
Volume 133 Number 1 March 2007 ISSN 0022-0531
ELSEVIER
JOURNAL OF
Economic
Theory
Editors: Alessandro Lizzeri and Karl Shell
Associate Editors
Administrative Editors: Delia Cleveland and Susan Witherow Schulze
Available online at
ScienceDirect
www.sciencedirect.com

Available online at www.sciencedirect.com
ScienceDirect

Journal of Economic Theory 133 (2007) 177–198

JOURNAL OF Economic Theory
www.elsevier.com/locate/jet

Self-tuning experience weighted attraction learning in games☆

Teck H. Ho[a,*], Colin F. Camerer[b], Juin-Kuan Chong[c,d]

[a] *Haas School of Business, University of California, Berkeley, Berkeley, CA 94720-1900, USA*
[b] *Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125, USA*
[c] *National University of Singapore, 1 Business Link, Singapore 117592, Singapore*
[d] *Sungkyunkwan University, 53 Myungryun-dong 3-ga, Jongro-gu, Seoul 110-745, Korea*

Received 4 November 2004; final version received 19 December 2005
Available online 13 February 2006

Abstract

Self-tuning experience weighted attraction (EWA) is a one-parameter theory of learning in games. It addresses a criticism that an earlier model (EWA) has too many parameters, by fixing some parameters at plausible values and replacing others with functions of experience so that they no longer need to be estimated. Consequently, it is econometrically simpler than the popular weighted fictitious play and reinforcement learning models. The functions of experience which replace free parameters "self-tune" over time, adjusting in a way that selects a sensible learning rule to capture subjects' choice dynamics. For instance, the self-tuning EWA model can turn from a weighted fictitious play into an averaging reinforcement learning as subjects equilibrate and learn to ignore inferior foregone payoffs. The theory was tested on seven different games, and compared to the earlier parametric EWA model and a one-parameter stochastic equilibrium theory (QRE). Self-tuning EWA does as well as EWA in predicting behavior in new games, even though it has fewer parameters, and fits reliably better than the QRE equilibrium benchmark.

JEL classification: C72; C91

Keywords: Learning; Experience weighted attraction; Quantal response equilibrium; Fictitious play; Reinforcement learning

☆ 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, Guillaume Frechette, and anonymous referees for helpful comments.

* Corresponding author.
E-mail addresses: hoteck@haas.berkeley.edu (T.H. Ho), camerer@hss.caltech.edu (C.F. Camerer), bizcjk@nus.edu.sg (J.-K. Chong).

0022-0531/$ - see front matter
doi:10.1016/j.jet.2005.12.008

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 in order 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: the value of additional parameters in a generalized model, and explaining apparent cross-game variation in learning rules.

The first unresolved issue is how to judge the value of incremental parameters when comparing models with different numbers of parameters. For example, the experience weighted attraction (EWA) learning model [8] econometrically nests the weighted fictitious play [21] and averaging reinforcement learning [19] models. As a result, the EWA model has two and three more parameters than weighted fictitious play and averaging reinforcement learning, respectively. There are many standard statistical procedures for comparing models with different numbers of parameters. One popular approach is generalized likelihood ratio tests which penalize theories for extra free parameters (e.g., the Akaike or Bayesian information criteria). Another approach is to judge the predictive power of a learning model out-of-sample (i.e., after parameter values have been estimated) in both existing and brand new games. The results from prior studies suggest that the additional parameters in EWA are helpful in improving predictions of behavior at least in some games, by either Bayesian information criteria or out-of-sample fit (e.g., [9]).

In this paper we take a different approach: by reducing the number of estimated parameters in the EWA model. Ideally, the methods used to reduce parameters are behaviorally principled and insightful. We reduce parameters in two ways: fixing parameters to specific numerical values; and replacing estimable parameters with functions of data.

Fixing a parameter value numerically can reduce the general EWA specification. There are two parameter restrictions which reduce general EWA learning models so that they still include many familiar special cases. It is also helpful to replace the free parameters which represent the initial strengths of strategies, rather than estimating them, because initial conditions are not of central interest in studying learning (their initial impact typically declines rapidly as learning takes place). We replace estimation of initial conditions with a priori numerical values derived from a cognitive hierarchy model that was designed to predict how people play one shot games [11], although other models could be used instead.

Replacing a free parameter with a function which depends on observed experience means the parameter does not have to be estimated from data. The functional value can also vary over time, games, and different players. The key for these functions to fit well and to generate insight lies in using behaviorally plausible principles to derive the functional forms. We use this functional approach to replace parameters that represent the decay weights placed on recent and distant experience, and the weight placed on foregone payoffs from strategies that were not chosen.

By replacing two EWA parameters and all initial attractions with sensibly determined numerical values and two EWA parameters with behaviorally disciplined functions of experience, the resulting self-tuning EWA model has only one free parameter, fewer than typical versions of reinforcement and weighted fictitious play models. The goal is to make the self-tuning EWA model as predictive as the original parametric EWA model while reducing the number of parameters as much as possible.

The second unresolved issue is how to model cross-game variation in parameters of learning theories. Several studies have shown that the best-fitting parameter values of learning rules vary significantly across games and across subjects (see for example [13,8]; for a comprehensive review, see [7]). Some learning theorists regard parameter variation across games as a "disappointment" (e.g., [20, p. 42]). We disagree and regard variation across games as a natural reflection of the heterogeneity in the learning environment (as determined by the game structure and the history of play). Modelling cross-game variation should be viewed as a scientific challenge.

One could abandon the search for a single rule that predicts well across games, and instead build up a catalog of which rules fit well in which games. We do otherwise, by developing a 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). [1] 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.

Since many familiar learning rules correspond to particular values of decay and foregone payoff weights, endogeneous changes in these parameters through their functional forms in self-tuning EWA also creates variation in rules over time, a reduced-form variant of "rule learning". [2] 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 or surprise). This change is like switching from a pure fictitious play belief learning rule (which weights all past experience equally) to a rapidly adjusting Cournot belief learning rule (which weights only the last period and ignores previous periods). 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. (Marcet and Nicolini [33] 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 or equal 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.

[1] Erev et al. [18] 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 Arifovic and Ledyard [2]—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.

[2] Variation over time in EWA might be more aptly called rule adjustment. It is different than directly reinforcing different rules according to payoffs of strategies those rules recommend, as in Stahl [46].

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.[3] 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 [49,33]), and consumer choice (e.g., [28]).

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 concludes.

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

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

[3] Since its first exposition in 2001, the self-tuning model has succeeded in many robustness tests in which other researchers applied the model to many games that are substantially different from the seven games we validated the model on in this paper. We highlight three pieces of work here: Kocher and Sutter [32] data stretches the model to explain choices made by *groups*; Cabrales et al. [6] and Sovik [44] are games of incomplete information. (All seven games we studied are games with complete information.) Kocher and Sutter [32]'s data are choices made by individual and three-person groups in p-beauty contests. Competing groups appear to learn much faster than individuals. Cabrales et al. [6] report data from a 2×2 coordination game with incomplete information, studying theories of "global games". In their games, players receive a private signal (knowing the distribution of private signals) which shows the payoffs in a signal-dependent stag-hunt coordination game. The game was designed so that iterated dominance leads subjects to select the risk-dominant outcome. Sovik [44] reports data from a zero-sum betting game with asymmetric information. Two players simultaneously choose between a safe outside option and a zero-sum game based on their information sets. The information sets were chosen such that the game is dominance solvable. Players should figure out that they should never bet because they can only bet against players with better information. The self-tuning model fits the data well in all three datasets. Specific estimation result is available in our longer working paper version or by request. We have also recently extended the self-tuning model to study sophistication and quantal response equilibrium in repeated trust and entry games [14].

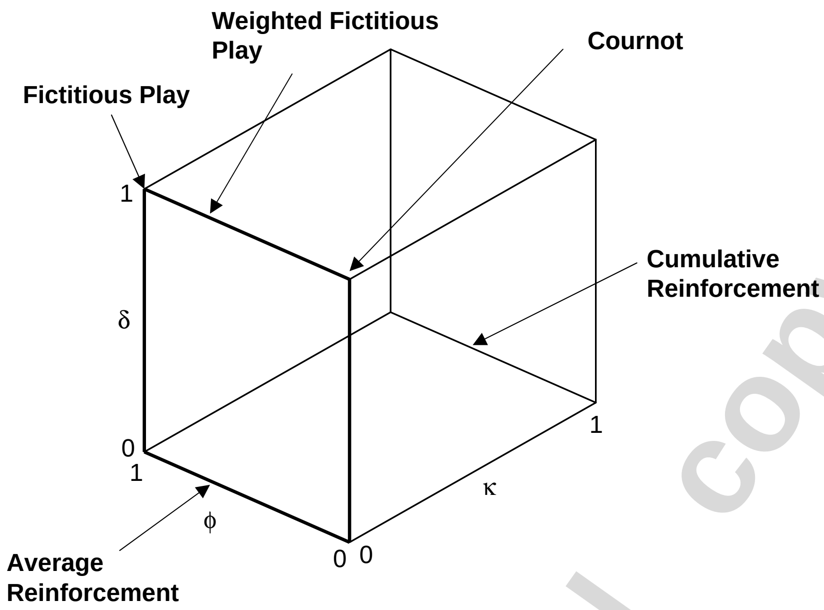

Fig. 1. The EWA learning cube.

where $I(x, y)$ is an indicator function (equal to zero if $x \neq y$ and one if $x = y$) (see [8]). 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 [40]. When $\delta = 0$ and $\kappa = 0$ it is like the averaging reinforcement model of Erev and Roth [19]. 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., $\delta > 0$). Fig. 1 is a cube that shows the universe of learning rules captured by the EWA learning model and the positions of various familiar cases.

As shown in Fig. 1, the front face of the cube ($\kappa = 0$) captures almost all familiar special cases except for the cumulative reinforcement model. The cumulative model has been supplanted by the averaging model (with $\kappa = 0$) because the latter seems to be more robust in predicting behavior in some games (see [19]). This sub-class of EWA learning models is the simplest model that

[4] See Harley [23], Roth and Erev [40], Sarin and Vahid [42] (cf. [5,17,3]). 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 [35,39,51]), which suggests that players do not simply reinforce chosen strategies.

[5] See also Cheung and Friedman [13, pp. 54–55], Fudenberg and Levine [21, pp. 1084–1085], Hopkins [30].

nests averaging reinforcement and weighted fictitous play (and hence Cournot and simple fictitious play) as special cases. It can also capture a weighted fictitious play model using time-varying belief weights (such as the stated-beliefs model explored by Nyarko and Schotter [37]), as long as subjects are allowed to use a different weight to decay lagged attractions over time (i.e., move along the top edge of the side in Fig. 1). There is also an empirical reason to set κ to a particular value. Our prior work suggests that κ does not seem to affect fit much (e.g., [8], [29]).

The initial experience $N(0)$ was included in the original EWA model so that Bayesian learning models are nested as a special case—$N(0)$ represents the strength of prior beliefs. We restrict $N(0) = 1$ here because its influence fades rapidly as an experiment progresses and most subjects come to experiments with weak priors anyway. The specification of initial attractions is not specific to a learning rule. All learning rules that use attraction as latent variable to predict behavior must face this problem. We use a "cognitive hierarchy" (CH) theory of games [11] to provide the initial attraction values because it has been shown to predict behavior well in many one-shot games. In the CH theory, each player assumes that his strategy is the most sophisticated. The CH model has a hierarchy of categories: step 0 players randomize; and step k thinkers best respond, assuming that other players are distributed over step 0 through step $k-1$. Since an average of 1.5 steps fits data well from many games, we use that value to set initial attractions. [6]

Consequently, we are left with 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. [7]

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 senses 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.

[6] 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 assume players use a decision rule that best responds 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 [47] and Costa-Gomes et al. [16] explore richer reasoning-step models.

[7] 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.

The core of the $\phi_i(t)$ change-detector function is a "surprise index", which is the difference between other players' most recently chosen strategies and their chosen strategies in all previous periods. First, define a cumulative 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 $\frac{\sum_{\tau=1}^{t} I(s_{-i}^k, s_{-i}(\tau))}{t}$.[8] The immediate "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 the squared deviations between the cumulative history vector $h_i^k(t)$ and the immediate history vector $r_i^k(t)$; that is,

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

In other words, the surprise index captures the degree of change of the most recent observation[9] from the historical average. Note that it 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). When surprise index is zero, we have a stationary environment; when it is one, we have a turbulent environment. The change-detecting decay rate is

$$\phi_i(t) = 1 - \tfrac{1}{2} S_i(t). \tag{3}$$

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

Some numerical boundary cases help illuminate how the change detection works. 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). 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 0.75, 0.56, 0.36, and 0.19.) 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. Note that since the observed behavior in period t is included in the cumulative history $h_i^k(t)$, $\phi_i(t)$ will never dip completely to zero (which could be a mistake because it erases all the history embodied in the lagged attraction). For example, if a player chose the same strategy for each of 19 periods and a new strategy in period 20, then $\phi_i(t) = 39/400 = 0.0975$.

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 $\phi_i(t) = 0.5 + \frac{1}{2t}$, which starts at 0.75 and asymptotes at 0.5 as t increases. Comparing the case where the previous strategy was the same, and the previous strategies

[8] 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.

[9] 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 $r_i^k(t) = \sum_{k=1}^{m_{-i}} \left[\frac{\sum_{\tau=t-W+1}^{t} I(s_{-i}^k, s_{-i}(\tau))}{W} \right]$.

were all different, it is evident that if the choice in period t is new, the value of $\phi_i(t)$ is higher if there were more variation in previous choices, and lower if there were less variation. This mimicks a hypothesis-testing approach (cf. [33]) in which more variation in previous strategies implies that players are less likely to conclude there has been a regime shift, and therefore do not lower the value of $\phi_i(t)$ too much.

Note that the change-detector function $\phi_i(t)$ is individual and time specific and it depends on information feedback. Nyarko and Schotter [37] show that a weighted fictitious play model that uses stated beliefs (instead of empirical beliefs posited by the fictitious play rule) can predict behavior better than the original EWA model in games with unique mixed-strategy equilibrium. One way to intrepret their result is that their model allows each subject to attach a different weight to previous experiences over time. In the same vein, the proposed change-detector function allows for individual and time heterogeneity by positing them theoretically.

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. [24], who show that such a regret-driven rule converges to correlated equilibrium.) To capture this property, define [10]

$$\delta_{ij}(t) = \begin{cases} 1 & \text{if } \pi_i(s_i^j, s_{-i}(t)) \geqslant \pi_i(t), \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

That is, subjects reinforce chosen strategies and all unchosen strategies with (weakly) better payoffs by a weight of one. They reinforce unchosen strategies with strictly 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) toward 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 [43]. 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.

[10] In games with unique mixed-strategy equilibrium, we use $\delta_{ij}(t) = \frac{1}{W}$ if $\pi_i(s_i^j, s_{-i}(t)) \geqslant \pi_i(t)$ 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 [53]). 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 and the Mookerjhee–Sopher games).

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". 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, triggering higher $\delta_{ij}(t)$ values (and reinforcement of superior payoffs) and guiding 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 parametric EWA model shows 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) and spreads choice probability over a wider range of strategies due to the logit response rule. This implies that the strategy chosen may no longer be optimal, leading $\delta_{ij}(t)$ to allocate 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 re-allocating attention when environmental change demands such action.

A theorist's instinct is to derive conditions when flexible learning rules choose parameters optimally, which is certainly a direction to explore in future research (cf. [25,31]). 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 to find a universally optimal rule that can always beat rules which adapt locally.

Our approach is more like the 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 [48]). 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 the properties of the self-tuning model. First, the use of simple fictitious play and reinforcement theories in empirical analysis is often justified by the fact that they have only a few free parameters. The self-tuning EWA is useful by this criterion because it requires estimating only one parameter, λ (which is difficult to do without 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, when opponent choices vary 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 [34]. [11] The goals are to see whether self-tuning EWA can fit and predict as well as parametric EWA and whether the ϕ and δ functions can vary sensibly across 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 in brand new games.

We use seven games: two matrix games with unique mixed-strategy equilibrium [36]; R&D patent race games [38]; a median-action order statistic coordination game with several players [50]; a continental-divide coordination game, in which convergence behavior is extremely sensitive to initial conditions [52]; a coordination game in which players choose whether to enter a large or small market [1]; dominance-solvable p-beauty contests [27]; and a price matching game called traveler's dilemma [12].

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 the 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 [40,22]).

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

$$\prod_{t=1}^{T} P_i^{s_i(t)}(t|\lambda). \tag{5}$$

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

[12] The other four games are: mixed-equilibrium games studied by Mookherjee and Sopher [36] which have four or six strategies, one of which is weakly dominated; the nine-player median-action game studied by Van Huyck et al. [50], 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 [12] 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 [1] in which *n* players simultaneously enter a large or small market and earn 2*n* (*n*) divided by the number of entrants if they enter the large (small) market.

Table 1
The seven games used in the estimation of self-tuning EWA, parametric EWA, and QRE

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
A. Games with a unique mixed-strategy equilibrium								
Mixed strategies Mookerjhee and Sopher [36]	2	4,6	0	80	40	Fixed	Stake size	A constant-sum game with unique mixed strategy equilibrium
Patent race Rapoport and Amaldoss [38]	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
B. Coordination games								
Continental divide Van Huyck et al. [52]	7	14	2	70	15	Fixed	None	A coordination game with two pure strategy equilibria
Median action Van Huyck et al. [50]	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 et al. [1]	3,6,9,18	2	n[a]	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.
C. Dominance solvable games								
Price matching (Traveller's dilemma) Capra et al. [12]	2	121[b]	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. [27]	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

[a] The number of equilibria depends on the number of players. $n = 3$, 15, 84, 18 564, respectively.
[b] Continuous strategies of 80–200 are discretized to 121 integer strategies.

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

$$L(\lambda) = \prod_{i}^{N} \left\{ \prod_{t=1}^{T} P_i^{s_i(t)}(t|\lambda) \right\}. \tag{6}$$

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 [10,11] using $\tau = 1.5$.[13] After each period, the functions $\phi_i(t)$ and $\delta_{ij}(t)$ are updated according to player *i*'s experience and apply 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 a random sample of 70% of the subjects in each game, we determine the value of λ that maximizes the joint likelihood. Then the value of λ is frozen and used to forecast behavior of the entire path of the remaining 30% of the subjects.[14] We convert payoffs to inflation-adjusted dollars (which is important for cross-game forecasting) and scale them by subtracting the lowest possible payoffs. Randomized bootstrap resampling is used to calculate parameter standard errors.[15]

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 a 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 over-fitting 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 where k is the number of parameters. 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 shows the choice probability implied by the average likelihood[16] 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 0.49.[17]

---

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

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

[15] In the first few periods of a game, players often learn rapidly. To capture this, we smooth the $\phi_i(t)$ function by starting at 0.5, and gently blending in the updated values according to $\hat{\phi}_i(t) \equiv 0.5/t + (t-1)\phi_i(t)/t$ in the empirical implementation.

[16] 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.

[17] The parameter estimates are reported in Table 4. Standard errors of these estimates are reported in Ho et al. [26].

Table 2
Model fit and validation using in-game estimates

Data set	Mixed strategies	Patent race	Continental divide	Median action	Pot games	p-Beauty contest	Price matching	Pooled[a]
Total sample size	3200	5760	1050	540	2217	1960	520	15 247
In-sample calibration								
Sample size	2240	4000	735	380	1478	1380	360	10 573
BIC (Bayesian information criterion)[b]								
Self-tuning EWA	−3206	−4367	−1203	−309	−938	−5190	−1451	−16 672
EWA	−3040	−4399	−1091	−293	−931	−4987	−1103	−16 646
QRE	−3442	−6686	−1923	−549	−1003	−6254	−1420	−21 285
Average probability								
Self-tuning EWA (%)	23.9	33.6	19.5	44.6	53.1	2.3	1.8	
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	
Out-of-sample validation								
Sample size	960	1760	315	160	739	580	160	4674
Log-likelihood								
Self-tuning EWA	−1394	−1857	−522	−94	−441	−2350	−585	−7246
EWA	−1342	−1876	−482	−89	−433	−2203	−532	−7366
QRE	−1441	−3006	−829	−203	−486	−2667	−607	−9240

[a] A common set of parameters, except game-specific λ, is estimated for all games. Each game is given equal weight in LL estimation.
[b] BIC (Bayesian Information Criterion) is given by $LL - (k/2)\log(NT)$ where k is the number of parameters, N is the number of subjects and T is the number of periods.

However, self-tuning EWA is better than EWA in the pooled estimation where a common set 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 (2×2 games with mixed equilibria [19]; and 5×5 symmetric games [46]). 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 [40] 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

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	−4602	−7114	−1723	−419	−1399	−8282	−2192
EWA	−4733	−6321	−1839	−457	−1365	−7806	−2140
QRE	−4667	−9132	−2758	−957	−1533	−8911	−2413
Average probability							
Self-tuning EWA (%)	23.7	29.1	19.4	46.0	53.2	1.5	1.5
EWA (%)	22.8	33.4	17.4	42.9	54.0	1.9	1.6
QRE (%)	23.3	20.5	7.2	17.0	50.1	1.1	1.0
Random (%)	20.4	18.3	7.1	14.3	50.0	1.0	0.8

Table 4
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	0.11	0.08	0.29	0.15	0.10	0.29	0.36	0.32
λ	4.13	9.24	4.45	5.64	7.34	2.39	10.17	5.87
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

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. [18]

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 [38]. 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

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

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.

The game has an interesting strategic structure. The strong player can guarantee a payoff of five by investing the entire endowment $i_{\text{strong}} = 5$ (out-spending the weak player), which strictly dominates investing zero ($i_{\text{strong}} = 0$). Eliminating the strong player's dominated strategy $i_{\text{strong}} = 0$ then makes $i_{\text{weak}} = 1$ a dominated strategy for the weak player (since she can never win by investing one unit). Iterating in this way, the strong player deletes $i_{\text{strong}} \in \{0, 2, 4\}$ and the weak player deletes $i_{\text{weak}} \in \{1, 3\}$ 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 two groups where random matching occurs within group. Since aggregate 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. Figs. 2a–d show the empirical transition matrix and predicted transition frequencies across the five strategies ($i_{\text{strong}} \in \{0, 1, \ldots, 5\}$) 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 $i_{\text{strong}} = 5$, 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 $i_{\text{strong}} = 5$, which explains the persistence of investing and the low rate of switching. The low value of δ, which is estimated to be 0.30 in EWA and averages 0.08 in self-tuning EWA, is one way of expressing why strong players are sluggish in switching down from $i_{\text{strong}} = 5$.

3.4. Games with multiple pure strategy equilibria: continental divide game

Van Huyck et al. [52] 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 $M \leqslant 7$, the best response to a median M is strictly between M and 3. For high medians $M \geqslant 8$, 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.

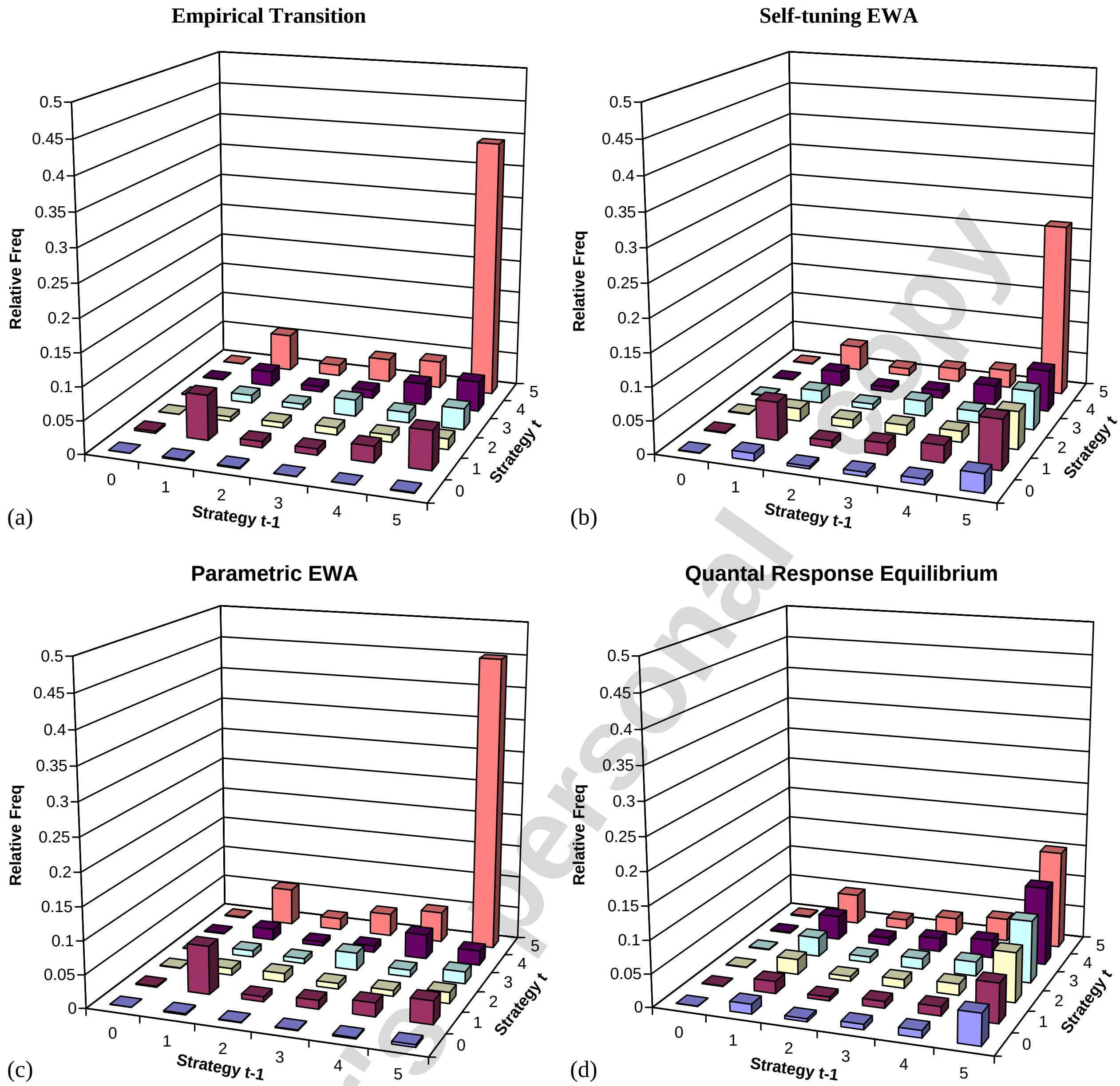

Fig. 2. Actual and predicted choice transition matrices in patent race.

Figs. 3a–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. (Fig. 3a hides strong path-dependence: groups which had first-period $M \leqslant 7$ ($M \geqslant 8$) *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

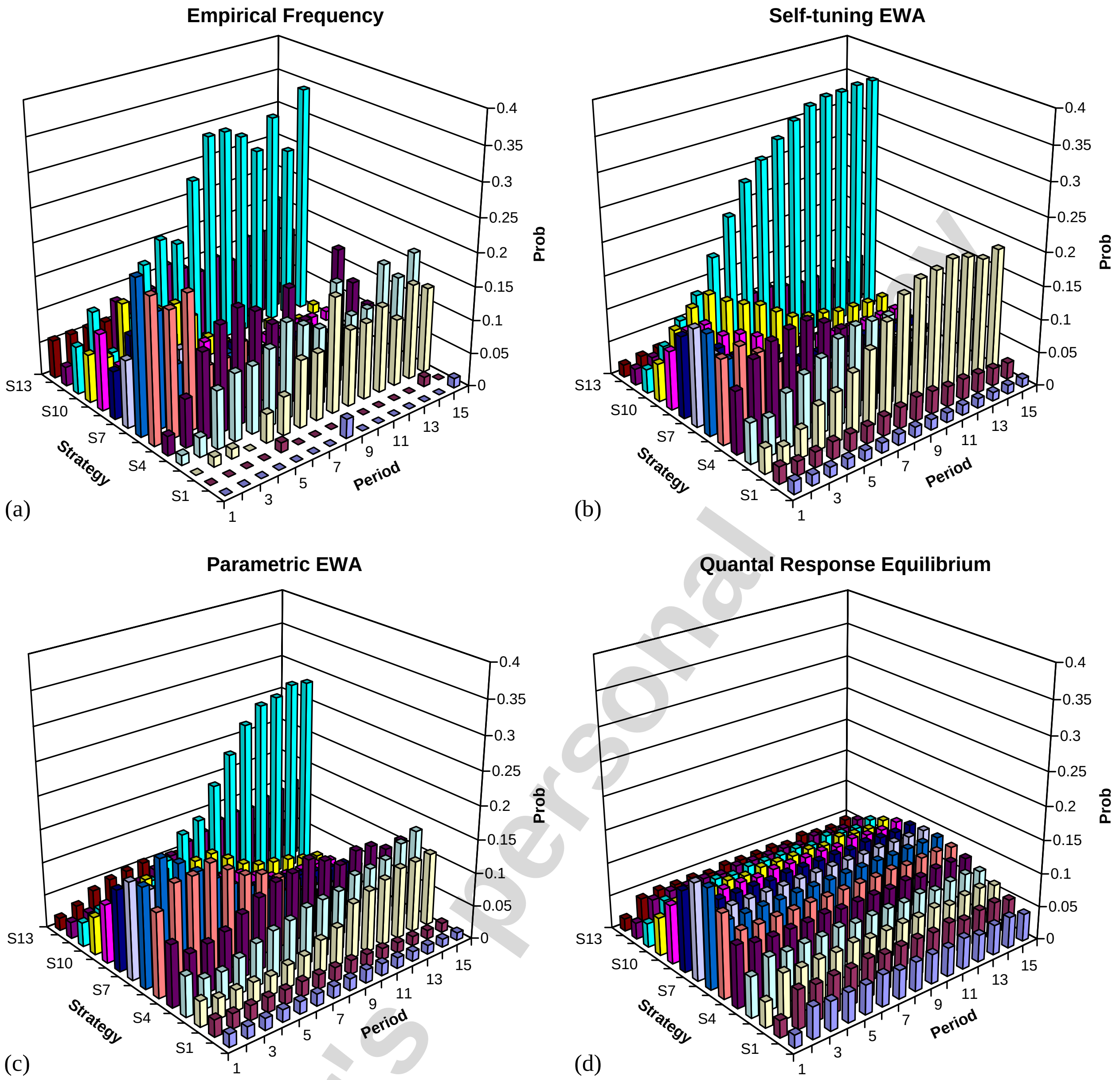

Fig. 3. Actual and predicted choice frequencies in continental divide.

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. [27], 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 toward equilibrium.

Figs. 4a–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.)

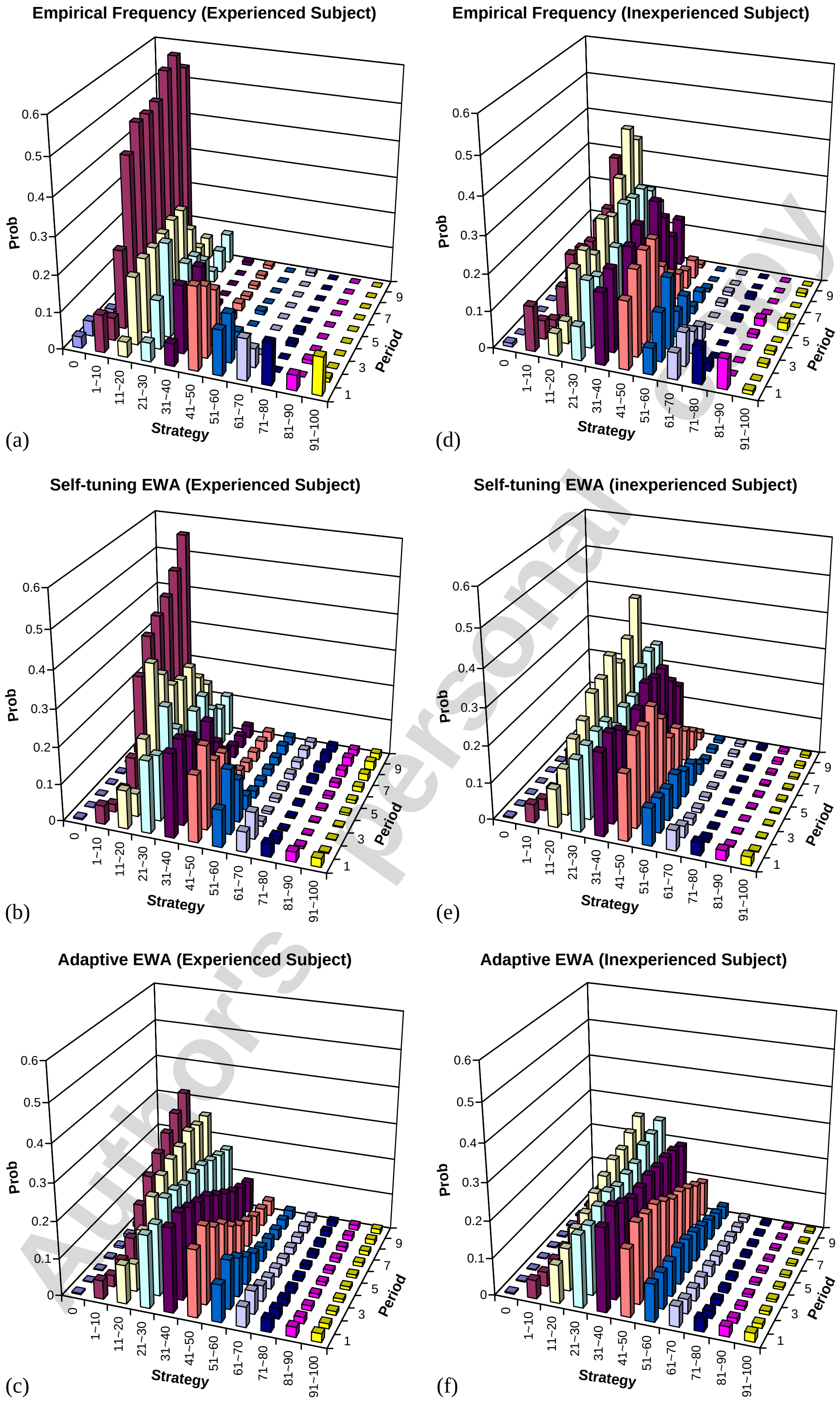

Fig. 4. Actual and predicted choice frequencies in p-beauty contest played by experienced and inexperienced subjects.

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 more responsive to surprises than the 0.83 cross-game estimate of EWA. The EWA's sluggish response to surprises explains why it is less accurate in experienced sessions because experienced subjects generate more surprises than their inexperienced counterparts by choosing smaller numbers. The $\delta_{ij}(t)$ function, which by definition is totally responsive to better strategies, gives more weights to smaller numbers too. In combination, the two functions help to explain the faster convergence of experienced subjects toward the equilibrium of zero.

4. 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 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 more complex models do not necessarily fit and predict better.

There are two key results.

First, self-tuning EWA fits and predicts slightly worse than parametric EWA in all seven games. Since self-tuning EWA generates functional parameter values for ϕ and δ, which vary sensibly across games, it predicts better than other QRE and parametric EWA when games are pooled and common parameters are estimated. Self-tuning EWA therefore represents one solution to the problem of flexibly generating EWA-like parameters across games.

Second, the functions in self-tuning EWA seem to be robust across games. Our working paper added three brand new games which are not reported here (after the first version was written and circulated, and people were invited to submit games for analysis) 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 [45]) and explain "teaching" behavior in repeated games [9,15].

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 models that use estimated parameters from history will do poorly. Since many naturally occurring economic games do have unanticipated shocks which change payoffs, exploring how people and models perform in games like this is worthwhile.

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, participants in market institutions like eBay auctions learn to bid late to hide their information about an object's common value [4]. Consumers learn over time what products they like [28]. Learning in financial markets can generate excess volatility and returns predictability, which are otherwise anomalous in rational expectations models [49]. Sargent [41] 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] W. Amaldoss, J.-K. Chong, T.-H. Ho, Choosing the right pond: EWA learning in games with different group sizes, working paper, Department of Marketing, University of California, Berkeley, 2005.

[2] J. Arifovic, J. Ledyard, Computer testbeds: the dynamics of groves-ledyard mechanisms, unpublished manuscript, 2002.

[3] B. Arthur, Designing economic agents that act like human agents: a behavioral approach to bounded rationality, Amer. Econ. Rev. 81 (2) (1991) 353–359.

[4] P. Bajari, A. Hortacsu, Winner's curse, reserve prices and endogenous entry: empirical insights from ebay auctions, Rand J. Econ. 34 (2) (2003) 329–355.

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

[6] A. Cabrales, R. Nagel, R. Armenter, Equilibrium selection through incomplete information in coordination games: an experimental study, working paper, Universitat Pompeu Fabra, 2002.

[7] C.F. Camerer, Behavioral Game Theory: Experiments on Strategic Interaction, first ed., Princeton University Press, Princeton, NJ, 2003.

[8] C.F. Camerer, T.-H. Ho, Experience-weighted attraction learning in normal-form games, Econometrica 67 (1999) 827–874.

[9] C.F. Camerer, T.-H. Ho, J.-K. Chong, Sophisticated EWA learning and strategic teaching in repeated games, J. Econ. Theory 104 (1) (2002) 137–188.

[10] C.F. Camerer, T.-H. Ho, J.-K. Chong, Models of thinking learning and teaching games, Amer. Econ. Rev. 93 (2) (2003) 192–195.

[11] C.F. Camerer, T.-H. Ho, J.-K. Chong, A cognitive hierarchy model of thinking in games, Quart. J. Econ. 119 (3) (2004) 861–898.

[12] M. Capra, J. Goeree, R. Gomez, C. Holt, Anomalous behavior in a traveler's dilemma, Amer. Econ. Rev. 89 (3) (1999) 678–690.

[13] Y.-W. Cheung, D. Friedman, Individual learning in normal form games: some laboratory results, Games Econ. Behav. 19 (1997) 46–76.

[14] J.-K. Chong, C.F. Camerer, T.-H. Ho, A learning-based model of repeated games with incomplete information, Games Econ. Behav. 2006, in press.

[15] D. Cooper, J. Kagel, Learning and transfer in signaling games, working paper, Department of Economics, Case Western Reserve University, 2004.

[16] M. Costa-Gomes, V. Crawford, B. Broseta, Cognition and behavior in normal-form games: an experimental study, Econometrica 69 (2001) 1193–1235.

[17] J. Cross, A Theory of Adaptive Learning Economic Behavior, Cambridge University Press, New York, 1983.

[18] I. Erev, Y. Bereby-Meyer, A.E. Roth, The effect of adding a constant to all payoffs: experimental investigation and a reinforcement learning model with self-adjusting speed of learning, J. Econ. Behav. Organ. 39 (1999) 111–128.

[19] I. Erev, A.E. Roth, Predicting how people play games: reinforcement learning in experimental games with unique mixed-strategy equilibria, Amer. Econ. Rev. 88 (1998) 848–881.

[20] G. Frechette, Learning in a multilateral bargaining experiment, working paper, Harvard University, 2003.

[21] D. Fudenberg, D. Levine, The Theory of Learning in Games, first ed., MIT Press, Cambridge, MA, 1998.

[22] J. Goeree, C. Holt, Ten little treasures of game theory and ten intuitive contradictions, Amer. Econ. Rev. 91 (6) (2001) 1402–1422.

[23] C. Harley, Learning the evolutionary stable strategies, J. Theor. Biol. 89 (1981) 611–633.

[24] S. Hart, A. Mas-Colell, A general class of adaptive strategies, J. Econ. Theory 98 (2001) 26–54.

[25] D. Heller, R. Sarin, Parametric adaptive learning, working paper, University of Chicago, 2000.

[26] T.-H. Ho, C.F. Camerer, J.-K. Chong, Functional EWA: a one-parameter theory of learning in games, working paper, California Institute of Technology, Division of the Humanities and Social Sciences Paper #1122, 2001.

[27] T.-H. Ho, C.F. Camerer, K. Weigelt, Iterated dominance and iterated best-response in p-beauty contests, Amer. Econ. Rev. 88 (1998) 947–969.

[28] T.-H. Ho, J.-K. Chong, A parsimonious model of sku choice, J. Marketing Res. 40 (3) (2003) 351–365.

[29] T.-H. Ho, X. Wang, C.F. Camerer, Individual differences in the EWA learning with partial payoff information, working paper, Department of Marketing, University of California, Berkeley, CA, 2004.

[30] E. Hopkins, Two competing models of how people learn in games, Econometrica 63 (2002) 759–778.

[31] J. Josephson, A numerical analysis of the evolutionary stability of learning rules, working paper, Stockholm School of Economics SSE/EFI paper no. 474, ⟨http://swopec.hhs.se/hastef/abs/hastef0474.htm⟩, 2001.

[32] M. Kocher, M. Sutter, The decision maker matters: individual versus group behavior in experimental beauty-contest games, Econ. J. 115 (2005) 200–223.

[33] A. Marcet, J.P. Nicolini, Recurrent hyperinflations and learning, Amer. Econ. Rev. 93 (5) (2003) 1476.

[34] R.D. McKelvey, T.R. Palfrey, Quantal response equilibria for normal-form games, Games Econ. Behav. 10 (1995) 6–38.

[35] D. Mookherjee, B. Sopher, Learning behavior in an experimental matching pennies game, Games Econ. Behav. 7 (1994) 62–91.

[36] D. Mookherjee, B. Sopher, Learning and decision costs in experimental constant-sum games, Games Econ. Behav. 19 (1997) 97–132.

[37] Y. Nyarko, A. Schotter, An experimental study of belief learning using elicited beliefs, Econometrica 70 (2002) 971–1005.

[38] A. Rapoport, W. Amaldoss, Mixed strategies and iterative elimination of strongly dominated strategies: an experimental investigation of states of knowledge, J. Econ. Behav. Organ. 42 (2000) 483–521.

[39] A. Rapoport, I. Erev, Coordination, “magic”, and reinforcement learning in a market entry game, Games Econ. Behav. 23 (1998) 146–175.

[40] A. Roth, I. Erev, Learning in extensive-form games: experimental data and simple dynamic models in the intermediate term, Games Econ. Behav. 8 (1995) 164–212.

[41] T. Sargent, The Conquest of American Inflation, first ed., Princeton University Press, Princeton, NJ, 1999.

[42] R. Sarin, F. Vahid, Predicting how people play games: a simple dynamic model of choice, Games Econ. Behav. 34 (2001) 104–122.

[43] R. Selten, R. Stoecker, End behavior in sequences of finite prisoner’s dilemma supergames: a learning theory approach, J. Econ. Behav. Organ. 7 (1986) 47–70.

[44] Y. Sovik, Impossible bets: an experimental study, working paper, University of Oslo, 2001.

[45] D.O. Stahl, Sophisticated learning and learning sophistication, working paper, University of Texas at Austin, 1999.

[46] D.O. Stahl, Rule learning in symmetric normal-form games: theory and evidence, Games Econ. Behav. 32 (1) (2000) 105–138.

[47] D.O. Stahl, P. Wilson, On players models of other players: theory and experimental evidence, Games Econ. Behav. 10 (1995) 213–254.
[48] R. Sutton, A. Barto, Reinforcement Learning: An Introduction, first ed., MIT Press, Boston, 1998.
[49] A.G. Timmerman, How learning in financial markets generates excess volatility and predictability in stock prices, Quart. J. Econ. 108 (1993) 1135–1145.
[50] J. Van Huyck, R. Battalio, R. Beil, Tacit cooperation games, strategic uncertainty, and coordination failure, Amer. Econ. Rev. 80 (1990) 234–248.
[51] J. Van Huyck, R. Battalio, F. Rankin, Selection dynamics and adaptive behavior without much information, working paper, Department of Economics, Texas A&M University, 2005.
[52] J. Van Huyck, J. Cook, R. Battalio, Adaptive behavior and coordination failure, J. Econ. Behav. Organ. 32 (1990) 483–503.
[53] N.T. Wilcox, Heterogeneity and learning principles, working paper, University of Houston, 2003.