59 (0.56,0.64)	0.48 (0.33,0.66)
ϕ	1.02 (0.99,1.04)	0.77 (0.76,0.81)	0.76 (0.71,0.81)
ρ	1.00 (0.98,1.00)	0.98 (0.98,1.00)	0.98 (0.98,1.00)
λ	0.41 (0.34,0.54)	1.11 (1.10,1.15)	1.20 (1.13,1.36)
τ		1.07 (1.04,1.11)	1.05 (0.95,1.15)
μ_1		0.00 (0.00,0.00)	0.00 (0.00,0.01)
μ_2		0.46 (0.41,0.50)	0.57 (0.48,0.72)
α		0.00 (0.00,0.00)	
Sender			
$A_S^{m_1 t_1}(0)$	$\underline{15.00}$	$\underline{15.00}$	$\underline{15.00}$
$A_S^{m_2 t_1}(0)$	9.90 (9.04,10.72)	12.52 (12.51,12.52)	12.79 (12.77,12.81)
$N_S^{m_1}(0)$	50.00 (49.91,50.00)	29.32 (29.32,29.32)	32.35 (32.35,32.35)
$A_S^{m_1 t_2}(0)$	14.72 (13.94,15.24)	14.59 (14.58,14.59)	14.57 (14.48,14.57)
$A_S^{m_2 t_2}(0)$	$\underline{15.00}$	$\underline{15.00}$	$\underline{15.00}$
$N_S^{m_2}(0)$	32.91 (32.81,32.94)	50.00 (50.00,50.00)	50.00 (50.00,50.00)
Receiver			
$A_R^{a_1 m_1}(0)$	$\underline{30.00}$	$\underline{30.00}$	$\underline{30.00}$
$A_R^{a_2 m_1}(0)$	25.12 (24.61,25.74)	27.38 (27.38,27.38)	27.55 (27.52,27.57)
$A_R^{a_3 m_1}(0)$	15.00 (15.00,15.00)	15.00 (15.00,15.00)	15.00 (15.00,15.00)
$N_R^{m_1}(0)$	50.00 (49.91,50.00)	29.32 (29.32,29.32)	32.35 (32.35,32.35)
$A_R^{a_1 m_2}(0)$	20.08 (19.38,20.32)	17.20 (17.19,17.21)	16.97 (16.93,17.03)
$A_R^{a_2 m_2}(0)$	21.78 (21.16,21.87)	17.81 (17.80,17.82)	17.52 (17.46,17.57)
$A_R^{a_3 m_2}(0)$	$\underline{15.00}$	$\underline{15.00}$	$\underline{15.00}$
$N_R^{m_2}(0)$	32.91 (32.81,32.94)	50.00 (50.00,50.00)	50.00 (50.00,50.00)

almost always do. Because of the receivers' actions, $m_1|t_1$ is almost always updated with its highest payoff, 45. What happens to the attraction of $m_2|t_1$? Since m_2 is rarely chosen by t_1s, most of the updating of that attraction comes from unrealized type updating when t_2s choose m_2. Because the unrealized type reinforcement is estimated to be strong ($\hat{\delta} = 0.69$), and given how receivers respond to m_2, the attraction for $m_2|t_1$ usually gets reinforcement of 0, and sometimes $\delta \cdot 30$ (about 21). Because $m_1|t_1$ is typically getting reinforced by 45, and $m_2|t_1$ is typically getting reinforced by 0 or 21, EWA predicts $m_2|t_1$ gets more and more infrequently chosen over time.

The unchosen message models correct this subtle ‘overlearning’ predicted by EWA by using choices of a different message by a different type, $m_1|t_2$, as a cognitive opportunity to think again about the possible payoffs from m_2 for t_1s. In the convex model, for example, μ_2 is estimated to be 0.57 (and all the weight is on an unchosen message’s highest payoff), so that when $m_1|t_2$ is chosen, $m_2|t_1$ is reinforced by 0.57 (45), around 26. Updating of ‘distant’ choices is like a reminder that a message which is rarely chosen by a particular type may yield a good payoff after all. Since that message-type combination is not directly reinforced very often, this indirect reinforcement is necessary to maintain a substantial probability that it may occur in the future. The baseline model does not allow this kind of reinforcement, and hence, overpredicts how quickly $m_1|t_1$ is distinguished by direct experience. The difference is small in percentage terms, but is important in statistical estimation and gives a substantial predictive advantage (especially out-of-sample) to the unchosen message models.[25]

The unchosen message model estimates for Game 5 are presented in Table 8. As with Game 3, the behavior of both of Game 5’s alternative models is similar to EWA, and very similar to each other. This again suggests that if there is some significant pattern in the data not captured by EWA, these models capture it in the same way. Table 6 shows, however, that the improvement in fit from modeling the unchosen messages is barely worth the extra degrees of freedom. AIC favors the unchosen message models, but BIC does not, and the out-of-sample statistics are similar for EWA and the unchosen message models.

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 (or 128) periods is not enough to converge to equilibrium.

Using these data, we tested our adaptation of EWA to signalling games. The baseline model updates the attractions to a sender’s unrealized *type*. This allows the sender to make all valid inferences given that receivers are 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 underestimate the speed of learning.

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

[25] While the frequencies of $m_2|t_2$ are small, they can have a large impact on estimation (particularly when log likelihood is the fit measure). Because the logarithms of small positive numbers can be hugely negative, it makes a big difference whether a model predicts that a rare event is impossible, or just very unlikely.

Table 8. Parameter estimates for Game 5 (underlined values are fixed for identification or for model restrictions and bootstrapped 95% confidence intervals are in parentheses)

	EWA	Convex	Soph
δ	0.54 (0.45,0.63)	0.51 (0.49,0.52)	0.50 (0.48,0.54)
ϕ	0.65 (0.59,0.71)	0.70 (0.69,0.72)	0.70 (0.68,0.71)
ρ	0.46 (0.39,0.54)	0.86 (0.86,0.87)	0.84 (0.84,0.85)
λ	0.09 (0.07,0.11)	0.28 (0.26,0.32)	0.25 (0.23,0.30)
τ		0.57 (0.56,0.58)	0.48 (0.45,0.49)
μ_1		0.05 (0.04,0.05)	0.00 (0.00,0.01)
μ_2		0.22 (0.20,0.23)	0.14 (0.09,0.14)
α		0.00 (0.00,0.00)	
Sender			
$A_S^{m_1t_1}(0)$	18.25 (18.25,18.26)	23.83 (23.83,23.83)	23.30 (23.30,23.30)
$A_S^{m_2t_1}(0)$	$\underline{30.00}$	$\underline{30.00}$	$\underline{30.00}$
$N_S^{m_1}(0)$	0.62 (0.59,0.66)	5.01 (5.01,5.01)	4.36 (4.36,4.36)
$A_S^{m_1t_2}(0)$	$\underline{30.00}$	$\underline{30.00}$	$\underline{30.00}$
$A_S^{m_2t_2}(0)$	11.34 (11.34,11.34)	24.41 (24.41,24.41)	23.79 (23.79,23.79)
$N_S^{m_2}(0)$	3.37 (3.37,3.37)	7.09 (7.09,7.09)	6.32 (6.32,6.32)
Receiver			
$A_R^{a_1m_1}(0)$	$\underline{30.00}$	$\underline{30.00}$	$\underline{30.00}$
$A_R^{a_2m_1}(0)$	37.26 (37.26,37.26)	31.87 (31.87,31.87)	32.09 (32.09,32.09)
$A_R^{a_3m_1}(0)$	0.00 (0.00,0.01)	0.00 (0.00,0.00)	0.00 (0.00,0.01)
$N_R^{m_1}(0)$	0.62 (0.59,0.66)	5.01 (5.01,5.01)	4.36 (4.36,4.36)
$A_R^{a_1m_2}(0)$	41.88 (41.88,41.88)	23.77 (23.77,23.77)	24.88 (24.88,24.88)
$A_R^{a_2m_2}(0)$	$\underline{15.00}$	$\underline{15.00}$	$\underline{15.00}$
$A_R^{a_3m_2}(0)$	43.26 (43.26,43.26)	24.06 (24.06,24.06)	25.17 (25.16,25.17)
$N_R^{m_2}(0)$	3.37 (3.37,3.37)	7.09 (7.09,7.09)	6.32 (6.32,6.32)

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 about one in Game 3 and about 0.5 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 may also suggest that updating same-type, different-message attractions does not allow quick-enough convergence to type-conditional messages.

Finally, we hoped to gain some insight into how subjects reinforce unchosen messages. The two 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 sophisticated payoff model (indeed its AIC and BIC are higher for both games). However, the models are so similar that it is difficult to conclude the intuition behind the mirror sophistication model is more compelling than that of the convex combination model. Thus, while we have been able to determine that senders do update the unchosen message attractions, 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 either of the unchosen message models and expect to do adequately.

Using formal learning models can provide insight into how unintuitive equilibria might arise in natural markets. One example of historical convergence to an unintuitive equilibrium is dividend policy of firms (see Bhattacharya, 1979). From a tax point of view, firms should not pay dividends because they are taxed as regular income of investors; if the firms' cash were instead reinvested, the result would be higher investor capital gains, which are taxed at a lower rate. So why do firms pay dividends? Suppose there are two types of firms: low-quality ones, which do not always have enough cash to meet a regular dividend payment (and cannot borrow to finance it), and high-quality ones which have plenty of cash and good business prospects. Decades ago, security analysts were less able to learn about a firm's financial health from accounting data and company sources than they are today; and struggling firms were less able to borrow. In this era, regular dividend payments signal a firm's financial health: low-quality firms cannot afford to commit to dividends (and often miss regular payments), but high-quality firms can. This part of corporate history corresponds to a temporary separating-equilibrium phase in which low- and high-quality firms are distinguished by their dividend policies.

But as firms realized how important dividends are, and credit markets developed, low-quality firms soon realized they had to pay dividends (or else reveal their type) and could borrow to do so. So a pooling equilibrium emerged in which all firms paid dividends. However, this equilibrium is unintuitive if high-quality firms have good investment opportunities and would prefer to plow dividend payments into those investments. If security analysts have the (newly-developed) capacity to guess a firm's investment prospects, these high-quality firms could conceivably benefit from cutting the dividend (if the capital markets interpreted this as a signal of having good opportunities). However, low-quality firms cannot benefit as much (if security analysts can see they have few good opportunities). Thus, firms may be 'forced' to continue to pay dividends if they think capital markets will interpret a cut as a signal of low-quality—since only low-quality firms did not pay dividends in the past. In this analytical narrative, dividend policy is an unintuitive equilibrium which emerged because the trace of the past, in which only low-quality firms did not pay dividends, inhibits high-quality firms from breaking the pooling equilibrium (even though their perception of the capital markets' likely reaction does not obey the intuitive criterion).

Although we cannot identify the exact form of unchosen message updating, we have replicated earlier results that empirical histories which conflict with logical refinements can be generated, and that these histories interfere with convergence to more refined equilibria. The Brandts-Holt dynamic provides the intuition for how this conflict might arise. Adding the flexibility of EWA improves our understanding (and predictive accuracy) considerably. Therefore, carefully specified formal learning models provide insight into how agents combine information about the history of play with the payoff table to make strategic decisions and help us to understand when play might be inconsistent with logical refinements.

References

Anderson, C.M., Camerer, C.F.: Experience-weighted attraction learning in sender-receiver signaling games. California Institute of Technology Social Science working paper number 1058 (1999)

Aptech: Constrained maximum likelihood (1995)

Banks, J., Camerer, C.F., Porter, D.: An experimental analysis of Nash refinements in signaling games. Games and Economic Behavior **6**(1), 1–31 (1994)

Bhattacharya, S.: Imperfect information, divided policy and 'the bird in the hand' fallacy. Bell Journal of Economics **10**, 259–270 (1979)

Binmore, K., Shaked, A., Sutton, J.: Testing noncooperative bargaining theory: A preliminary study. American Economic Review **75**, 1178–1180 (1985)

Brandts, J., Holt, C.: An experimental test of equilibrium dominance in signaling games. American Economic Review **82**(5), 1350–1365 (1992)

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

Brandts, J., Holt, C.: Naive bayesian learning and adjustment to equilibrium in signaling games. University of Virginia Department of Economics working paper (1994)

Brown, G.: Iterative solution of games by fictitious play. In: Activity, analysis of production and allocation. New York: Wiley 1951

Camerer, C.F., Ho, T., Chong, K.: Sophisticated learning and strategic teaching in repeated games. Caltech working paper (2000)

Camerer, C.F., Hsia, D., Ho, T.: Ewa learning in coordination games and bilateral call markets. Caltech working paper (2000)

Camerer, C.F., Ho, T.: EWA learning in games: Heterogeneity, time-variation, and probability form. Journal of Mathematical Psychology **42**, 305–326 (1998)

Camerer, C.F., Ho, T.: Experience-weighted attraction learning in games: Estimates from weak-link games. In: Budescu, D., Erev, I., Zwick, R. (eds.) Games and human behavior, pp. 31–52. Mahwah, NJ: Lawrence Erlbaum 1999a

Camerer, C.F., Ho, T.: Experience-weighted attraction learning in normal form games. Econometrica **67**(3), 827–874 (1999b)

Camerer, C.F., Ho, T., Wang, X.: Heterogeneous EWA learning with payoff learning in centipede games. California Institute of Technology Social Science working paper (1999)

Carlin, B., Louis, T.: Bayes and empirical Bayes methods for data analysis. New York: Chapman and Hall 1996

Cho, K., Kreps, D.: Signaling games and stable equilibria. Quarterly Journal of Economics **102**(2), 179–221 (1987)

Cooper, D., Garvin, S., Kagel, J.: Signalling and adaptive learning in an entry limit pricing game. RAND Journal of Economics **28**(4), 662–683 (1997a)

Cooper, D., Garvin, S., Kagel, J.: Adaptive learning vs. equilibrium refinements in an entry limit pricing game. Economic Journal **107**(442), 553–575 (1997b)

Costa-Gomes, M., Crawford, V., Broseta, B.: Cognition and behavior in normal-form games. Econometrica (forthcomming)

Cournot, A.: Researches in the mathematical principles of the theory of wealth (translated by Bacon, N.). London: Haffner 1960

Efron, B., Tibshirani, R.: An introduction to the bootstrap. New York: Chapman and Hall 1993

Erev, I., Roth, A.: Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. American Economic Review **88**, 848–881 (1998)

Fudenberg, D., Levine, D.: The theory of learning in games. Cambridge, MA: MIT Press 1998

Gibbons, R.: Applied game theory for economists. Princeton, NJ: Princeton University Press 1992

Harley, C.: Learning the evolutionarily stable strategy. Journal of Theoretical Biology **89**, 611–633 (1981)

Hsia, D.: Learning in call markets. USC Department Economics (1998)

Morgan, J., Sefton, M.: An experimental investigation of unprofitable games. Princeton University Woodrow Wilson School manuscript (1998)

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

Roth, A., Erev, I., Slonim, R., Barron, G.: Equilibrium and learning in economic environments. The predictive value of approximations. Harvard University working paper (2000)

Samuelson, L.: Analogies, adaptations, and anomalies. University of Wisconsin working paper (2000)

Tirole, J.: The theory of industrial organization. Cambridge, MA: MIT Press 1988

Vriend, N.: Will reasoning improve learning? Economic Letters **55**, 9–18 (1997)

Weber, R.: Learning without feedback in beauty contests. Carnegie Melon University working paper (2000)

Weiszacker, G.: Ignoring the rationality of others: Evidence from experimental normal-form games. Harvard Business School working paper (2000)