The Economics of Learning Models: A Self-tuning Theory of Learning in Games

Teck H. Ho, Colin F. Camerer, and Juin-Kuan Chong[1]

March 14, 2004

---

[1]Ho: Haas School of Business, University of California, Berkeley, Berkeley, CA 94720-1900, hoteck@haas.berkeley.edu. Camerer: Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125, camerer@hss.caltech.edu. Chong: National University of Singapore, 1 Business Link, Singapore 117951, bizcjk@nus.edu.sg. Direct correspondence to the first author. Thanks to participants in the 2000 Southern Economics Association meetings, the Wharton School Decision Processes Workshop, the University of Pittsburgh, the Berkeley Marketing Workshop, the Nobel Symposium on Behavioral and Experimental Economics (December 2001) and C. Mónica Capra, David Cooper, Vince Crawford, Ido Erev, and Guillaume Frechette for helpful comments.

Abstract

Self-tuning experience weighted attraction (EWA) is a one-parameter theory of learning in games. It replaces the key parameters in an earlier model (EWA) with functions of experience that "self-tune" over time. The theory was tested on seven different games, and compared to the earlier model and a one-parameter stochastic equilibrium theory. The more parsimonious self-tuning EWA does as well as EWA in predicting behavior in new games, and reliably better than an equilibrium benchmark. The economic value of a learning theory is measured by how much more subjects would have earned in an experimental session if they followed the theory's recommendations. Economic values for several learning and equilibrium theories were estimated (controlled for boomerang effects of following a model's advice in one period, on future earnings). Most models have economic value. Self-tuning EWA adds the most value.

1 Introduction

The power of equilibrium models of games comes from their ability to produce precise predictions using only the structure of a game and assumptions about players' rationality. Statistical models of learning, on the other hand, often need data to calibrate free parameters, and use a specific criterion (e.g., maximized likelihood) to judge the relative accuracy of models. The calibrated models can then be used to generate predictions about behavior in new games. This calibration-prediction process, while standard in econometrics, gives rise to two unresolved issues for learning models: Specifying an accuracy or value criterion, and explaining apparent cross-game variation in learning rules.

The first unresolved issue is how to judge the relative performance of learning theories. Various criteria, such as mean squared deviation, hit rate, log-likelihood, and the value of theories in terms of equivalent empirical observation (Alvin Roth et al., 2002) have been used to measure predictive accuracy and show when models fit well and badly. These statistics are useful. But Schelling (1960, p. 98) noted that, "a normative theory must produce strategies that are at least as good as what people can do without them." Schelling's definition suggests a new measure of the value of theories: The *economic value* of a theory is defined as how much more (or less) players would earn if they use the theory to forecast what others will do, and best-respond given that forecast, compared to how much they actually earn.

At first blush, economic value is simply a business person's measure: It answers the question, How much money is your theory worth? But economic value is also a measure of the degree of disequilibrium, in economic terms. If players' beliefs are not in equilibrium, then (by definition) those players are misforecasting what others will do. A theory with more accurate beliefs will have positive economic value (and an equilibrium theory can have *negative* economic value if it misleads players). On the other hand, if players are in equilibrium, then an equilibrium theory will advise them to make the same choices they would make anyway, and hence will have zero economic value. A non-equilibrium theory which advises them to choose differently will have *negative* economic value. So computing economic value can be seen as a novel way to evaluate how close to equilibrium players are– in dollar terms. Economic value can also be interpreted as the "fitness advantage" of theories, which is a useful input to evolutionary game theory.[2]

[2]In replicator dynamics models, the growth rate of organisms which use a strategy is proportional to the difference between that strategy's fitness (against the population average) and the population-weighted strategy fitness (e.g., Weibull, 1995). If fitness is equated to experiment payoff, this fitness advantage is exactly the same

The distinction between statistical accuracy and economic value is important because theories which are statistically much more accurate than the implicit theories players use may not add much economic value. And theories which are only small statistical improvements may be very valuable (e.g., in high-stakes decision making or forecasting in large markets). This paper computes both statistical accuracy and economic value and compares them.

The second unresolved issue is how to respond to cross-game variation in parameters of learning theories. Some studies have shown that the best-fitting parameter values of learning rules vary significantly across games and across subjects (see for example Yin-Wong Cheung and Daniel Friedman, 1997 and Colin Camerer and Teck Ho, 1999; for a comprehensive review, see Camerer, 2003). Some learning theorists regard parameter variation as a "disappointment" (e.g., Guillaume Frechette, 2003, p. 42). We disagree and regard variation as a scientific challenge rather than a disappointment.

One reaction to parameter variation across games is to abandon the search for a single rule that predicts well across games, and instead build up a catalog of which rules fit in which games. Rather than give up hope, we propose a general learning model which flexibly "self-tunes" the parameters of our earlier EWA model across game, subjects, and time (inspired by self-tuning control and Kalman filtering[3]). The model posits a single family of self-tuning *functions* which can be applied to a wide variety of games. The functions are the same in all games, but the interaction of game structure and experience will generate different parameter values from the functions, both across games and across time within a game. We find that the self-tuning functions can reproduce parameter variation well enough that the model out-predicts models which estimate separate parameters for different games, in forecasting behavior in new games. The model also addresses the concern that the earlier EWA model has too many parameters, because it only has one free parameter that must be fixed a priori or estimated.

Self-tuning EWA also addresses the challenge of predicting behavior in a brand new game

as the economic value of a theory relative to the average subject's implicit theory.

[3]Erev, Bereby-Meyer and Roth (1999) use a response sensitivity parameter which is "self-adjusting" in a similar way. They divide the fixed parameter (λ in the notation below) by the average absolute deviation of received payoffs from historically average payoff. As a result, when equilibration occurs the payoff variance shrinks and the adjusted λ rises, which sharpens convergence to equilibrium. (This remedies a problem noted by Jasmina Arifovic and John Ledyard, 2002– in games with many strategies learning models predict less convergence in strategy frequencies than is observed.) Their model also has the interesting property that when the game changes and received payoffs suddenly change, the adjusted λ rises, flattening the profile of choice probabilities which causes players to "explore" their new environment more.

where we do not have prior data and hence must arbitrarily specify free parameters. Since parameters self-adjust based on information feedback, they will "repair" poor initial estimates and refine them so that the accuracy of the model will not deteriorate over time. The self-tuning functions can also potentially do something that people do well– respond to changes in structural parameters of games over time (like demand shocks or the entry of new players)– which is something that models which estimate parameters from history do poorly.

Another interesting property of self-tuning EWA is that it can automatically transform into familiar special cases in different games, and at different points of time in a game. One key function is the decay rate $\phi_i(t)$, which weights previous experience. The self-tuning value of $\phi_i(t)$ falls when another player's behavior changes sharply (a kind of deliberate forgetting resulting from change-detection). This change is like switching from a fictitious play belief learning rule (which weights all past experience equally) to a more rapidly-adjusting Cournot belief learning rule (which weights only the previous period). The change captures the idea that if their opponents are suddenly behaving differently than in the past, players should ignore their distant experience and concentrate only on what happened recently. (Albert Marcet and Juan Pablo Nicolini, 2004, use a similar change-detection model to explain learning in repeated monetary hyperinflations).

The second key function is the weight given to foregone payoffs, $\delta_{ij}(t)$, in updating the numerical attractions of strategies. This weight is one for strategies that yield better payoffs than the payoff a player actually received, and zero for strategies that yield worse than actual payoffs. If players start out choosing bad strategies, then the weights on most alternative strategies are 1, and the rule is approximately the same as belief learning (i.e., reinforcing all strategies' payoffs equally). But when players are in a strict pure-strategy equilibrium, all other strategies have worse payoffs and so the rule is equivalent to choice reinforcement.

While the self-tuning EWA model is honed on data from several game experiments, it should be of interest to economists of all sorts because learning is important for virtually every area of economics. Much as in physical sciences, in the lab we can see clearly how various theories perform in describing behavior (and giving valuable advice) before applying those theories to more complex field applications. Since self-tuning EWA is designed to work across many economic domains, sensible extensions of it could be applied to field settings such as evolution of economic institutions (e.g., internet auctions or pricing), investors and policymakers learning about equity market fluctuations or macroeconomic phenomena (as in Allan Timmerman, 1993, or Marcet and Nicolini, 2004), and consumer choice (e.g., Ho and Juin-Kuan Chong, 2003).

The paper is organized as follows. In the next section, we describe the self-tuning EWA model and its parametric precursor. In section 3 self-tuning EWA is used to fit and predict data from seven experimental data sets. The hardest kind of prediction is to estimate free parameters on one sample of games and predict play in a new game. Self-tuning EWA does well in this kind of cross-game prediction. Section 4 estimates the economic value for the self-tuning EWA model, three other learning models (EWA, weighted fictitious play, and reinforcement) and an equilibrium benchmark (quantal response equilibrium). Many models have positive economic value, and self-tuning EWA adds the most value. Section 6 concludes. An appendix shows how to correct the economic value of a theory's advice in period t for possible "boomerang" effects that advice can have on earnings from future periods (addressing the "Lucas critique" applied to learning in games).

2 Self-tuning EWA

First, some notation is necessary. For player i, there are m_i strategies, denoted s_i^j (the j-th strategy for player i), which have initial attractions denoted $A_i^j(0)$. Strategies actually chosen by i in period t, and by all other players (who are denoted $-i$) are $s_i(t)$ and $s_{-i}(t)$ respectively. Player i's ex-post payoff of choosing strategy s_i^j in time t is $\pi_i(s_i^j, s_{-i}(t))$ and the actual payoff received is $\pi_i(s_i(t), s_{-i}(t)) \equiv \pi_i(t)$.

For player i, strategy j has a numerical attraction $A_i^j(t)$ after updating from period t experience. $A_i^j(0)$ are initial attractions before the game starts. Attractions determine choice probabilities in period t+1 through a logistic stochastic response function, $P_i^j(t+1) = \frac{e^{\lambda \cdot A_i^j(t)}}{\sum_{k=1}^{m_i} e^{\lambda \cdot A_i^k(t)}}$, where λ is the response sensitivity. Note that $\lambda = 0$ is random response and $\lambda = \infty$ is best-response.

In the parametric EWA model, attractions are updated by

$$A_i^j(t) = \frac{\phi \cdot N(t-1) \cdot A_i^j(t-1) + [\delta + (1-\delta) \cdot I(s_i^j, s_i(t))] \cdot \pi_i(s_i^j, s_{-i}(t))}{N(t-1) \cdot \phi \cdot (1-\kappa) + 1} \tag{2.1}$$

where $I(x, y)$ is an indicator function (equal to zero if $x \neq y$ and one if $x = y$) (see Camerer and Ho, 1999). That is, previous attractions are multiplied by an experience weight $N(t-1)$, decayed by a weight ϕ, incremented by either the payoff received (when $I(s_i^j, s_i(t)) = 1$) or by δ times the foregone payoff (when $I(s_i^j, s_i(t)) = 0$), and pseudo-normalized.

EWA is a hybrid of the central features of reinforcement and fictitious play (belief learning) models. This hybrid is useful if actual learning mixes components of those simpler rules. A hybrid is also useful for evaluating the statistical and economic advantages of complicating the simpler models by adding components (and, consequently, for finding out when simple rules are adequate approximations and when they are not).

Standard reinforcement models assume that only actual choices are reinforced (i.e., $\delta = 0$).[4] When $\delta = 0$ and $\kappa = 1$ the rule is like the cumulative reinforcement model of Roth and Erev (1995). When $\delta = 0$ and $\kappa = 0$ it is like the averaging reinforcement model of Erev and Roth (1998). A central insight from the EWA formulation is that weighted fictitious play belief learning is *exactly* the same as a generalization of reinforcement in which all foregone payoffs are reinforced by a weight $\delta = 1$ (when $\kappa = 0$).[5] Intuitively, the EWA form allows *both* the stronger focus on payoffs that are actually received, as in reinforcement (i.e., $\delta < 1$) *and* the idea that foregone payoffs usually affect learning when they are known, as in fictitious play (i.e., $\delta > 0$).

For parsimony, we set $\kappa = 0$ and $N(0) = 1$ in self-tuning EWA.[6] Initial attractions $A_i^j(0)$ are determined by a "cognitive hierarchy" model of reasoning (with parameter $\tau = 1.5$; see Camerer, Ho and Chong, 2003, 2004).[7]

---

[4] See Calvin Harley (1981), Roth and Erev (1995), Sarin and Vahid (2001) (cf. Richard Bush and Frederick Mosteller, 1955; John Cross, 1983; Patrick McAllister, 1991; Brian Arthur, 1991). Choice reinforcement is most sensible when players do not know the foregone payoffs of unchosen strategies. However, several studies show that providing foregone payoff information affects learning (See Dilip Mookerjhee and Barry Sopher (1994), Amnon Rapoport and Erev (1998), and John Van Huyck, Ray Battalio and Frederick Rankin (1996)), which suggests that players do not simply reinforce chosen strategies. Rajiv Sarin and Farshid Vahid, 2003, show that "spilling" over payoff reinforcement to neighboring strategies can explain the rapid pace of learning in the Van Huyck et al. (1996) game.

[5] See also Cheung and Friedman, 1997, pp. 54-55; Drew Fudenberg and David Levine, 1998, pp. 1084-1085; Ed Hopkins, 2002.

[6] We set $\kappa = 1$ because it does not seem to affect fit much (e.g., Camerer and Ho, 1999; Ho, Xin Wang, and Camerer, 2003). $N(0)$ was included in the original EWA so that Bayesian learning models are nested as a special case– $N(0)$ represents the strength of prior beliefs. It is fixed to one here because its influence frequently fades rapidly as an experiment progresses.

[7] There are at least three other ways to pin down the initial attractions $A_i^j(0)$. You can either use the first-period data to "burn-in" the attractions, assume all initial attractions are equal (which leads to uniformly-distributed first-period choices), or use a decision rule like assuming that players best-respond to a uniform distribution. The cognitive hierarchy approach uses a specific mixture of the latter two rules but adds further steps of iterated thinking in a precise way. Stahl and Wilson (1995) and Costa-Gomes, Crawford and Broseta (2001) explore richer reasoning-step models.

These simplifications leave three free parameters– ϕ, δ, and λ. To make the model simple to estimate statistically, and self-tuning, the parameters ϕ and δ are replaced by deterministic functions $\phi_i(t)$ and $\delta_{ij}(t)$ of player i's experience with strategy j, up to period t. These functions determine numerical parameter values for each player, strategy, and period, which are then plugged into the EWA updating equation above to determine attractions in each period. Updated attractions determine choice probabilities according to the logit rule, given a value of λ. Standard maximum-likelihood methods for optimizing fit can then be used to find which λ fits best.[8]

2.1 The change-detector function $\phi_i(t)$

The decay rate ϕ which weights lagged attractions is sometimes called "forgetting" (an interpretation which is carried over from reinforcement models of animal learning). While forgetting obviously does occur, the more interesting variation in $\phi_i(t)$ across games, and across time within a game, is a player's perception of how quickly the learning environment is changing. The function $\phi_i(t)$ should therefore "detect change". When a player sense that other players are changing, a self-tuning $\phi_i(t)$ should dip down, putting less weight on distant experience. As in physical change detectors (e.g., security systems or smoke alarms), the challenge is to detect change when it is really occurring, but not falsely mistake small fluctuations for real changes too often.

The core of the $\phi_i(t)$ change-detector function is a "surprise index", which is the difference between other players' recent strategies and their strategies in previous periods. To make exposition easier, we describe the function for games with pure-strategy equilibria (it is suitably modified for games with mixed equilibria, as noted below). First define a history vector, across the other players' strategies k, which records the historical frequencies (including the last period t) of the choices by other players. The vector element $h_i^k(t)$ is $\frac{\sum_{\tau=1}^{t} I(s_{-i}^k, s_{-i}(\tau))}{t}$.[9] The recent 'history' $r_i^k(t)$ is a vector of 0's and 1's which has a one for strategy $s_{-i}^k = s_{-i}(t)$ and 0's for all other strategies s_{-i}^k (i.e., $r_i^k(t) = I(s_{-i}^k, s_{-i}(t))$). The surprise index $S_i(t)$ simply sums up

[8] If one is interested only in the hit rate– the frequency with which the predicted choice is the same as what a player actually picked– then it is not necessary to estimate λ. The strategy that has the highest attraction will be the predicted choice. The response sensitivity λ only dictates how *frequently* the highest-attraction choice is actually picked, which is irrelevant if the statistical criterion is hit rate.

[9] Note that if there is more than one other player, and the distinct choices by different other player's matter to player i, then the vector is an $n-1$- dimensional matrix if there are n players.

the squared deviations between the cumulative history vector $h_i^k(t)$ and the immediate recent history vector $r_i^k(t)$; that is,

$$S_i(t) = \sum_{k=1}^{m_{-i}} (h_i^k(t) - r_i^k(t))^2. \tag{2.2}$$

Note that this surprise index varies from zero (when the last strategy the other player chose is the one they have always chosen before) to two (when the other player chose a particular strategy 'forever' and suddenly switches to something brand new). The change-detecting decay rate is:

$$\phi_i(t) = 1 - \frac{1}{2} \cdot S_i(t). \tag{2.3}$$

Because $S_i(t)$ is between zero and two, ϕ is always (weakly) between one and zero.

The numerical boundary cases illuminate intuition: If the other player chooses the strategy she has always chosen before, then $S_i(t) = 0$ (player i is not surprised) and $\phi_i(t) = 1$ (player i does not decay the lagged attraction at all, since what other players did throughout is informative). If the other player chooses a new strategy which was never chosen before in a very long run of history, $S_i(t) = 2$ and $\phi_i(t) = 0$ (player i decays the lagged attraction completely and 'starts over'). Note that since the observed behavior in period t is included in the history $h_i^k(t)$, $\phi_i(t)$ will typically not dip to zero. For example, if a player chose the same strategy for each of nine periods and a new strategy in period 10, then $S_i(t) = (.9 - 0)^2 + (1 - .1)^2 = 2 \cdot .81$ and $\phi_i(t) = 1 - .5(2 \cdot .81) = .19$.

In games with mixed equilibria (and no pure equilibria), a player should expect other players' strategies to vary. Therefore, if the game has a mixed equilibrium with W strategies which are played with positive probability, the surprise index defines recent history over a window of the last W periods (e.g., in a game with four strategies that are played in equilibrium, $W = 4$). Then $r_i^k(t) = \sum_{k=1}^{m_{-i}} [\frac{\sum_{\tau=t-W+1}^{t} I(s_{-i}^k, s_{-i}(\tau))}{W}]$.

A sensible property of $S_i(t)$ is that the surprisingness of a new choice should depend not only on how often the new choice has been chosen before, but also on how *how variable* previous choices have been. Incorporating this property requires $\phi_i(t)$ to be larger when there is more dispersion in previous choices, which is guaranteed by squaring the deviations between current and previous history. (Summing absolute deviations between $r_i(t)$ and $h_i(t)$, for example, would not have this property.) If previously observed relative frequencies of strategy k are denoted f_k, and the recent strategy is h, then the surprise index is $(1 - f_h)^2 + \sum_{k \neq h} (f_k - 0)^2$. Holding f_h

constant, this index is minimized when all frequencies f_k with $k \neq h$ are equal. In the equal-f_k case, the surprise index is $S_i(t) = (m_{-i} - 1)/m_{-i}$ and $\phi_i(t) = (m_{-i} + 1)/2m_{-i}$, which has a lower bound of .5 in games with large m_{-i} (many strategies).

The opposite case is when an opponent has previously chosen a single strategy in every period, and suddenly switches to a new strategy. In that case, $\phi_i(t)$ is $\frac{2t-1}{t^2}$. This expression declines gracefully toward zero as the string of identical choices up to period t grows longer. (For $t = 2$, 3, 5 and 10 the $\phi_i(t)$ values are .75, .56, .36, and .19.) The fact that the ϕ values decline with t expresses the principle that a new choice is bigger surprise (and should have an associated lower ϕ) if it follows a *longer* string of *identical* choices which are different from the surprising new choice. It also embodies the idea that dipping $\phi_i(t)$ too low is a mistake which should be avoided because it erases the history embodied in the lagged attraction. So $\phi_i(t)$ only dips low when opponents have been reliably choosing one strategy for a very long time, and then switch to a new one.

Another interesting special case is when unique strategies have been played in every period up to $t - 1$, and another unique strategy is played in period t. (This is often true in games with large strategy spaces.) Then $\phi_i(t) = .5 + \frac{1}{2t}$, which starts at .75 and asymptotes at .5 as t increases.

In the first few periods of a game, $\phi_i(t)$ will not dip much below one (because the t-th period experience is included in the recent history $r_i(t)$ vector and also folded into the cumulative history $h_i(t)$). But in these periods players often learn rapidly. Since it makes sense to start with a low value of $\phi_i(0)$ to express players' responsiveness in the first few periods, in the empirical implementation we smooth the $\phi_i(t)$ function by starting at 0.5, and gently blending in the updated values according to $\hat{\phi}_i(t) \equiv .5/t + (t-1)\phi_i(t)/t$.

2.2 The attention function, $\delta_{ij}(t)$

The parameter δ is the weight on foregone payoffs. Presumably this is tied to the attention subjects pay to alternative payoffs, ex-post. Subjects who have limited attention are likely to focus on strategies that would have given higher payoffs than what was actually received, because these strategies present missed opportunities (cf. Sergiu Hart and Andreu Mas-Collel (2001), who show that such a regret-driven rule converges to correlated equilibrium.) To capture this property, define[10]

$$\delta_{ij}(t) = \begin{cases} 1 & \text{if } \pi_i(s_i^j, s_{-i}(t)) > \pi_i(t), \\ 0 & \text{otherwise.} \end{cases} \tag{2.4}$$

That is, subjects reinforce chosen strategies (where the top inequality necessarily binds) and all unchosen strategies with better payoffs (where the inequality is strict) with a weight of one. They reinforce unchosen strategies with equal or worse payoffs by zero.

Note that this $\delta_{ij}(t)$ can transform the self-tuning rule into special cases over time. If subjects are strictly best-responding (ex post), then no other strategies have a higher ex-post payoff so $\delta_{ij}(t) = 0$ for all strategies j which were not chosen, which reduces the model to choice reinforcement. However if they always choose the worst strategy, then $\delta_{ij}(t) = 1$, which corresponds to weighted fictitious play. If subjects neither choose the best nor the worst strategy, the updating scheme will push them (probabilistically) towards those strategies that yield better payoffs, as is both characteristic of human learning and normatively sensible.

The updating rule is a natural way to formalize and extend the "learning direction" theory of Selten and Stoecker (1986). Their theory consists of an appealing property of learning: Subject move in the direction of ex-post best-response. Broad applicability of the theory has been hindered by defining 'direction' only in terms of numerical properties of ordered strategies (e.g., choosing 'higher prices' if the ex-post best response is a higher price than the chosen price). The self-tuning $\delta_{ij}(t)$ defines the 'direction' of learning set-theoretically, as shifting probability toward the set of strategies with higher payoffs than the chosen ones.

The self-tuning $\delta_{ij}(t)$ also creates the "exploration-exploitation" shift in machine learning (familiar to economists from multi-armed bandit problems). In low-information environments, it makes sense to explore a wide range of strategies, then gradually lock in to a choice with a good historical relative payoffs. In self-tuning EWA, if subjects start out with a poor choice, many unchosen strategies will be reinforced by their (higher) foregone payoffs, which shifts choice probability to those choices and captures why subjects "explore". As equilibration occurs, only the chosen strategy will be reinforced, thereby producing an "exploitation" or "lock-in" (cf.

[10] In games with unique mixed-strategy equilibrium, we use $\delta_{ij}(t) = \frac{1}{W}$ if $\pi_i(s_i^j, s_{-i}(t)) \geq \pi_i(t)$ and 0 otherwise. This modification is driven by the empirical observation that estimated δ's are often close to zero in mixed games (which might also be due to misspecified heterogeneity, see Nathaniel Wilcox 2003). Using only $\delta_{ij}(t)$ without this adjustment produces slightly worse fits in the two mixed-equilibrium games examined below where the adjustment matters (patent-rate games the Mookerjhee-Sopher games).

Erev et al., 1999). This is behaviorally very plausible. The updating scheme also helps to detect any change in environment. If a previously optimal response becomes inferior because of an exogenous change, other strategies will have higher ex-post payoffs, which trigger higher $\delta_{ij}(t)$ values (and reinforcement of superior payoffs), which guides players to re-explore better strategies.

The self-tuning $\delta_{ij}(t)$ function can also be seen as a reasonable all-purpose rule which conserves a scarce cognitive resource– attention. The hybrid EWA model showed that weighted fictitious play is equivalent to generalized reinforcement in which all strategies are reinforced. But reinforcing many strategies takes attention. As equilibration occurs, the set of strategies which receive positive $\delta_{ij}(t)$ weights shrinks so attention is conserved when spreading attention widely is no longer useful. When an opponent's play changes suddenly, the self-tuning $\phi_i(t)$ value drops. This change reduces attractions (since lagged attractions are strongly decayed) which spreads choice probability over a wider range of strategies due to the logit response rule. This implies that the strategy which happens to be chosen may no longer be optimal, which re-allocates $\delta_{ij}(t)$ attention over a wider range of better-responses. Thus, the self-tuning system can be seen as procedurally rational (in Herbert Simon's language) because it follows a precise algorithm and is designed to express the basic features of how people learn– by exploring a wide range of options, locking in when a good strategy is found, but reevaluating when environmental change demands reallocation of attention.

A theorist's instinct is to derive conditions under which flexible learning rules choose parameters optimally, which is certainly a direction to explore in future research (cf. Dana Heller and Sarin, 2000; Jens Josephson, 2001). However, broadly-optimal rules will likely depend on the set of games an all-purpose learning agent encounters, and also may depend sensitively on how cognitive costs are specified (and should also jibe with data on the details of neural mechanisms, which are not yet well-understood). So it is unlikely that a universally optimal rule will be found that can always beat rules which adapt locally.

Our approach is more like exploratory work in machine learning. Machine learning theorists try to develop robust heuristic algorithms which learn effectively in a wide variety of low-information environments (see Sutton and Barto 1998). Good machine learning rules are not provably optimal but perform well on tricky test cases and natural problems like those which good computerized robots need to perform (navigating around obstacles, hill-climbing on rugged landscapes, difficult pattern recognition, and so forth).

Before proceeding to estimation, it is useful to summarize some of the properties of a self-tuning model. First, the use of simple fictitious play and reinforcement theories in empirical analysis are often justified by the fact that they have few free parameters. The self-tuning EWA is useful by this criterion as well because it requires estimating only one parameter, λ (which is difficult to do without in empirical work). Second, the functions in self-tuning EWA naturally vary across time, people, games, and strategies. The potential advantage of this flexibility is that the model can predict across new games better than parametric methods. Whether this advantage is realized will be examined below. Third, the self-tuning parameters can endogenously shift across rules. Early in a game, where opponent choices are varying a lot and players are likely to make ex-post mistakes, the model automatically generates low values of $\phi_i(t)$ and high $\delta_{ij}(t)$ weights– it resembles Cournot belief learning. As equilibration occurs and behavior of other players stabilizes, $\phi_i(t)$ rises and $\delta_{ij}(t)$ falls– it resembles reinforcement learning. The model therefore keeps a short window of history (low ϕ) and pays a lot of attention (high δ) when it should, early in a game, and conserves those cognitive resources by remembering more (high ϕ) and attending to fewer foregone strategies (low δ) when it can afford to, as equilibration occurs.

3 Self-tuning EWA predictions within and across games

In this section we compare in-sample fit and out-of-sample predictive accuracy of self-tuning EWA, its predecessor (EWA) where parameters are freely estimated, and the one-parameter Quantal Response Equilibrium model benchmark (Richard McKelvey and Thomas Palfrey, 1995).[11] The goal is to see whether self-tuning EWA functions can produce game-specific parameters which are similar to values estimated separately in different games. In addition, we use a jackknife approach by estimating a common set of parameters on $n-1$ of the n games and use the estimated parameters to predict choices in the remaining game, to judge how well models predict across games (cf. Erev and Roth, 1998).

We use seven games: Two matrix games with unique mixed strategy equilibrium (Dilip Mookerjhee and Barry Sopher, 1997); R&D patent race games (Amnon Rapoport and Wilfred Amaldoss, 2000); a median-action order statistic coordination game with several players (John Van Huyck, Ray Battalio, and Richard Beil, 1990); a continental-divide coordination game,

[11] Our working paper also reports fit statistics and estimates from belief learning and reinforcement models.

in which convergence behavior is extremely sensitive to initial conditions (Van Huyck, Joseph Cook, and Battalio, 1997); a coordination game in which players choose whether to enter a large or small market (Amaldoss and Ho, 2001); dominance-solvable p-beauty contests (Ho, Camerer, and Keith Weigelt, 1998); and a price matching game called traveler's dilemma (Monica Capra, Jacob Goeree, Rosario Gomez and Charles Holt, 1999).

Table 1 summarizes features of these games and the data. Three of the games are described in detail below.[12] Many different games are studied because a main goal is to see how well self-tuning EWA can explain cross-game variation. Sampling widely is also a good way to test robustness of any model of learning or equilibrium. Models that are customized to explain one game are insightful, but not as useful as games which explain disparate patterns with one general model (see Roth and Erev, 1995; Goeree and Holt, 2001).

3.1 Estimation method

Consider a game where N subjects play T rounds. For a given player i, the likelihood function of observing a choice history of $\{s_i(1), s_i(2), \ldots, s_i(T-1), s_i(T)\}$ is given by:

$$\Pi_{t=1}^{T} P_i^{s_i(t)}(t|\lambda) \tag{3.1}$$

The joint likelihood function $L(\lambda)$ of observing all players' choice is given by

$$L(\lambda) = \Pi_i^N \{\Pi_{t=1}^{T} P_i^{s_i(t)}(t|\lambda)\} \tag{3.2}$$

To determine the predicted probabilities $P_i^{s_i(t)}(t|\lambda)$, we start with initial attractions $A_i^j(0)$ (which are the same for all i) determined by the predictions of the cognitive hierarchy model (Camerer, Ho and Chong, 2003, 2004) using $\tau = 1.5$.[13] After each period, the functions $\phi_i(t)$

[12] The other four games are: Mixed-equilibrium games studied by Mookerjhee and Sopher (1997) which have four or six strategies, one of which is weakly-dominated; the nine-player median-action game studied by Van Huyck et al. (1990), in which players choose integer strategies 1-7 and earn payoffs increasing linearly in the group median and decreasing linearly in the squared deviation from the median; a traveler's dilemma game (Capra et al., 1999) in which players choose numbers from 80 to 200 and each player receives the a payoff equal to the minimum of the chosen numbers and the player who chose the low number receives a bonus of R from the player who chose the high number; and a coordination game (Amaldoss and Ho, 2001) in which n players simultaneously enter a large or small market and earn $2n$ (n) divided by the number of entrants if they enter the large (small) market.

[13] In the games we study, for example, one-step behavior predicts choices of 35 in beauty contests, 7 in continental-divide games, 4 in median-action games, the large pot in entry-choice games, 5 and 4 in patent-race games for strong and weak players, and $200 - 2R$ in traveler's dilemma games.

and $\delta_{ij}(t)$ are updated according to player i's experience and applied to the self-tuning EWA model (fixing λ) to determine updated attractions according to the EWA rule. The updated attractions produce predicted probabilities $P_i^{s_i(t)}(t|\lambda)$. Using the first 70% of the subjects in each game, we determine the value of λ that maximizes the total likelihood over the 70% of the subjects. Then the value of λ is frozen and used to forecast behavior of the entire path of the remaining 30% of the subjects.[14] Payoffs were converted to inflation-adjusted dollars (which is important for cross-game forecasting) and scaled by subtracting the lowest possible payoffs. Randomized bootstrap resampling is used to calculate parameter standard errors.

In addition to self-tuning EWA, we estimated both the parametric EWA model and the one-parameter quantal response equilibrium (QRE) model. QRE is a static no-learning benchmark, which is tougher competition than Nash equilibrium.

3.2 Model fit and predictive accuracy

The first question to address is how well models fit and predict on a game-by-game basis (i.e., when EWA parameters are estimated separately for each game). As noted, to limit overfitting we estimate parameters using 70% of the subjects (in-sample calibration) and use those estimates to predict choices by the remaining 30% (out-of-sample validation). For in-sample estimation we report a Bayesian information criterion (BIC) which subtracts a penalty $\frac{k \cdot ln(NT)}{2}$ from the $L(\lambda)$ value. For out-of-sample validation we report the log-likelihood ($L(\lambda)$) on the hold-out sample of 30% of the subjects.

Table 2 shows the results. The Table also show the choice probability implied by the average likelihood[15] compared to the probability if choices were random. Across games, self-tuning EWA predicts a little worse than EWA in out-of-sample prediction, though generally more accurately than QRE. This is not surprising since the parametric EWA model uses four extra free parameters (δ, ϕ, κ and $N(0)$) in each game. The correlation between the ϕ function of self-tuning and the ϕ estimates of EWA is 0.77 and the corresponding correlation for the δ is 0.52.[16] However, self-tuning EWA is better in the pooled estimation where a common set

[14] We also tried using the first 70% of the observations from each subject, then forecasted the last 30%. The results are similar.

[15] That is, divide the total log likelihood by the number of subject-periods. Exponentiating this number gives the geometric mean ("average") predicted probability of the strategies that are actually played.

[16] The parameter estimates are reported in Table A.1.

of parameters (except λ, which is always game-specific) was used for EWA. QRE fits worst in both individual games and pooled estimation, which is no surprise because it does not allow learning.

A more challenging robustness test is to estimate all parameters on six of the seven games, then use those parameters to predict choices in the remaining seventh game for all subjects. This is done for each of the seven games, one at a time. Cross-game prediction has been used by others but only within similar games in a class (2x2 games with mixed equilibria, Erev and Roth, 1998; and 5x5 symmetric games, Stahl, forthcoming). Our results test whether fitting a model on a coordination game, say, can predict behavior in a game with mixed equilibrium. This is the most ambitious use of learning models across games since Roth and Erev (1995) who demonstrated the importance of this kind of cross-game forecasting.

Table 3 reports results from the cross-game prediction. By this measure, self-tuning EWA has the highest cross-game likelihood in three games; EWA is highest in four other games. QRE is the least accurate in six out of the seven games.

Likelihood values summarize the model fit over time, strategies and subjects; but they do not allow one to gauge how model fit changes over time and across strategies. To get a more nuanced feel for the fit between data and models, the next section produces graphs using predicted and relative frequencies for three games which are exemplars of three classes: The patent race game has a unique mixed-strategy equilibrium, the continental divide coordination game has multiple Pareto-ranked pure equilibria, and the beauty contest games are dominance-solvable.[17]

3.3 Games with unique mixed strategy equilibrium: Patent races

In the patent race game two players, one strong and one weak, are endowed with resources and compete in a patent race. The strong player has an endowment of 5 units and the weak player has an endowment of 4 units (Rapoport and Amaldoss, 2000). They simultaneously invest an integer amount up to their endowments, i_{strong} and i_{weak}. The player whose investment is strictly larger earns 10 minus their investment. A player whose investment is less than or equal to the other player's investment earns no payoff, and ends up with their endowment minus the investment.

[17] Corresponding graphs for *all* games can be seen at http://www.bschool.nus.edu.sg /depart/mk /bizcjk /fewa.htm so readers can draw their own conclusions about other games.

The game has an interesting strategic structure. The strong player can guarantee a payoff of five by investing the entire endowment $i_{strong} = 5$ (out-spending the weak player), which strictly dominates investing zero ($i_{strong} = 0$). Eliminating the strong player's dominated strategy $i_{strong} = 0$ then makes $i_{weak} = 1$ a dominated strategy for the weak player (since she can never win by investing one unit). Iterating in this way, the strong player deletes $i_{strong} \in \{0, 2, 4\}$ and the weak player deletes $i_{weak} \in \{1, 3\}$ by iterated application of strict dominance. The result is a unique mixed equilibrium in which strong players invest five 60% of the time and play their other two (serially) undominated strategies of investing one and three 20% of the time, and weak players invest zero 60% of the time and invest either two or four 20% of the time.

Thirty six pairs of subjects played the game in a random matching protocol 160 times (with the role switched after 80 rounds); the 36 pairs are divided into 2 groups where random matching occurs within group. Since overall choice frequencies do not change visibly across time, and are rather close to equilibrium predictions, our plots show frequencies of *transitions* between period $t-1$ and period t strategies to focus on changes across time. Figures 1a-d show the empirical transition matrix and predicted transition frequencies across the five strategies ($i_{strong} \in \{0, 1, \cdots 5\}$) for strong players, using the within-game estimation and pooling across all subjects. (Weak-player results are similar.)

The key features of the data are the high percentage of transitions from 5 to 5, almost 40%, and roughly equal numbers of transitions (about 5%) from 1 to 1, and from 1 to 5 or vice versa. The figures show that QRE does not predict differences in transitions at all. The challenge for explaining transitions is that after investing $i_{strong} = 5$, about 80% of the time the strong player knows that the weak player invested only 0 or 2. Most learning models predict that strong players should therefore invest less, but the figures show that about half the time the strong players invest 5 again. The self-tuning and parametric EWA models explain the relatively low rate of downward transitions by multiplying the high foregone payoffs, in the case where the strong player invested 5 and the weak player invested nothing or two, by a relatively low value of δ. This means the attractions for low i_{strong} are not updated as much as the chosen strategy $i_{strong} = 5$, which explains the persistence of investing and the low rate of switching. The low value of δ, which is estimated to be .30 in EWA and averages .16 in self-tuning EWA, is one way of expressing why strong players are sluggish in switching down from $i_{strong} = 5$.

3.4 Games with multiple pure strategy equilibria: Continental divide game

Van Huyck et al. (1997) studied a coordination game with multiple equilibria and extreme sensitivity to initial conditions, which we call the continental divide game (CDG).

Subjects play in cohorts of seven people. Subjects choose an integer from 1 to 14, and their payoff depends on their own choice and on the median choice of all seven players. The payoff matrix is constructed so that there are two pure equilibria (at 3 and 12) which are Pareto-ranked (12 pays \$1.12 and 3 pays \$.60). The best-response correspondence bifurcates in the middle: For all $M \leq 7$, the best response to a median M is strictly between M and 3. For high medians $M \geq 8$, the best response is strictly between M and 12. The payoff at 3 is about half as much as at 12. This game captures the possibility of extreme sensitivity to initial conditions (or path-dependence).

Their experiment used 10 cohorts of seven subjects each, playing for 15 periods. At the end of each period subjects learned the median, and played again with the same group in a partner protocol.

Figures 2a-d show empirical frequencies (pooling all subjects) and model predictions. The key features of the data are: Bifurcation over time from choices in the middle of the range (5-10) to the extremes, near the equilibria at 3 and 12; and late-period choices are more sharply clustered around 12 than around 3. (Figure 2a hides strong path-dependence: Groups which had first-period $M \leq 7$ ($M \geq 8$) *always* converged toward the low (high) equilibrium.) Notice also that strategies 1-4 are never chosen in early periods, but are frequently chosen in later periods; and oppositely, strategies 7-9 are frequently chosen in early periods but never chosen in later periods. A good model should be able to capture these subtle effects by "accelerating" low choices quickly (going from zero to frequent choices in a few periods) and "braking" midrange choices quickly (going from frequent 7-9 choices to zero).

QRE fits poorly because it predicts no movement. Self-tuning EWA and parametric EWA are fairly accurate and able to explain the key features of the data– viz., convergence toward the two equilibria, sharper convergence around 12 than around 3, and rapid increase in strategies 1-4 and extinction of 7-9. Both this game and the game above show how self-tuning EWA is able to reproduce the predictive power of EWA without having to estimate parameters.

3.5 Games with dominance-solvable pure strategy equilibrium: p-Beauty Contests

In the p-beauty contests of Ho et al. (1998), seven players simultaneously choose numbers in [0,100]. The player whose number is closest to a known fraction (either 0.7 or 0.9) of the group average wins a fixed prize. Their experiment also manipulated experience of subjects because half of them played a similar game before. Initial choices are widely dispersed and centered around 45. When the game is repeated, numbers gradually converge toward the equilibrium 0 and experienced subjects converge much faster towards equilibrium.

Figures 3a-f show empirical frequencies and model predictions of self-tuning EWA and EWA broken down by experience of subjects. (As in the earlier plots, QRE fits badly so it is omitted.) Self-tuning EWA tracks behavior about as accurately as EWA for inexperienced subjects, and is substantially more accurate for experienced subjects. The cross-game EWA estimate of ϕ used is 0.83, which is significantly higher than the in-game estimate ϕ of 0.31. The cross-game estimate of δ used is 0.29, lower than the in-game estimate δ of 0.70. These cross-game values create the sluggishness in responding to surprises (ϕ is too high) and to the many better strategies available (δ is too low).

The average self-tuning function $\phi_i(t)$ is 0.58, which is relatively more responsive to surprises than the 0.83 cross-game estimate. Coupled with the $\delta_{ij}(t)$ function, which by definition is totally responsive to better strategies, this helps to explain why self-tuning EWA predicts behavior better. The sluggish response to surprise of EWA also explains why EWA is less accurate than its self-tuning cousin in experienced session; experienced subjects create more surprises than their inexperienced counterparts as they move faster to convergence.

4 Economic Value of Models

Most criteria used to judge fit and predictive accuracy of models are purely statistical, or are roughly equivalent to familiar statistics. But economic applications of theories call for a financial measure of what good theories are worth. In this section we define a theory's economic value as the increase in an average subject's profit from substituting learning theory recommendations– i.e., best responses based on the theory's prediction about what others will do– for actual choices.

To define the economic value of a learning theory formally, we need to start with some notation. Denote the aggregate choice frequency of strategy s_{-i}^k of the potential pool of player i's opponents in round t by $f_k(t)$. Player i's expected ex-ante payoff of choosing $s_i(t)$, given the ex-post frequencies of strategy choices $f_k(t)$, is $E(\pi_i(s_i(t)) = \sum_{k=1}^{m_{-i}} f_k(t) \cdot \pi_i(s_i(t), s_{-i}^k)$.

To measure economic value, use the estimated model parameters from the previous section (out-of-game estimates) and a player's observed experience through period t to generate model predictions about what others will do in $t+1$. That prediction is a probability distribution over choices by others. Optimizing using this predicted distribution produces a choice with the highest expected value. A model x that prescribes an action $s_i(t|x)$ receives an actual payoff of $\pi_i(s_i(t|x), s_{-i}(t))$ and has an expected payoff, with random matching across the population, of $v_x(t) \equiv E(\pi_i(s_i(t|x) = \sum_{k=1}^{K} f^k(t) \cdot \pi_i(s_i(t|x), s_{-i}^k)$. Since the typical player earns $v(s_i(t)) \equiv \sum_{s_i(t)} E(\pi_i(s_i(t))$, the added economic value from model x is $\Delta EV(x) = \sum_t v_x(t) - v(s_i(t))$.

A reasonable benchmark is where players guess accurately the *distribution* of opponents' plays in the population (i.e., they know $f_k(t)$) but do not know exactly what the opponent they are paired with will do. Then their optimal strategy is $s_i(t|RE) = \text{argmax}_j \sum_{k=1}^{K} f^k(t) \cdot \pi_i(s_i^j, s_{-i}^k)$. This "rational expectations" benchmark yields an expected payoff of $v_{RE}(t) \equiv E(\pi_i(s_i(t|RE)) = \sum_{k=1}^{K} f^k(t) \cdot \pi_i(s_i(t|RE), s_{-i}^k)$.

A common misconception about the economic value of learning models is that a good learning model will necessarily have little economic value, because if it captures learning then players will anticipate it and reduces its value. This misconception assumes players are "sophisticated" (Milgrom and Roberts, 1991)– i.e., players understand how others learn. If players are learning from experience, but are unsophisticated about the possibility that others are learning, then using the learning rule can have economic value. Put simply, if players learn according to a predictable rule, they must not be sophisticated. Put differently, if players are sophisticated about learning of others, then a sophisticated learning model (as in Camerer, Ho and Chong, 2002, or Stahl, 1999) will not have economic value. So if a learning model has economic value compared to the average player, the model's economic value is evidence of the degree of *un*sophistication of players.

In estimating the economic value of learning models, there is a potential subtle "boomerang" effect of current choices on future value (a la the Lucas critique in macroeconomic models). Consider a game like the "traveller's dilemma" (Capra et al., 1999) which we rename a game of price-matching with loyalty. In this game two players choose prices p_1, p_2 from the interval

[80, 120]. Call the low price $p_{low} \equiv min(p_1, p_2)$. The low-price player earns $p_{low} + R$ and the high-price player earns $p_{low} - R$; the value R represents a reward to the low-price firm and a penalty for the high-price firm (if the players name the same price both earn p_{low}).

A player who uses a good learning model to forecast accurately how other players will behave generally undercuts prices. As a result, the economic value from the learning model in a particular period is usually positive. However, choosing a low undercutting price in period t may lead other players to choose even lower prices in period $t+1$, which lowers the future value of the player who undercut price in t. Thus, in evaluating a model's economic value, it is important to account for the effects of a current decision on *future* payoffs. Note that if there are many subjects in a random-matching protocol, this boomerang effect will be minimal because one player's period t choice will not affect future earnings much (since a player will not face his period t opponent again). Furthermore, the boomerang effect can be positive (e.g., in coordination games, where a player who forecasts accurately speeds up the rate of mutually-beneficial coordination). The Appendix describes how we carefully account for boomerang effects.[18]

Table 4 shows the economic value of self-tuning EWA, parametric EWA, QRE, weighted fictitious play belief learning (Fudenberg and Levine, 1988), and an average reinforcement model with adjustment for payoff variation (Erev and Roth, 1998). The table reports the percentage improvement (inclusive of boomerang effects) of each model over the actual payoff. Only self-tuning and EWA consistently add positive value. Self-tuning EWA model adds the most value in four out of seven games. Interestingly, belief learning, reinforcement and QRE models have negative value in two to three of the seven games, implying that subjects' beliefs are more accurate than forecasts implied in these models in these three games.

[18] An even better test is to have human subjects compete with computerized learners, who continually update their forecasts based on a particular learning model (e.g., self-tuning EWA with a prespecified value of λ). Then the economic value of the computerized players' learning rule is easily calculated as how much more (or less) the computerized players earn, compared to the average human subject. Computerizing learning-rule players is nontrivial since it requires players to receive feedback and continually update learning rules in real-time. It lies naturally beyond the scope of this paper.

5 Conclusion

Learning is clearly important for economics. Equilibrium theories are useful because they suggest a possible limit point of a learning process and permit comparative static analysis. But if learning is slow, or the time path of behavior selects one equilibrium out of many, a precise theory of equilibration is crucial for knowing which equilibrium will result, and how quickly.

The theory described in this paper, self-tuning EWA, replaces the key parameters in the EWA learning models with functions that change over time in response to experience. One function is a "change detector" ϕ which goes up (limited by one) when behavior by other players is stable, and dips down (limited by zero) when there is surprising new behavior by others. When ϕ dips down, the effects of old experience (summarized in attractions which cumulate or average previous payoffs) is diminished by decaying the old attraction by a lot. The second "attention" function δ is one for strategies that yield better than actual payoff and zero otherwise. This function ties sensitivity to foregone payoffs to attention, which is likely to be on strategies that give better than actual payoff ex post. Self-tuning EWA is more parsimonious than most learning theories because it has only one free parameter– the response sensitivity λ.

We report fit and prediction of data from seven experimental games using self-tuning EWA, the parameterized EWA model, and quantal response equilibrium (QRE). Both QRE and self-tuning EWA have one free parameter, and EWA has five. We report both in-sample fit (penalizing more complex theories using the Bayesian information criterion) and out-of-sample as well as out-of-game predictive accuracy, to be sure that many complex models do not necessarily fit better.

There are three key results.

First, self-tuning EWA fits and predicts slightly worse than EWA in all seven games; and it produces a functional parameter values for ϕ and δ which roughly track the estimated values of fixed parameters across games. Self-tuning EWA therefore represents one solution to the central problem of flexibly generating EWA-like parameters across games. Because self-tuning EWA generates sensible cross-game parameter variation automatically, it fits and predicts better than other models when games are pooled and common parameters are estimated.

Second, we propose a new criterion for judging the usefulness of theories, called economic value. A theory's economic value is the incremental profit a subject would earn from following the theory's advice rather than making their own choices. Most learning models add economic value at least in some games. Self-tuning EWA add the most economic value in a majority of games. Similar conclusions are drawn when boomerang and Lucas-critique effects are carefully controlled by a new statistical methodology.

Third, the functions in self-tuning EWA are robust across games. This paper added three brand new games (after the first version was written and circulated) to test robustness. The basic conclusions are replicated in these games, which have incomplete information and choices are made by groups rather than individuals (see our working paper).

A next step in this research is to find some axiomatic underpinnings for the functions. Extending the ϕ function to exploit information about ordered strategies might prove useful. And since self-tuning EWA is so parsimonious, it is useful as a building block for extending learning theories to include sophistication (players anticipating that others are learning; see Stahl, 1999) and explain "teaching" behavior in repeated games (Camerer, Ho and Chong 2002; Cooper and Kagel, 2001).

The theory is developed to fit experimental data, but the bigger scientific payoff will come from application to naturally-occurring situations. If learning is slow, a precise theory of economic equilibration is just as useful for predicting what happens in the economy as a theory of equilibrium. For example, institutions for matching medical residents and medical schools, and analogous matching in college sororities and college bowl games, developed over decades (Roth and Xing, 1994). Bidders in eBay auctions learn to bid late to hide their information about an object's common value (Bajari and Hortacsu, 1999). Consumers learn over time what products they like (Ho and Chong, 2003). Learning in financial markets can generate excess volatility and returns predictability, which are otherwise anomalous in rational expectations models (Timmerman, 1993). Sargent (1999) argues that learning by policymakers about expectational Phillips' curves and the public's perceptions of inflation explains macroeconomic behavior in the last couple of decades. Good theories of learning should be able to explain these patterns and help predict how new institutions will evolve, how rapidly bidders learn to wait, and which new products will succeed. Applying self-tuning EWA, and other learning theories, to field domains is therefore an important goal of future research.

References

[1] Amaldoss, Wilfred and Teck-Hua Ho, "EWA Learning in Games with Different Group Sizes," Marketing Department Working Paper, The Wharton School, 2001.

[2] Arifovic, Jasmina and John Ledyard, "Computer Testbeds: The Dynamics of Groves-Ledyard Mechanisms," unpublished manuscript, 2002.

[3] Arthur, Brian, "Designing Economic Agents That Act Like Human Agents: A Behavioral Approach to Bounded Rationality," *American Economic Review*, 81(2), (1991), pp. 353-359.

[4] Bajari, Patrick and Ali Hortacsu, "Winner's Curse, Reserve Prices and Endogenous Entry: Empirical Insights from Ebay Auctions," Working Paper, Department of Economics, Stanford University, 1999.

[5] Bush, Robert and Frederick Mosteller, *Stochastic models for learning*, New York: Wiley, 1955.

[6] Camerer, Colin F., *Behavioral Game Theory: Experiments on Strategic Interaction*, Princeton:Princeton University Press, 2003.

[7] Camerer, Colin F. and Teck-Hua Ho, "Experience Weighted Attraction Learning in Normal Form Games," *Econometrica*, 67, (1999), pp. 827-873.

[8] Camerer, Colin F., Teck-Hua Ho and Juin-Kuan Chong, "Sophisticated EWA Learning and Strategic Teaching in Repeated Games," *Journal of Economic Theory*, 104, (2002), pp. 137-188.

[9] Camerer, Colin F., Teck-Hua Ho and Juin-Kuan Chong, "Models of Thinking, Learning and Teaching Games," *The American Economic Review*, 93(2), (2003), pp. 192-195.

[10] Camerer, Colin F., Teck-Hua Ho and Juin-Kuan Chong, "A Cognitive Hierarchy Model of Games," *Quarterly Journal of Economics*, (2004), forthcoming.

[11] Capra, Monica, Jacob Goeree, Rosario Gomez and Charles Holt, "Anomalous Behavior in a Traveler's Dilemma," *American Economic Review*, 89(3), (1999), pp. 678-690.

[12] Cheung, Yin-Wong and Daniel Friedman, "Individual Learning in Normal Form Games: Some Laboratory Results," *Games and Economic Behavior*, 19, (1997), pp. 46-76.

[13] Cooper, David and John Kagel, "Learning and Transfer in Signaling Games," working paper, Department of Economics, Case Western Reserve University, 2001.

[14] Costa-Gomes, M. and V. Crawford and B. Broseta, "Cognition and Behavior in Normal-form Games: An Experimental Study," *Econometrica*, 69, (2001), 1193-1235.

[15] Erev, Ido and Alvin E. Roth, "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria," *American Economic Review*, 88(4), (1998), pp. 848-81.

[16] Erev, Ido, Yoella Bereby-Meyer and Alvin E. Roth, "The Effect of Adding a Constant to All Payoffs: Experimental Investigation, and a Reinforcement Learning Model with Self-Adjusting Speed of Learning," *Journal of Economic Behavior and Organization*, 39, (1999), pp. 111-128.

[17] Frechette, Guillaume, "Learning in A Multilateral Bargaining Experiment," Harvard University working paper, September 12, 2003. http://www.people.hbs.edu/ gfrechette/ print/ frechette_learning.pdf

[18] Fudenberg, Drew and David K. Levine, *The Theory of Learning in Games*, Boston: MIT Press, 1998.

[19] Goeree, J. K. and C. A. Holt, "Ten little treasures of game theory, and ten intuitive contradictions," *American Economic Review*, 91(6), (2001), pp. 1402-1422

[20] Hart, Seigiu and Andreu Mas-Colell, "A General Class of Adaptive Strategies," *Journal of Economic Theory*, 98, (2001), pp. 26-54.

[21] Harley, Calvin, "Learning the Evolutionary Stable Strategies," *Journal of Theoretical Biology*, 89, (1981), pp. 611-633.

[22] Heller, Dana and Rajiv Sarin, "Parametric Adaptive Learning," University of Chicago Working Paper, 2000.

[23] Ho, Teck-Hua, Colin F. Camerer, Colin and Juin-Kuan Chong, "Economic Value of fEWA: A Functional Theory of Learning in Games," 2001. http://www.bschool.nus.edu.sg /depart/mk/bizcjk/fewa.htm

[24] Ho, Teck-Hua, Colin F. Camerer and Keith Weigelt, "Iterated Dominance and Iterated Best Response in Experimental "p-Beauty Contests"," *American Economic Review*, 88, (1998), pp. 947-969.

[25] Ho, Teck-Hua and Juin-Kuan Chong, "A Parsimonious Model of SKU Choice," *Journal of Marketing Research*, 40(3), (2003), pp. 351-65.

[26] Ho, Teck-Hua, Xin Wang and Colin F. Camerer, "Individual Differences in the EWA Learning with Partial Payoff Information," Marketing Department Working Paper 99-010, The Wharton School, 2002.

[27] Hopkins, Edward, "Two Competing Models of how People Learn in Games," *Econometrica*, 70(6), (2002), pp. 2141-2166.

[28] Josephson, Jens, "A Numerical Analysis of the Evolutionary Stability of Learning Rules," Stockholm School of Economics SSE/EFI paper no. 474, 2001. http://swopec.hhs.se/hastef/abs/hastef0474.htm.

[29] Marcet, Albert and Juan Pablo Nicolini, "Recurrent Hyperinflations and Learning," *American Economic Review*, (2004), forthcoming.

[30] McAllister, Patrick H., "Adaptive Approaches to Stochastic Programming," *Annals of Operations Research*, 30, (1991), pp. 45-62.

[31] McKelvey, Richard and Thomas Palfrey, "Quantal Response Equilibria for Normal Form Games," *Games and Economic Behavior*, 10, (1995), pp. 6-38.

[32] Milgrom, Paul, and John Roberts, "Adaptive and Sophisticated Learning in Normal Form Games," *Games and Economic Behavior*, 3, (1991), pp. 82-101.

[33] Mookerjhee, Dilip, and Barry Sopher, "Learning Behavior in an Experimental Matching Pennies Game," *Games and Economic Behavior*, 7, (1994), pp. 62-91.

[34] Mookerjhee, Dilip, and Barry Sopher, "Learning and Decision Costs in Experimental Constant-sum Games," *Games and Economic Behavior*, 19, (1997), pp. 97-132.

[35] Rapoport, Amnon and Ido Erev, "Coordination, "magic", and reinforcement learning in a market entry game," *Games and Economic Behavior*, 23, (1998), pp. 146-175.

[36] Rapoport, Amnon and Wilfred Amaldoss, "Mixed Strategies and Iterative Elimination of Strongly Dominated Strategies: An Experimental Investigation of States of Knowledge," *Journal of Economic Behavior and Organization*, 42, (2000), pp. 483-521.

[37] Roth, Alvin E. and Xiaolin Xing, "Jumping the Gun: Imperfections and Institutions Related to the Timing of Market Transactions," *American Economic Review*, 84, (1994), pp. 992-1044.

[38] Roth, Alvin E. and Ido Erev, "Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term," *Games and Economic Behavior*, 8(1), (1995), pp. 164-212.

[39] Roth, Alvin E., Greg Barron, Ido Erev and Robert Slonim, "Equilibrium and Learning in Economic Environments: the Predictive Value of Approximations," Harvard University Working Paper, 2002.

[40] Sarin, Rajiv and Farshid Vahid, "Predicting How People Play Games: A Simple Dynamic Model of Choice," *Games and Economic Behavior*, 34, (2001), pp. 104-122.

[41] Sarin, Rajiv and Farshid Vahid, "Strategy Similarity and Coordination," *Economic Journal*, forthcoming, (2003).

[42] Sargent, Thomas, *The Conquest of American Inflation*, Princeton: Princeton University Press, 1999.

[43] Selten, Reinhard and Rolf Stoecker, "End Behavior in Sequences of Finite Prisoner's Dilemma Supergames: A Learning Theory Approach," *Journal of Economic Behavior and Organization*, 7, (1986), pp. 47-70.

[44] Stahl, Dale O., "Sophisticated Learning and Learning Sophistication," University of Texas at Austin Working Paper, 1999.

[45] Stahl, Dale O, "Local Rule Learning: Theory and Evidence," *Games and Economic Behavior*, forthcoming.

[46] Stahl, Dale O. and P. Wilson, "On Players Models of Other Players: Theory and Experimental Evidence," *Games and Economic Behavior*, 10, (1995), 213-54.

[47] Sutton, Richard and Andrew Barto, *Reinforcement Learning: An Introduction*, Boston: MIT Press, 1998.

[48] Timmerman, Allan G, "How Learning in Financial Markets Generates Excess Volatility and Predictability in Stock Prices.' *Quarterly Journal of Economics*, 108, (1993), pp. 1135-1145.

[49] Van Huyck, John, Raymond Battalio and Richard Beil, "Tacit Cooperation Games, Strategic Uncertainty, and Coordination Failure," *The American Economic Review*, 80, (1990), pp. 234-248.

[50] Van Huyck, John, Raymond Battalio and Frederick Rankin, "Selection Dynamics and Adaptive Behavior Without Much Information," Working Paper, Department of Economics, Texas A & M University, 1996.

[51] Van Huyck, John, Joseph Cook and Raymond Battalio, "Adaptive Behavior and Coordination Failure," *Journal of Economic Behavior and Organization*, 32, (1997), pp. 483-503.

[52] Weibull, J., *Evolutionary Game Theory*, Cambridge: MIT Press, 1995.

[53] Wilcox, N., "Heterogeneity and Learning Models," University of Houston working paper, 2003. http://www.uh.edu/ñwilcox/papers/hetero_and_learning_W.pdf

6 Appendix: Adjusting for boomerang effects in computing economic value

The procedure described in section 4 allows us to estimate how much subjects would have earned if they were to adopt a model's prescription. The procedure works if subjects play the game once and their choices have no future ramifications (Camerer, Ho and Chong, 2003, 2004). The procedure has two potential problems (a la Lucas's critique) if the game is repeated.

First, if player i were to follow a model's prescription in period t, then her opponent would have changed his actions in future rounds creating a boomerang effect that could have changed the evolution of plays (i.e., $f^k(\tau), \forall \tau > t$ would have changed). We call this "cross-sectional boomerang effect".

Second, the adoption of a model's prescription in period t could have influenced the player's payoffs from period $t+1$ to T. For instance, following a model's recommendation could help the player in the immediate round but hurt her in future rounds because it leads to a pareto inferior outcome (e.g., a lower number in traveller's dilemma). We call this "longitudinal boomerang effect". Consequently, it is possible that a subject who adopted a theory's recommendation would make choices that would reduce long-run payoffs.

It is possible to control for these boomerang effects statistically. First we divide actual subject choices into those that are consistent with a theory's recommendation and those that are not, and see which choices earn more. If the advice is good, the subjects who "coincidentally" followed the advice should earn more and this should control for "cross-sectional boomerang effect".

The "longitudinal boomerang effect" is controlled by estimating both current and future ramifications of a theory's recommendation so that its overall economic impact can be assessed. We describe how this can be done next.

Let's consider the following set of simple regression equations:

$$\pi_i(t) \quad = \quad \alpha_{0,t} + \alpha_{1,t} \cdot I(s_i(t), s_i(t|x)), \quad t = 1, \ldots, T \tag{6.1}$$

where $I(s_i(t), s_i(t|x))$ is an indicator function and is equal to 1 if $s_i(t) = s_i(t|x)$ and 0 otherwise. A model x adds positive economic value iff $\sum_{t=1}^{T} \alpha_{1,t} > 0$. The parameters $\alpha_{1,t} (t = 1, \ldots, T)$ control for cross-sectional boomerang effect because they are derived based on differences in the performance of choices that are consistent with the model's prescription and those that are not. Since subjects do not change their choices as a result of the model's prescription, there is no cross-sectional boomerang effect.

The above set of equations however does not control for longitudinal boomerang effect. Consequently, the economic value of a model in period t, $\alpha_{1,t}$, can be biased in two ways. Except for $\alpha_{1,1}$, $\alpha_{1,t}$ might include longitudinal boomerang effect from following the model's prescriptions from period 1 to $t-1$. In addition, α_{1t} (except for $\alpha_{1,T}$) does not include the economic value of following the model's period t recommendation in periods $t+1$ and beyond. The following set of multiple regressions will fix both problems:

$$\pi_i(t) \quad = \quad \alpha_{0,t} + \sum_{\tau=1}^{t} \alpha_{\tau,t} \cdot I(s_i(\tau), s_i(\tau|x)), \quad t = 1, \ldots, T \tag{6.2}$$

This set of equations regresses players' period t payoffs with whether or not their choices are consistent with the model's prescriptions from period 1 to t. $\alpha_{\tau,t}$ captures the increase in payoff in period t by adopting the model's prescription in period $\tau < t$. The average economic value of a model's period t prescription is:

$$\alpha_t = \sum_{\tau=t}^{T} \alpha_{t,\tau} \tag{6.3}$$

Note that α_1 has T terms $(\alpha_{1,1} + \alpha_{1,2} + \ldots, +\alpha_{1,T})$, α_T has one term $(\alpha_{T,T})$, and α_t in general has $T - t + 1$ terms. Hence, learning theories can be ranked by the following overall economic value:

$$\Delta EV(x) = \sum_{\tau=1}^{T} (1 - f(t|x))\alpha_t \tag{6.4}$$

where $f(t|x)$ is the ex-post frequency of strategic choices in period t that are consistent with model x. The economic values are adjusted for instances where strategic choices are consistent with the model where no additional value is generated. This overall economic value can be broken into the current period effect $(\sum_{\tau=1}^{T}(1 - f(t|x))\alpha_{\tau,\tau})$ and the boomerang effect $((1 - f(t|x)) \cdot (\sum_{\tau=1}^{T} \alpha_t - \sum_{\tau=1}^{T} \alpha_{\tau,\tau}))$.

Note that the regressions look *backward*, using the historical coefficients $\alpha_{\tau,t}$ of period τ matching of model predictions with actual choices on period t payoffs to measure the effect of τ choices on future periods t. These coefficients are then used to measure the *future* effects of choosing the model-advised choice in period τ on future period payoffs. The approach makes sense because the only way to estimate a model behavior's future period effects is to look back from future period payoffs to the model behaviors which preceded those payoffs. A numerical example can help make this clear. Consider the continental divide game (Van Huyck et al., 1997). Table A.2 summarizes choices of 3 subjects in the first 3 periods of a game. The table also includes the result of the regression analysis to decompose the payoff into current effect and boomerang effects. For example, in period 3, subject 1 chooses 12 that yields a payoff of 118. This payoff of 118 can be decomposed into $\alpha_{0,3} + \alpha_{1,3} + \alpha_{2,3} + \alpha_{3,3} = 117 + 0 - 3 + 4$ where $\alpha_{1,3}$ and $\alpha_{2,3}$ control for boomerang effect from period 1 and 2 respectively. In this example, the model's prescription yields an economic value of 9 which consists of $6, -1$ and 4 for period 1, 2, and 3 respectively. There is no boomerang effect in period 1 but period 2 has a negative boomerang effect of -3. In other words, although subjects gains 2 units by following model's period 2 description, these subjects actually make 3 units less in period 3.

Table A.3 shows the results of adjusted economic value and its decomposition. Economic value is positive for many games, which shows that there is some degree of disequilibrium and models can, in principle, help people make more money. Across the five models, self-tuning EWA performs the best in three out of the five games. In fact, it is the only model that has positive economic value in median action and continental divide game. It also adds the most value in the pot game. These results suggest that self-tuning EWA has a clear edge in enhancing economic value in coordination games. By this metric, EWA and belief learning perform better than reinforcement model; both are better than reinforcement model in four out of five games. QRE is the worst in three out of the five games.

Table 1: A Description of the Seven Games Used in the Estimation of Various Learning Models

Game	Number of Players	Number of Strategies	Number of Pure Strategy Equilibria	Number of Subjects	Number of Rounds	Matching Protocol	Experimental Treatment	Description of Games
Mixed Strategies Mookerjhee and Sopher (1997)	2	4,6	0	80	40	Fixed	Stake Size	A constant-sum game with unique mixed strategy equilibrium.
Patent Race Rapoport and Amaldoss (2000)	2	5,6	0	36	80	Random	Strong vs Weak	Strong (weak) player invests between 0 and 5 (0 and 4) and the higher investment wins a fixed prize.
Continental Divide Van Huyck et al. (1997)	7	14	2	70	15	Fixed	None	A coordination game with two pure strategy equilibria
Median Action Van Huyck et al. (1990)	9	7	7	54	10	Fixed	None	A order-statistic game with individual payoff decreases in the distance between individual choice and the median
Pot Games Amaldoss and Ho (2001)	3,6,9,18	2	1	84	25 (manual) 28 (computer)	Fixed	Number of Players	An entry game where players must decide which of the two ponds of sizes 2n and n they wish to enter. Payoff is the ratio of the pond size and number of entries.
Price Matching (Traveller's Dilemma) Capra et al. (1999)	2	121[1]	1	52	10	Random	Penalty Size	Players choose claims between 80 and 200. Both players get lower claim but the high-claim player pays a penalty to the low-claim player.
p-Beauty Contest Ho et al. (1998)	7	101	1	196	10	Fixed	Experienced vs. Inexperienced	Players simultaneously choose a number from 0 to 100 and the winner whose number is closet to p (<1) times the group average

Note 1: Continuous strategies of 80 to 200 are discretized to 121 integer strategies

Table 2: Model Fit

Data Set	Mixed Strategies	Patent Race	Continental Divide	Median Action	Pot Games	p-Beauty Contest	Price Matching	Pooled [2]
Total Sample Size	3200	5760	1050	540	2217	1960	520	15247
<u>In-sample Calibration</u>								
Sample Size	2240	4000	735	380	1478	1380	360	10573
BIC (Bayesian Information Criterion) [1]								
Self-tuning EWA	-3199	-4367	-1207	-287	-938	-5265	-1408	-16690
EWA	-3040	-4399	-1091	-293	-931	-4987	-1103	-16646
QRE	-3442	-6686	-1923	-549	-1003	-6254	-1420	-21285
Average Probability								
Self-tuning EWA	24.0%	33.6%	19.5%	47.3%	53.1%	2.2%	2.0%	
EWA	26.0%	33.5%	23.2%	48.1%	53.9%	2.7%	4.9%	
QRE	21.6%	18.8%	7.3%	23.8%	50.8%	1.1%	2.0%	
Random	20.4%	18.3%	7.1%	14.3%	50.0%	1.0%	0.8%	
<u>Out-of-sample Validation</u>								
Sample Size	960	1760	315	160	739	580	160	4674
Log-likelihood								
Self-tuning EWA	-1406	-1857	-515	-91	-441	-2392	-581	-7315
EWA	-1342	-1876	-482	-89	-433	-2203	-532	-7377
QRE	-1441	-3006	-829	-203	-486	-2667	-607	-9240

Note 1: BIC (Bayesian Information Criterion) is given by LL - (k/2)*log(N*T) where k is the number of parameters, N is the number of subjects and T is the number of periods.

Note 2: A common set of parameters, except game-specific λ, is estimated for all games. Each game is given equal weight in LL estimation.

Table 3: Out-of-sample Prediction Using Out-game Estimates

Data Set	Mixed Strategies	Patent Race	Continental Divide	Median Action	Pot Games	p-Beauty Contest	Price Matching
Total Sample Size	3200	5760	1050	540	2217	1960	520
Log-likelihood							
Self-tuning EWA	-4618	-7074	-1720	-385	-1398	-8468	-2179
EWA	-4733	-6321	-1839	-457	-1365	-7806	-2140
QRE	-4667	-9132	-2758	-957	-1533	-8911	-2413
Average Probability							
Self-tuning EWA	23.6%	29.3%	19.4%	49.0%	53.2%	1.3%	1.5%
Random	20.4%	18.3%	7.1%	14.3%	50.0%	1.0%	0.8%

Table 4: Economic Value

Data Set	Mixed Strategies	Patent Race	Continental Divide [2]	Median Action [2]	Pot Games	p-Beauty Contest [2]	Price Matching
Actual Payoff	1560	2657	872	432	304	519	192
Economic Value Achieved as a Percetage of Actual Payoff [1]							
%V(Rational Expectation)	17.7%	47.9%	1.9%	0.9%	35.9%	585.4%	19.3%
%V(Self-tuning EWA)	11.8%	3.8%	1.4%	0.7%	6.9%	53.6%	15.5%
%V(EWA)	3.9%	5.0%	0.8%	0.5%	4.7%	49.8%	15.2%
%V(Belief Based)	5.5%	1.4%	-0.7%	-0.5%	3.4%	48.9%	13.9%
%V(Reinforcement)	11.9%	5.0%	-5.2%	-0.1%	7.3%	-27.6%	9.7%
%V(QRE)	2.5%	1.3%	-8.6%	-0.5%	8.1%	-60.2%	12.0%

Note 1: We assume that each bionic subject use the respective model to predict other's behavior and best responds with the strategy that yields the highest expected payoff.

Note 2: The expected value of each strategy in these games is computed with 1000 simulated instances for a given round due to high computational burden for actual derivation.

Table A.1: Parameter Estimates

Data Set	Mixed Strategies	Patent Race	Continental Divide	Median Action	Pot Games	p-Beauty Contest	Price Matching	Pooled
Self-tuning EWA								
ϕ function	0.89	0.89	0.69	0.85	0.80	0.58	0.63	0.76
δ function	0.09	0.16	0.44	0.18	0.23	0.34	0.29	0.25
λ	3.98	9.24	4.43	5.30	7.34	2.39	10.20	5.86
EWA								
ϕ	0.97	0.91	0.72	0.73	0.86	0.31	0.80	0.79
δ	0.19	0.30	0.90	0.94	0.00	0.70	0.40	0.41
κ	0.82	0.15	0.77	0.99	0.92	0.91	0.80	0.28
N0	0.67	0.73	0.36	0.14	0.00	0.17	0.39	0.77
λ	0.34	4.27	13.83	18.55	1.61	2.57	9.88	6.33
QRE								
λ	1.12	0.81	1.83	28.45	3.37	0.69	29.83	9.44

Table A.2: An Example of Economic Value Decomposition

Period, t	1	2	3
Median Outcome, $s_{-i}(t)$	8	10	13
Model Choice, $s_i(t\|x)$	10	11	12
Subject 1			
Subject Choice, $s_1(t)$	10	11	12
Actual Payoff, $\pi[s_1(t)\ s_{-1}(t)]$	89	100	118
Subject 2			
Subject Choice, $s_2(t)$	10	10	13
Actual Payoff, $\pi[s_2(t)\ s_{-2}(t)]$	89	98	117
Subject 3			
Subject Choice, $s_3(t)$	10	11	11
Actual Payoff, $\pi[s_3(t)\ s_{-3}(t)]$	89	100	114

$\alpha_{\tau,t}$	t=1	t=2	t=3	α_τ
τ=0	83	98	117	
τ=1	6	0	0	6
τ=2		2	-3	-1
τ=3			4	4

Table A.3: Economic Value - Current Period and Boomerang Effects

Data Set	Mixed [1] Strategies	Patent [1] Race	Continental Divide	Median Action	Pot Games	p-Beauty [2] Contest	Price [2] Matching
Actual Payoff	1560	2657	872	432	304	519	192
Number of Subjects	80	36	70	54	84	196	52
Total Economic Value (As % of Observed Payoff)							
%V(Rational Expectation)	37.8%	-11.0%	4.3%	0.2%	114.3%	-	-
%V(Self-tuning EWA)	23.8%	13.7%	4.2%	0.5%	36.2%	-	-
%V(EWA)	28.6%	8.6%	-1.4%	-0.5%	15.6%	-	-
%V(Belief Based)	36.1%	20.1%	-10.7%	-1.3%	22.2%	-	-
%V(Reinforcement)	24.1%	7.6%	-17.0%	-1.2%	21.9%	-	-
%V(QRE)	2.2%	19.9%	-24.0%	-1.3%	24.8%	-	-
Current Period Effect							
%V(Rational Expectation)	16.8%	10.0%	2.5%	0.5%	52.1%	-	-
%V(Self-tuning EWA)	13.6%	3.9%	2.4%	0.5%	12.6%	-	-
%V(EWA)	9.3%	3.4%	1.1%	0.4%	10.9%	-	-
%V(Belief Based)	12.1%	4.0%	-2.9%	-1.1%	9.6%	-	-
%V(Reinforcement)	9.6%	3.3%	-4.8%	-0.4%	12.8%	-	-
%V(QRE)	7.3%	4.0%	-6.5%	-1.4%	15.0%	-	-
Boomerang Effect							
%V(Rational Expectation)	21.0%	-21.0%	1.7%	-0.2%	62.2%	-	-
%V(Self-tuning EWA)	10.2%	9.8%	1.8%	0.0%	23.7%	-	-
%V(EWA)	19.3%	5.2%	-2.5%	-0.8%	4.7%	-	-
%V(Belief Based)	24.1%	16.2%	-7.8%	-0.2%	12.6%	-	-
%V(Reinforcement)	14.6%	4.3%	-12.2%	-0.9%	9.1%	-	-
%V(QRE)	-5.1%	15.9%	-17.5%	0.0%	9.7%	-	-

Note 1: For these 2 datasets, due to the fact that there are more rounds than subjects, instead of regressing for all T rounds, we only regress on the next/last 10 and 20 rounds respectively for mixed strategies and patent race

Note 2: The number of strategic choices consistent with model recommendation is less than 10% of total observations. Hence, we do not perform the same analysis for this game.

Figure 1 Transition Matrices for Patent Race

Figure 1a: Empirical Transition

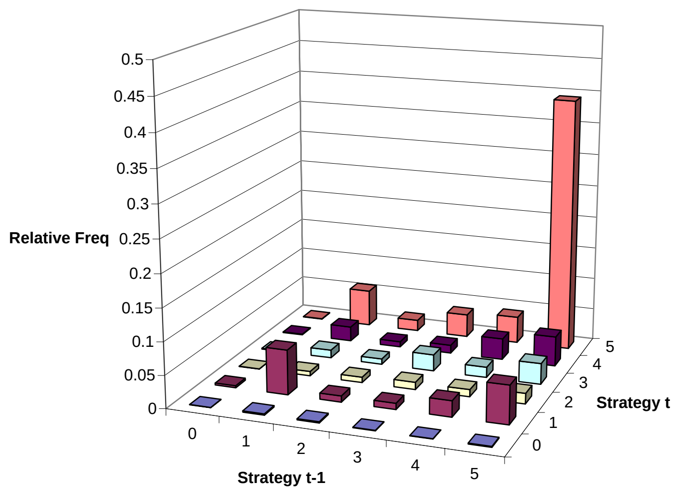

Figure 1b: Self-tuning EWA

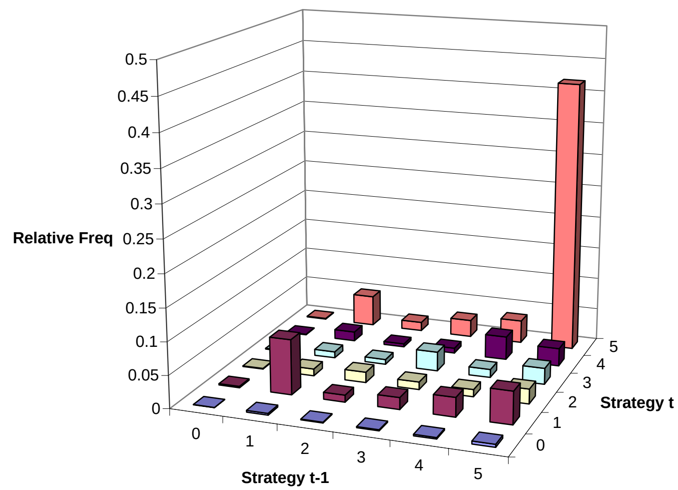

Figure 1c: Parametric EWA

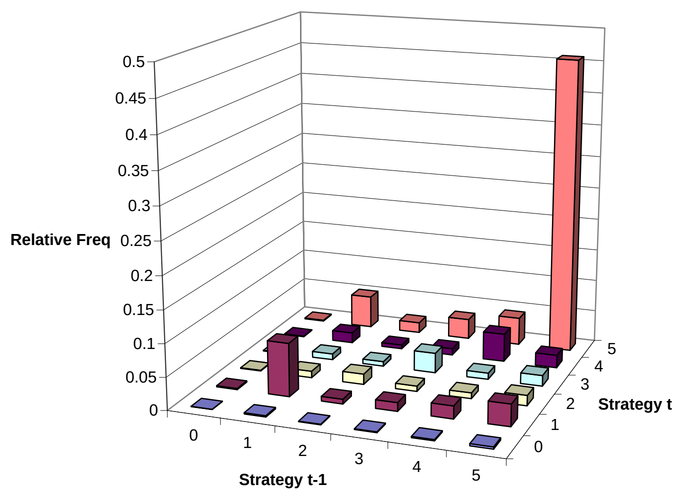

Figure 1d: Quantal Response Equilibrium

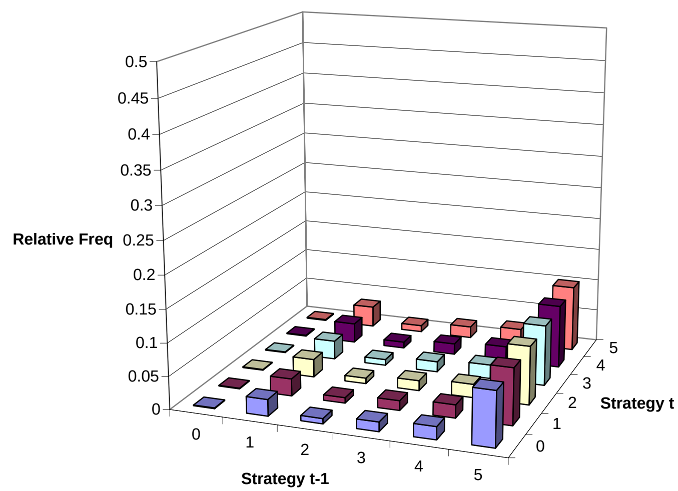

Figure 2: Empirical Frequency and Model Predictions for Continental Divide

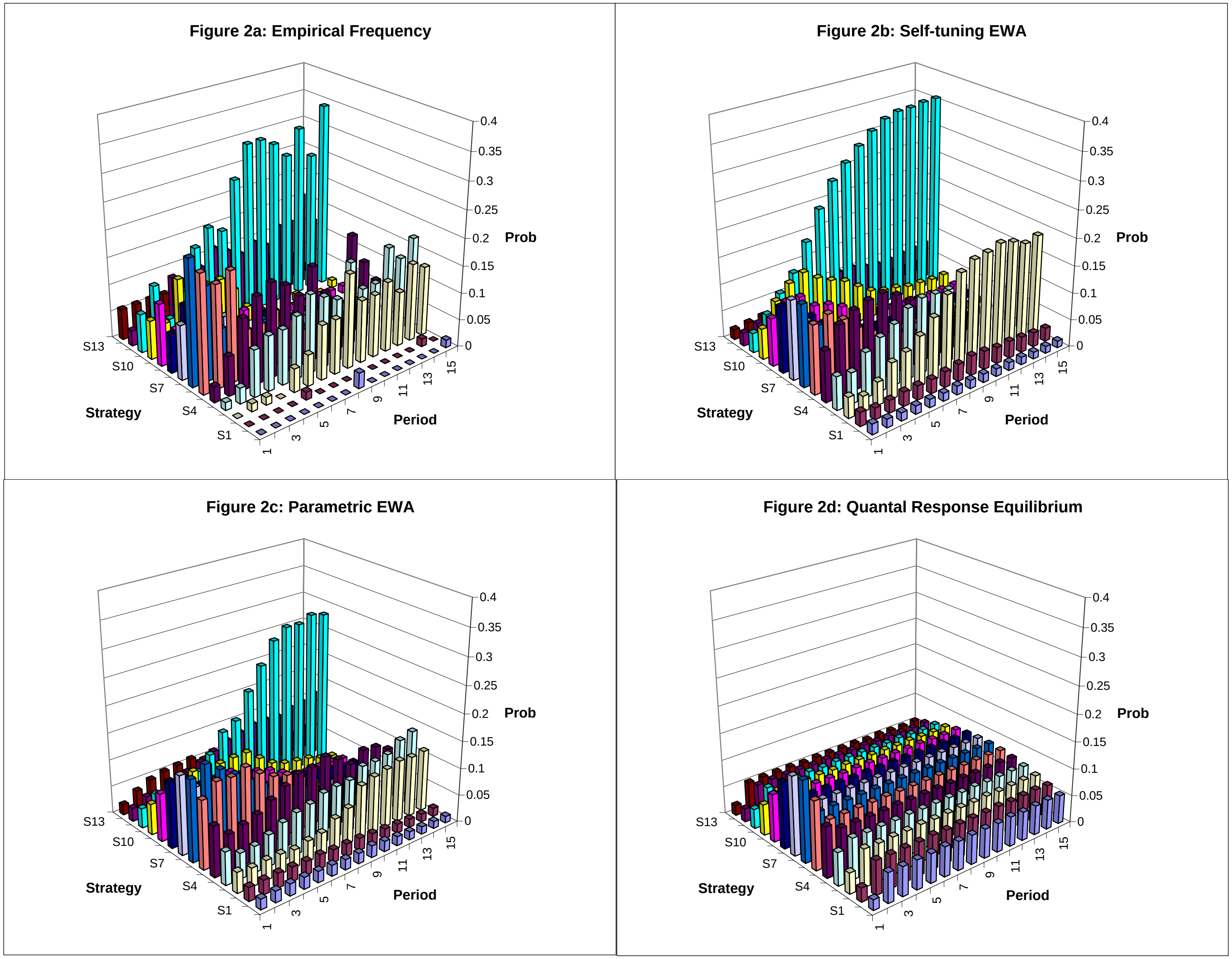

Figure 3: Empirical Frequency and Model Predictions for p-Beauty Contest

Figure 3a: Empirical Frequency (Experienced Subject)

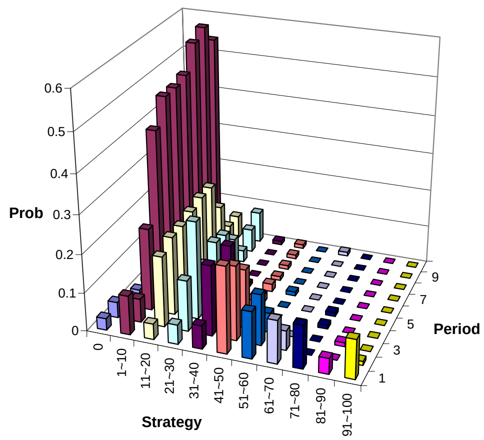

Figure 3d: Empirical Frequency (Inexperienced Subject)

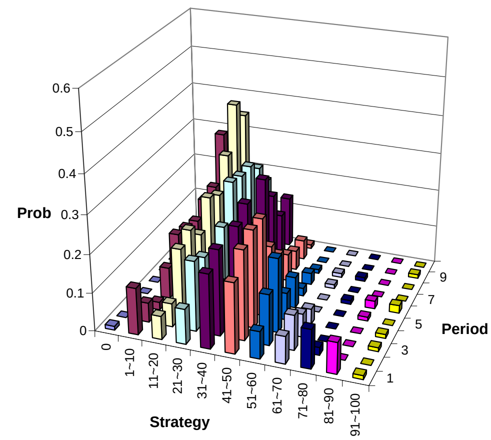

Figure 3b: Self-tuning EWA (Experienced Subject)

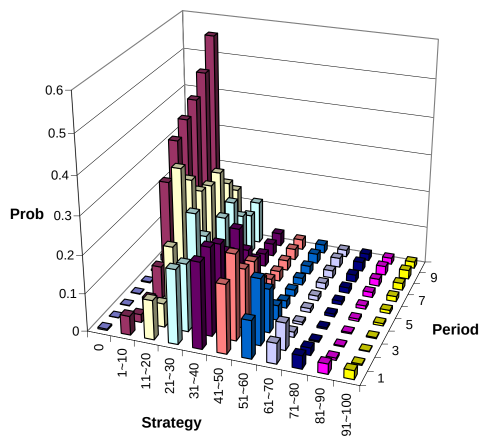

Figure 3e: Self-tuning EWA (inexperienced Subject)

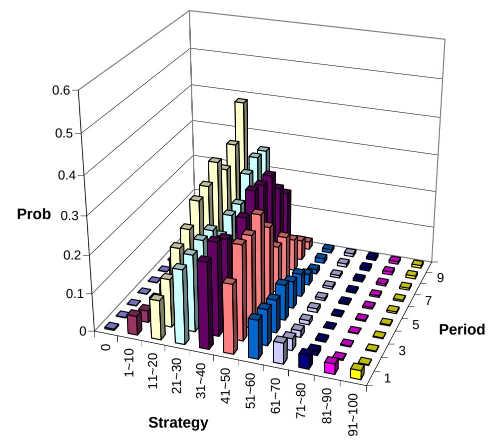

Figure 3c: Adaptive EWA (Experienced Subject)

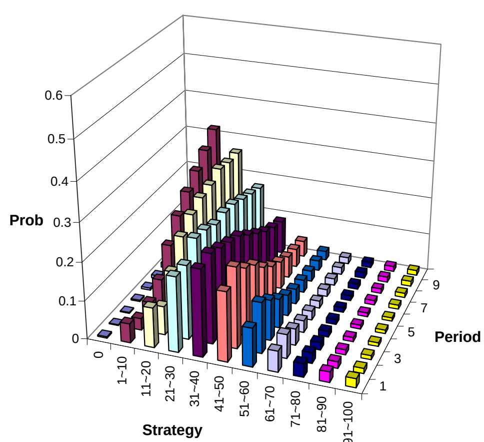

Figure 3f: Adaptive EWA (Inexperienced Subject)