Quant Mark Econ (2012) 10:475–503
DOI 10.1007/s11129-012-9125-8

Dynamic learning in behavioral games: A hidden Markov mixture of experts approach

Asim Ansari · Ricardo Montoya · Oded Netzer

Received: 30 July 2011 / Accepted: 12 July 2012 / Published online: 5 August 2012

Abstract Over the course of a repeated game, players often exhibit learning in selecting their best response. Research in economics and marketing has identified two key types of learning rules: belief and reinforcement. It has been shown that players use either one of these learning rules or a combination of them, as in the Experience-Weighted Attraction (EWA) model. Accounting for such learning may help in understanding and predicting the outcomes of games. In this research, we demonstrate that players not only employ learning rules to determine what actions to choose based on past choices and outcomes, but also change their learning rules over the course of the game. We investigate the degree of state dependence in learning and uncover the latent learning rules and learning paths used by the players. We build a non-homogeneous hidden Markov mixture of experts model which captures shifts between different learning rules over the course of a repeated game. The transition between the learning rule states can be affected by the players' experiences in the previous round of the game. We empirically validate our model using data from six games that have been previously used in the literature. We demonstrate that one can obtain a richer understanding of how different learning rules impact the observed strategy choices of players by accounting for the latent dynamics in the learning rules. In addition, we show that such an approach can improve our ability to predict observed choices in games.

A. Ansari (✉) · O. Netzer
Columbia Business School, Columbia University, New York, NY USA
e-mail: maa48@columbia.edu

O. Netzer
e-mail: on2110@columbia.edu

R. Montoya
Industrial Engineering Department, University of Chile, Santiago, Chile
e-mail: rmontoya@dii.uchile.cl

Keywords Learning · Behavioral game theory · Experimental economics · Hidden Markov mixture of experts model · Bayesian estimation

JEL Classification c5 · c7 · c11 · D83

1 Introduction

Research in economics and marketing often relies on the notion that agents operate in equilibrium. However, it is unlikely that such an equilibrium is reached instantaneously in a single period game. It may be more intuitive to think of equilibrium as a limiting state achieved after an evolutionary process that unfolds over time (Camerer et al. 2003). The question of how equilibrium arises has been the focus of attention in the rich literature dealing with learning in behavioral games (e.g., Camerer 2003; Fudenberg and Levine 1998). This literature suggests that over the course of repeated competitive interactions, agents exhibit dynamic behavior by learning from their own actions and outcomes and from those of their competitors (Kunreuther et al. 2009; Roth and Erev 1995). Several types of learning rules have been explored, including belief, reinforcement, and imitation learning. In belief learning, players form their beliefs based on the opponent's prior decisions. In reinforcement learning, strategies that paid off in the past get reinforced. Imitation learning implies that players may learn by imitating the action of others. Previous research has demonstrated via experiments that accounting for such learning rules helps in understanding and predicting the outcomes of games. Specifically, it has been shown that players' choice of game strategies is consistent with the above learning rules or with generalizations of these, such as the Experience-Weighted Attraction (EWA) learning model (Camerer and Ho 1999).

In this research we argue and demonstrate that in repeated strategic decision making, an additional source of dynamics may arise from shifts in the learning rules that are used by a player. That is, over the course of a repeated game, players can not only learn what actions to choose based on past choices and outcomes (e.g., Kunreuther et al. 2009), but can also change their learning rules over time (Stahl 2003). For example, a player may shift from exploration to exploitation behavior depending upon the outcome feedback in the game. We extend previous studies that have examined the existence of multiple types of learning by allowing players to switch over time between different latent learning rules. The use of particular learning rules and the dynamics in their deployment over a repeated game may depend upon the type of game, the outcomes of the game in the previous rounds, and may also vary across players. Accordingly, we address the following questions: (i) Do players dynamically shift between learning rules over repeated plays of a game? (ii) Can past outcomes trigger such transitions between learning rules? (iii) Do different games exhibit different learning dynamics? and (iv) Is there heterogeneity in the learning dynamics across players?

The learning rule used by a player on any given round is not observable as experimental data reveals only the strategies chosen by the players and the resulting outcomes. We therefore build a non-homogeneous hidden Markov mixture of experts (NH-HMME) model to capture the latent transitions among different learning rules (belief and reinforcement learning) over the course of the game. Such transitions between different learning rules could be induced by the outcome of the game over time. Conditional on a state, players make strategy choices that are probabilistically consistent with their current learning rule.

A proper accounting of such rule dynamics is not possible unless we allow for unobserved sources of heterogeneity that govern the different choices (i.e., learning rules and game strategies) made by players (Wilcox 2006). We use a Bayesian approach (Kunreuther et al. 2009) to specify a hierarchical NH-HMME model that captures heterogeneity in the parameters that underly each learning rule as well as in the parameters that govern the rule transition process. Our approach also allows us to probabilistically uncover the deployment of these rules and the specific learning paths used by different players over the course of a game.

We empirically validate our model using data from six repeated games: Continental divide coordination game (Van Huyck et al. 1997), the median-action order statistic coordination game with several players (Van Huyck et al. 1990), mixed strategies (Mookherjee and Sopher 1997), R&D patent race (Rapoport and Amaldoss 2000), pot games (Amaldoss et al. 2005), and *p*-Beauty contest (Ho et al. 1998). We demonstrate varying degrees of learning rule dynamics across games and players and show that our modeling approach provides an enriched understanding of how learning unfolds over the course of the game. We find that past outcomes, in particular whether the player left money on the table in the previous round of the game, have a strong and heterogeneous effect on the learning rules employed by players. In some cases, these outcomes may trigger a shift in the learning rule, whereas, in other situations, they may reinforce the learning rule that is in use. We also find that these additional insights about the learning process come at no cost in terms of predictive ability when compared to several benchmark models that are common in the literature.

The rest of the paper is organized as follows. Section 2 provides an overview of the existing learning models. Section 3 describes our NH-HMME model that represents the dynamics in learning rules. Section 4 presents the results of the proposed model and compares it to several benchmark models using data from the six behavioral games. Finally, Section 5 concludes the paper and discusses limitations and future research directions.

2 Learning models

The literature on behavioral games has proposed multiple approaches by which agents can learn from experience. Some of the learning rules that have been studied include: reinforcement learning (Erev and Roth 1998; Roth and Erev 1995), belief learning (Camerer and Ho 1998, 1999), fictitious play (Brown 1951), imitation (Hück et al. 1999), and anticipatory (sophistication) learning (Selten 1991). Hybrid models that combine multiple learning rules such as the EWA model (Camerer and Ho 1999; Camerer et al. 2002a) have generally demonstrated improved predictive ability over models that utilize a single learning rule. In this research, we focus on the two elementary learning rules most prominently studied in the literature: belief learning and reinforcement learning. We investigate how agents may switch between these two learning rules over the course of a repeated game and compare our proposed approach to popular alternatives including the EWA model, which also combines and generalizes both these learning rules. We briefly review these learning rules next.

2.1 Game setting

We follow the standard notation to specify the game setting (see e.g., Camerer and Ho 1999). Players are indexed by i ($i = 1, 2, \ldots, n$). Player i's strategy space consists of discrete choices indexed by j ($j = 1, 2, \ldots, J$). The game is repeated for several rounds indexed by t ($t = 1, 2, \ldots, T$). At each round, all players choose their strategies or actions. Let s_i^j denote strategy j for player i and $s_i(t)$ be the strategy actually chosen by player i in round t. After playing her own strategy, the player observes the strategies (choices) made by the other players, $s_{-i}(t)$. The vector of selected strategies $(s_i(t), s_{-i}(t))$ determines the payoff received by the player $\pi_i(s_i(t), s_{-i}(t))$. It is commonly assumed that each possible strategy j has an intrinsic value to the player, given by its "attraction". The attraction of strategy j for player i before the period t play is given by $A_{ij}(t-1)$. The players start the game with initial values for the attractions $A_{ij}(0)$, which are then updated at each round, based on the outcome of the round. The different learning rules vary with respect to how these attractions are updated over the course of the game.

2.2 Reinforcement learning

Reinforcement learning (Erev and Roth 1998; Roth and Erev 1995) relies on the notion that in choosing their game strategies, agents pay attention only to the history of their own payoffs. Specifically, these models suggest that chosen strategies are reinforced cumulatively by the received payoff. The reinforcement model is strongly rooted in the psychology literature dating back to "the law of effect" of Thorndike (1898), which suggests that actions that have led to positive outcomes in the past are likely to be repeated. In the reinforcement learning model, the attraction of strategy j for individual i *after* playing period t, $A_{ij}^{R}(t)$, is obtained via the following updating rule:[1]

$$A_{ij}^{R}(t) = \begin{cases} \phi_i A_{ij}^{R}(t-1) + \pi_i(s_i^j, s_{-i}(t)), & \text{If } s_i^j = s_i(t) \\ \phi_i A_{ij}^{R}(t-1), & \text{If } s_i^j \neq s_i(t) \end{cases}, \tag{1}$$

where, the parameter ϕ_i is the forgetting or recency parameter that depreciates player i's previous attraction. The second term in the first row in Eq. 1 captures the notion that chosen actions with higher payoffs are reinforced more and the second equation indicates that the unchosen actions are not reinforced at all.

Even though the reinforcement learning rule is considered as a "low rationality" rule because the player does not account for decisions made by other players, or as a "nearly pure-inertia" model (Erev and Haruvy 2005; Haruvy and Erev 2002), due to its lack of positive reinforcement of non-played strategies, it has been shown to provide robust predictions in mixed strategy equilibria games (Erev and Roth 1998; Roth and Erev 1995). While reinforcement learning stems from basic psychological principles and has been shown to fit well the data from several games, it forecloses the opportunity to learn from sources other than direct reinforcement of past actions. It has also been demonstrated that providing respondents with information about the other player's actions expedites convergence to equilibrium (Mookherjee and Sopher 1994; Van Huyck et al. 2007). This suggests that learning mechanisms that are based on competitive information can also underly observed strategy choices.

2.3 Belief learning

Belief learning models (Brown 1951; Fudenberg and Levine 1998) postulate that players keep track of past actions of other players and form beliefs based on the action history of opponents to determine the best response based on the expected payoffs. Belief models vary in terms of how far into the history the player looks in order to form her beliefs. For example, the Cournot model (Cournot 1960) assumes a single period history, whereas the fictitious play (Fudenberg and Levine 1998) assumes that all past actions are counted equally in forming beliefs. We follow the derivation in Camerer and Ho (1999) in defining a general belief model that nests several of the known belief models including Cournot and weighted fictitious play.

[1] The model could be extended to capture more elaborate reinforcement behaviors (e.g., Erev and Roth 1998; Roth and Erev 1995).

According to the belief learning rule, the attraction of strategy j for individual i can be updated by

$$A_{ij}^{B}(t) = \frac{\phi_i A_{ij}^{B}(t-1) N_i(t-1) + \pi_i(s_i^j, s_{-i}(t))}{\phi_i N_i(t-1) + 1}, \tag{2}$$

where, ϕ_i is a decay parameter for past attractions, π_i is the payoff function, and $N_i(t)$ represents the game experience, updated by $N_i(t) = \phi_i N_i(t-1) + 1$. The pre-game experience, $N_i(0)$, is a parameter that is estimated from data. For $\phi_i = 0$, Eq. 2 reduces to the Cournot model and for $\phi_i = 1$, Eq. 2 characterizes the fictitious play model with equal weight to each period in the history. Thus, while in the reinforcement learning model the player learns only from the history and consequences of her own actions, in the belief learning model the history of play for all players and the forgone payoffs are taken into consideration in learning. Previous research has compared these two types of learning models in terms of their ability to predict game outcomes. The evidence in this regard, however, is mixed (Camerer 2003, p. 304).

2.4 Experience-weighted attraction

The experience-weighted attraction (EWA) model (Camerer and Ho 1999) is one of the leading models of individual learning. It combines the most appealing aspects of reinforcement and belief models. In the EWA model, a player's learning occurs from considering both forgone payoffs of the unchosen strategies and the reinforcement of the chosen strategies. Accordingly, the EWA model nests the reinforcement and several belief learning models (e.g., Cournot and weighted fictitious play). While the EWA model combines the belief and reinforcement rules, it is not simply a convex combination of these rules. The EWA model provides a relatively parsimonious representation for several learning models. At the same time, its parameters make psychological sense and facilitate an understanding of the learning mechanism employed by players. A host of studies using 31 different datasets that span diverse game types (Camerer and Ho 1998, 1999; Camerer et al. 2002a) have demonstrated the superior fit and predictive ability of the EWA model relative to the reinforcement, belief and other learning models.

According to the EWA model, the attraction of strategy j for individual i can be updated by[2]

$$A_{ij}^{EWA}(t) = \frac{\phi_i A_{ij}^{EWA}(t-1) N_i(t-1) + [\delta_i + (1-\delta_i) I(s_i^j, s_i(t))]\pi_i(s_i^j, s_{-i}(t))}{\rho_i N_i(t-1) + 1}, \tag{3}$$

where, ϕ_i is a discount or decay factor of past attractions, δ_i measures the weight given to forgone payoffs relative to actual payoffs, ρ_i captures the

[2]Note that in some EWA papers ρ_i is replaced by $\phi_i(1-\kappa_i)$. We decided to keep the original notation used in Camerer and Ho (1999). The transformation of the results is straightforward.

decay of past experiences, and $I(s_i^j, s_i(t))$ is an indicator function equals 1, if $s_i(t)=s_i^j$, and 0, otherwise. $N_i(t)$ is updated according to $N_i(t) = \rho_i N_i(t-1) + 1$ and $N_i(0)$, the initial experience, is a parameter that is estimated. The EWA model reduces to the reinforcement and belief models under different restrictions on its parameters. When $\delta_i = 0$, $\rho_i = 0$ and $N_i(0) = 1$, it reduces to the reinforcement learning model shown in Eq. 1, and it is equivalent to the belief learning model, shown in Eq. 2, when $\delta_i = 1$ and $\phi_i = \rho_i$.

Several variants of the original EWA model (Camerer and Ho 1999) have been considered in the literature. For example, Camerer et al. (2002a) extend the EWA to allow for sophisticated learning behavior and strategic teaching. Ho et al. (2008) adapt the EWA to partial payoff information games. Camerer and Ho (1998) and Camerer et al. (2007) allow the parameters of the EWA model to vary over time. Camerer and Ho (1998) allow the parameters to monotonically increase or decrease over the course of the game, whereas, in Camerer et al. (2007) the EWA parameters vary over time as a function of player experience.

2.5 Rule learning

While the above learning rules focus on the updating of attraction functions, they assume that the learning rule itself is static. An alternative approach focuses on the dynamics in the usage of learning rules (*rule learning* approach Stahl 1999, 2000, 2001). The idea behind rule learning is that players can switch among learning rules over the course of a game. For instance, rules that have been used in the past and/or performed well in the past are more likely to be repeated by players. Salmon (2004) shows evidence for such rule switching, which he calls learning to learn behavior, by using experiments that varied the amount of information that is provided to players about the outcomes of the games and via interviews during the game. Because of computational difficulties in estimating the most general form of rule learning, applications involving rule learning have commonly studied population learning and have typically allowed only a limited number of rule transitions over the course of the game. Stahl (1999, 2000, 2001), using several experimental games, shows strong empirical evidence that players switch between multiple learning rules based on the previous outcomes of the game.

In this paper, we propose a hierarchical non-homogeneous Hidden Markov Mixture of Experts (NH-HMME) model for such rule learning. This approach allows players to transition between reinforcement and belief learning sub-models over the course of the game. The mixture of experts approach is popular in machine learning (Jordan and Jacobs 1994; Peng et al. 1996), where the sub-models are conditional distributions that are called experts. In our context, the sub-models, or experts, correspond to the two learning rules. The more familiar hidden Markov model can be thought of as a special case of the NH-HMME model, where the conditional distributions, or experts, come from the same family of distributions, but with different parameters.

The proposed NH-HMME model can be considered as a specific instantiation of the general approach for rule learning proposed by Stahl (2001, Eq. 2), providing a behaviorally sound structure to the dynamics of rule learning. In the NH-HMME model, players can change their learning rule after every round in the game instead of employing a limited number of possible switches in the learning rule (e.g., Stahl 2000, 2001; see Haruvy and Stahl 2012 for an exception). Additionally, the non-homogenoeus nature of the HMME captures the effect of past outcomes on the tendency to stay in or transition to other learning rules. The Markovian nature of the NH-HMME model allows us to capture the "law of effect" via the "stickiness" of the learning rule states. We now describe the proposed model in more detail.

3 Non-homogeneous hidden Markov mixture of experts model of learning rules

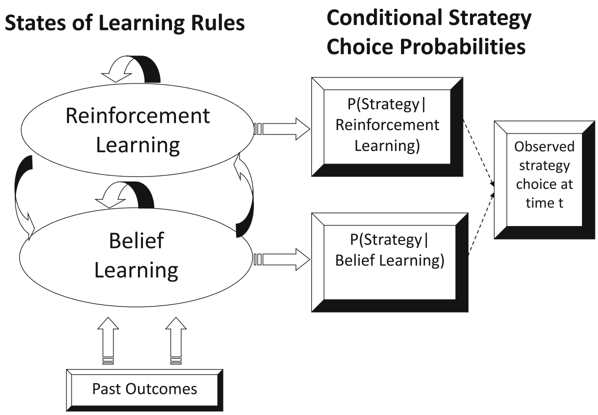
A HMME model involves a latent Markov process with unobserved states that represent the different learning rules that can be used by a player. We assume two states. One state represents the use of a reinforcement learning rule and the other the use of a belief learning rule. Players stochastically transition between these two states according to a first-order Markov process. We allow the transition probabilities between states to vary across players to reflect the heterogeneity in the intrinsic propensities of players to switch between learning rules. Additionally, to capture the effect of past outcomes on such transitions, we allow for non-homogeneous (time-varying) transition probabilities (Montoya et al. 2010; Netzer et al. 2008). The resulting model is a NH-HMME, which is a flexible model that can capture the game outcomes that may trigger a transition between the learning rules. For example, positive outcomes in previous rounds may encourage players to keep using the same learning rule, whereas disappointing outcomes may motivate changes in the learning rule. We schematically illustrate the proposed NH-HMME model in Fig. 1.

Let $s_i(t)$ be the game strategy (action) that is chosen by player i in round t. In the NH-HMME model, the joint probability of a sequence of decisions, $\{S_i(1) = s_i(1), \ldots, S_i(t) = s_i(t)\}$, up to time t, is a function of three main components: (1) the initial player-specific hidden states membership probabilities, given by a vector $\boldsymbol{\omega}_i$, (2) a set of player-specific time-varying transition probabilities between the different learning-rules (states), given by a transition matrix $\mathbf{Q}_{it}$, and (3) individual-level conditional probability models for choosing each strategy conditioned on the state (learning rule) represented by a diagonal matrix, $\mathbf{M}_{it}$. We describe our formulation of each of these components next.

3.1 Initial state membership probabilities

Let $z \in \{R, B\}$ denote a latent learning rule state ($z = R$, for Reinforcement and $z = B$, for Belief). Let ω_{iz} be the probability that player i is initially (i.e.,

Fig. 1 Non-homogeneous hidden Markov mixture of experts model of learning rules



in the first round of the calibration period) in state z, where $\omega_{iz} \geq 0$ and $\sum_{z\in\{R,B\}} \omega_{iz} = 1$. These initial probabilities may reflect previous game playing and rule usage experience by players and are inferred from the data. We collect these initial probabilities into a vector $\boldsymbol{\omega}_i$, such that

$$\boldsymbol{\omega}_i^{'} = [\omega_{iR}, \omega_{iB}] = [\omega_{iR}, 1 - \omega_{iR}]. \tag{4}$$

We reparameterize ω_{iR} in Eq. 4 using the inverse-logit transformation,

$$\omega_{iR} = \frac{\exp(\widetilde{\omega}_{iR})}{1 + \exp(\widetilde{\omega}_{iR})},$$

where, $\widetilde{\omega}_{iR}$ is an unconstrained parameter that represents the likelihood of beginning the repeated game with the reinforcement rule.

3.2 The Markov chain transition matrix

The transition matrix $\mathbf{Q}_{it}$ describes the law that governs player i's subsequent transitions between the states, i.e.,

$$\mathbf{Q}_{it} = \begin{bmatrix} q_{itRR} & q_{itRB} = 1 - q_{itRR} \\ q_{itBR} = 1 - q_{itBB} & q_{itBB} \end{bmatrix}. \tag{5}$$

We model $\mathbf{Q}_{it}$ as a function of the player's experience in previous rounds. Let $Z_{it} \in \{R, B\}$ be a random variable that denotes player i's state membership at time t and $\mathbf{x}_{it-1}$ be the vector of the covariates that represent the outcomes of the game that is observed by player i, prior to round t. Element $q_{itz'z}$ of the transition matrix denotes the probability that player i switches from learning-rule state z' in round $t-1$, to state z in round t, i.e.,

$$q_{itz'z} = P(Z_{it} = z | Z_{it-1} = z', \mathbf{x}_{it-1}), \tag{6}$$

where, $q_{itz'z} \geq 0$, $\sum_{z\in\{R,B\}} q_{itz'z} = 1$.

A player's propensity to transition from one state to another is a function of the observed outcomes in the previous rounds ($\mathbf{x}_{it-1}$) and intrinsic tendencies that are captured by random-effect intercepts. We can reparametrize each element in the transition matrix in Eq. 6 using the inverse-logit transformation to obtain:

$$q_{itRR} = \frac{\exp(\tau_{iR} + \boldsymbol{\rho}_{iR}^{'}\mathbf{x}_{it-1})}{1 + \exp(\tau_{iR} + \boldsymbol{\rho}_{iR}^{'}\mathbf{x}_{it-1})},$$

$$q_{itBB} = \frac{\exp(\tau_{iB} + \boldsymbol{\rho}_{iB}^{'}\mathbf{x}_{it-1})}{1 + \exp(\tau_{iB} + \boldsymbol{\rho}_{iB}^{'}\mathbf{x}_{it-1})} \qquad (7)$$

where, τ_{iz} represents the tendency for player i to remain in state z from one period to another, and $\boldsymbol{\rho}_{iz}$ is a vector of regression weights intended to capture the effect of previous outcomes on the propensity of player i to keep using the learning rule in state z. In our application, we explore alterative functions of the game outcomes that could affect the transition between learning rules.

3.3 Conditional choice model

Conditional on being in a particular state z, at round $t+1$, the probability of choosing a game strategy is specified using a logit model.[3] Specifically, we assume that player i chooses strategy $S_i(t+1)$, with a probability $P_{izt+1}(S_i(t+1))$ that depends on the learning rule being used. Specifically, we can write

$$P_{izt+1}(S_i(t+1) = j | Z_{it+1} = z) = \frac{\exp(\lambda_{iz} A_{ij}^{z}(t))}{\sum_k \exp(\lambda_{iz} A_{ik}^{z}(t))} \qquad (8)$$

where, $A_{ij}^{z}(t)$ is individual i's attraction for strategy j before playing round $t+1$ as given by the learning rule that state z represents. For state 1, which is the reinforcement learning state, the attraction *after* round t, $A_{ij}^{R}(t)$, is provided in Eq. 1. For state 2, which is the belief learning state, $A_{ij}^{B}(t)$ is provided in Eq. 2. The state-specific constant λ_{iz} scales these attractions.

To determine the probabilities of strategy choices in the first round, P_{iz1}, we follow Ho et al. (2002, Appendix 7) and define the vector of initial attractions $\mathbf{A}_i^{z}(0)$ based on the actual population frequency of choices in the first round of the game (given the estimated λ_i). To keep the models comparable, we do not use the first period to calibrate the other model's parameters and to compute the fit measures.[4] Unlike Ho et al. (2002), we allow the attractions in the first period to vary across players because of heterogeneity in λ_i.

[3] Alternative functional forms for the choice probabilities have been proposed in the literature (see e.g., Camerer and Ho 1998, 1999; Erev and Roth 1998). However, the logit formulation has consistently showed equal or better fit and prediction ability.

[4] Accordingly, the first period in the likelihood function in Eq. 9 corresponds to the second period of the game.

3.4 NH-HMME likelihood and heterogeneity

The likelihood function of observing a sequence of strategy choices over the T rounds for player i can be written as

$$L_{iT}(S_i(1), ..., S_i(T)) = \boldsymbol{\omega}_i^{'}\mathbf{M}_{i1} \prod_{t=2}^{T} \mathbf{Q}_{it}\mathbf{M}_{it}\mathbf{1}. \tag{9}$$

where $\boldsymbol{\omega}_i$ is given by Eq. 4, $\mathbf{Q}_{it}$ is given by Eq. 5, $\mathbf{M}_{it}$ is a 2×2 diagonal matrix with the state specific probabilities in Eq. 8 representing the diagonal elements, and $\mathbf{1}$ is a 2x1 vector of ones.

In any dynamic model it is imperative to capture individual differences to disentangle player heterogeneity from dynamics (Heckman 1981; Keane 1997). In the context of behavioral games, Stahl (1999, 2000, 2001) demonstrated significant heterogeneity in the employment of learning rules across players. Wilcox (2006) showed, based on simulated data, that ignoring heterogeneity in the EWA model can lead to downward bias in the δ parameter (which measures the weight given to foregone payoffs) in Eq. 3. This implies that the EWA model tends to overstate the weight of reinforcement learning relative to belief learning when heterogeneity is ignored, which is consistent with the finding of spurious state dependence (Heckman 1981). Ho et al. (2008) found similar results and a high degree of cross-player heterogeneity in the EWA model parameters.

We allow for individual differences in two distinct ways. First, we allow for the possibility that individuals differ in the parameters that govern the updating of attractions in the reinforcement and belief submodels. This enables us to capture differences in how previous attractions are depreciated (via ϕ_i and $N_i(0)$) and in the scale parameters for the two rules. Second, we allow the parameters of the transition matrix to be heterogeneous across players. This captures differences in the propensity to employ different learning rules as well as in the rule switching and retention probabilities.

Specifically, let $\boldsymbol{\theta}_i$ contain all the parameters in the state-specific attraction functions, the initial state probabilities and the transition matrix for player i. The elements of $\boldsymbol{\theta}_i$ are appropriate transformations of the original parameters such that each element varies over the entire real line. We assume that $\boldsymbol{\theta}_i$ is distributed multivariate normal, i.e., $\boldsymbol{\theta}_i \sim \mathcal{N}(\boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta)$. The elements in $\boldsymbol{\mu}_\theta$ contain the population mean of the player-specific parameters, whereas, $\boldsymbol{\Sigma}_\theta$ captures the variation and covariation among the parameters, both within and across states.[5] We estimate the individual-level parameters, as well as the

[5]One can allow for a richer specification of heterogeneity by using a mixture of normals (e.g., Allenby et al. 1998) or by using Mixtures of Dirichlet Process priors (MDP, e.g., Ansari and Mela 2003; Ansari and Iyengar 2006).

population quantities, using a hierarchial Bayesian approach (see Appendix A for details).[6]

3.5 Inferring latent states

One advantage of the NH-HMME formulation is that we can probabilistically infer the learning rule employed by a player at each round of the game, a posteriori. Given the parameter estimates and the entire sequence of T choices made by player i, we use the *forward-filtering-backward-smoothing* approach (McDonald and Zucchini 1997) to calculate the probability that player i is in state z at round t. This smoothing probability is given by

$$P(Z_{it} = z | S_i(1), \ldots, S_i(T)) \\ = \boldsymbol{\omega}_i' \mathbf{M}_{i1} (\prod_{k=2}^{t-1} \mathbf{Q}_{it} \mathbf{M}_{ik}) \mathbf{Q}_{it} \mathbf{m}_{itz}^{c} \mathbf{q}_{itz}^{r} \mathbf{M}_{it+1} (\prod_{k=t+2}^{T} \mathbf{Q}_{it} \mathbf{M}_{ik}) \mathbf{1} / L_{iT}, \quad (10)$$

where, $\mathbf{m}_{itz}^{c}$ is the z^{th} column of the matrix $\mathbf{M}_{it}$, $\mathbf{q}_{itz}^{r}$ is the z^{th} row of the matrix $\mathbf{Q}_{it}$, and L_{iT} is the likelihood of the observed sequence of actions up to time T as given in Eq. 9.

4 Application to six behavioral games

We apply the proposed approach to six behavioral games previously used in the literature: Continental divide coordination game (Van Huyck et al. 1997), the median-action order statistic coordination game with several players (Van Huyck et al. 1990), mixed strategies (Mookherjee and Sopher 1997), R&D patent race (Rapoport and Amaldoss 2000), pot games (Amaldoss et al. 2005) and p-Beauty contest (Ho et al. 1998).[7]

Details of each game can be found in Ho et al. (2007).[8] Because our focus is on studying the dynamics in the use of learning rules within a game, we include only games that were played for at least ten rounds. This allows us a reasonable chance of uncovering different dynamic patterns. These six games vary along multiple dimensions. They support either mixed or pure strategy equilibria, and differ in terms of whether they are dominance solvable or not. They also

[6] We conducted a series of simulations to analyze the empirical identification of the NH-HMME parameters (Salmon 2001). We found that for data that mimics the data in our empirical application, the model parameters can be correctly recovered (see Appendix B for details).

[7] For the p-Beauty game, we follow Camerer and Ho (1999) and assume that players knew only the winning number and neglected the effect of their own choice on the target number. We modified the computation of forgone payoffs accordingly.

[8] We thank Professor Teck-Hua Ho of University of California at Berkeley and the authors of the corresponding papers for generously providing us with the behavioral games experimental data.

differ in terms of the number of players, the number of rounds played and the number of game strategies.

4.1 NH-HMME specification

In applying the NH-HMME model to data, one needs to specify the covariates, $\mathbf{x}_{it-1}$, that influence the transition probabilities. There are many ways in which the outcomes of the past rounds (either from the previous round or from earlier rounds) of play may affect the choice of learning rules in any given period. For instance, players could be affected by their own past outcomes or those of other players.

We tested the effect of several covariates on the transition matrix. To test for reinforcement-type outcomes, we examined whether the fact that one obtains a positive payoff in the previous round (as measured by a binary indicator) or the actual magnitude of the payoff in the previous round, influences rule transitions. Nevo and Erev (2010) suggested that a surprising outcome may trigger a change in the strategy choices made by a player. In the spirit of their specification, we examined the impact of gains and losses in the previous round relative to the player's expectations (Erev and Haruvy 2012). These expectations are captured by a player's average payoffs up to that round. Finally, to represent the impact of the actions of other players, we also investigated the influence of past forgone payoffs (i.e., whether and how much money the player left on the table in the previous round, given the action of opponents).

We tested these alternative covariates using the six games discussed above. Our results indicate that a covariate that represents whether the player left money on the table in the previous round (Erev and Roth 2007) performed the best in terms of fit and interpretability. This is consistent with Nevo and Erev (2010) who find that forgone outcomes have an effect on future playing behavior when feedback is available. Specifically, this covariate, $\mathbf{ForgonePayoff}_{it-1}$, can be written as

$$\mathbf{ForgonePayoff}_{it-1} = I_{(0,\infty)}(\max_{k\in\{J\}}[\pi(s_i^k, s_{-i}(t-1))] - \pi(s_i(t), s_{-i}(t-1))), \quad (11)$$

where $I_{(0,\infty)}(a)$ is an indicator function equal to 1 if $a \in (0,\infty)$ and 0, otherwise. This variable captures player i's deviation from the (a posteriori) optimal strategy in the previous round, given the strategies chosen by the other players. Such forgone payoffs in the previous round could trigger a transition into (or increase the likelihood of staying in) a belief learning state. Moreover, foregone payoffs are likely to have a stronger effect for players who are in the belief state, as these players are in a state of heightened sensitivity to such forgone payoffs.

We also studied covariate specifications involving a longer history of play. However, consistent with previous research, which suggests that the effect of past outcomes decays fast and can be concisely captured by the outcomes in the previous round of the game (Crawford 1995), we found that these covariate specifications were not well supported by the data.

4.2 Benchmark models

In addition to the NH-HMME model, we estimated ten benchmark models. These include,

1. **Reinforcement Learning**: In this model, all players are assumed to follow reinforcement learning, as in Eq. 1, for all rounds of the game.
2. **Belief Learning**: In this model, all players are assumed to follow belief learning, as in Eq. 2, throughout the game.
3. **EWA**: In this model, all players are assumed to follow the EWA learning model with static parameters, as in Eq. 3, over the course of the game.
4. **EWA time-varying** : This model follows Camerer and Ho (1998), who allow the parameters of the EWA model to monotonically increase or decrease over the course of the game.
5. **Mixture of Experts (ME)**: In this model, players can use either the reinforcement or the belief learning rule throughout the game and are not allowed to transition between these rules over the course of the game. This model is, therefore, a nested version of the proposed NH-HMME model and is obtained by setting the transition matrix to identity.
6. **HMME**: This model is a nested version of the proposed NH-HMME model in which the transition matrix is not a function of the covariates, and therefore, is constant over time.
7. **Forward HMME**: This is a change-point model in which players can only move from reinforcement to belief learning but cannot transition the other way. In other words, only forward transitions are allowed in the transition matrix in Eq. 5 and the belief learning state is an absorbing state of the Markov chain.
8. **Backward HMME**: This is a change-point model in which players can only move from belief to reinforcement learning but cannot transition the other way. In other words, only backward transitions are allowed in the transition matrix in Eq. 5 and the reinforcement learning state is an absorbing state of the Markov chain.
9. **HMM of 2 EWAs**: This is a hidden Markov model with two states, where each of the states represents a different parametrization of the EWA model.
10. **HMM of 3 EWAs**: This is a hidden Markov model with three states, where each of the states represents a different parametrization of the EWA model.

It is important to note that all models capture individual differences by allowing the player-specific attraction parameters as well as state transition parameters (where applicable) to vary according to a population distribution. This ensures a fair comparison of the benchmark models to our proposed model.

4.3 Parameter inference

We estimate all models using hierarchical Bayesian methods. These are eminently suitable in our context as they allow borrowing of strength across players. Such shrinkage is crucial for proper estimation and improved predictive performance as most behavioral game data sets contain a limited number of rounds of play. The highly nonlinear nature of the model in tandem with the heterogeneity specification results in a posterior distribution that is not available in closed form. We therefore use Markov chain Monte Carlo (McMC) methods for parameter inference. In particular, to facilitate rapid mixing of the resulting Markov chain, we use an Adaptive random walk Metropolis-Hastings algorithm (Atchadé 2006). In each instance, we ran the chain for 200,000 iterations of which, the first 100,000 are discarded as burn-in and the remaining 100,000 iterations are used for posterior inference. We used the first 70 % of the observations of each player for calibration and the last 30 % for validation in each game (Ho et al. 2008). Convergence was assessed by monitoring the trace plots of the McMC output and by running multiple parallel chains following Gelman and Rubin (1992). All parameters satisfied the Gelman and Rubin convergence criterion.

4.4 Results

4.4.1 Model selection

Table 1 summarizes the overall in-sample and out-of-sample fit and predictive ability of the estimated models averaged across the six games. For model comparison, we use the log-marginal density (LMD) computed on the calibration data. Additionally, we report two metrics commonly used in the related literature: in-sample and out-of-sample hit rates (HR) and mean squared deviation (MSD) between the chosen strategies and the choice probabilities. The LMD appropriately accounts for both model fit and complexity and the HR and MSD measures allow us to assess the predictive performance of

Table 1 Average model fit and prediction measures

	LMD	MSD		Hit rate	
Learning models		In	Out	In	Out
Reinforcement	−2,082.1	0.030	0.026	44.4	47.0
Belief	−2,017.4	0.030	0.028	42.6	43.3
EWA	−1,950.5	0.029	0.026	45.5	47.1
EWA time varying	−1,939.5	0.029	0.028	45.2	46.1
ME	−1,993.5	0.030	0.026	44.7	46.1
Forward-HMME	−1,977.9	0.030	0.026	45.3	46.8
Backward-HMME	−1,990.2	0.030	0.026	44.7	46.2
HMM 2 EWAs	−1,939.4	0.029	0.026	45.5	47.7
HMM 3 EWAs	−1,984.0	0.030	0.026	44.3	47.8
HMME	−1,941.0	0.029	0.026	46.3	47.6
NH-HMME	**−1,925.2**	**0.029**	**0.026**	**46.7**	**48.2**

different models on the calibration and holdout data. The measures for each game are weighted by the number of players and the number of periods in the game to compute the aggregate measures across games.

Table 1 reveals that (i) the NH-HMME model performs better than or equal to the alternative models for all fit and prediction measures. (ii) consistent with previous research (e.g., Camerer and Ho 1999) models involving a single learning rule (Belief or Reinforcement) perform worse than models that combine multiple rules (iii) generally, the NH-HMME, HMME, and time-varying EWA models fit and predict the data better than static models, suggesting that the observed data exhibit dynamics in learning rules; (iv) constrained versions of the HMME model (forward and backward only) both fit and predict the data worse than the HMME model; (v) the more complex models (HMM of two and three EWAs), which allow for a different mixture of belief and reinforcement learning in each state, do not add much beyond the standard EWA or the HMME in terms of penalized fit, as their LMD statistics are comparable. However, they slightly improve predictive ability.

Examining the fit and predictive performance measures (LMD, MSD and Hit rates) of the different models for individual games (see Table 7 in Appendix C) reveals that, while there is no one single model that performs uniformly better than others across all games, the proposed model "wins", relative to other models on the most number of games. Consistent with the findings of Camerer and Ho (1998), we find that the time-varying EWA has fairly good fit for several games, but this comes at the cost of poor predictive ability for some of the games (e.g., the continental divide game).

Overall, we conclude that the fit and predictive performance of the NH-HMME model is the best among the studied models and that both heterogeneity and dynamics are important for modeling data from behavioral games. For the NH-HMME model, incorporating past outcomes improves the fit and predictive ability of the dynamic model, and, as we discuss later, it also enhances our understanding of the underlying learning process.

4.4.2 Parameter estimates

We now look at the parameter estimates of the proposed NH-HMME model. Table 2 reports the posterior means and the 95 % central posterior intervals for the population mean of the NH-HMME model's parameters for each of the six games.[9] We focus our discussion on the parameters of substantive interest, i.e., those that inform about the learning rule dynamics.

The high values of the diagonal elements of the transition matrix for all games, (q_{RR} and q_{BB}), suggest that the states are relatively sticky. That is, players exhibit, on average, a high propensity to repeat the learning rule from

[9]To ensure that N is non-decreasing over time, we follow the previous literature by imposing $N^0 \leq 1/(1-\rho)$, where $N^0 = N(t=0)$, for the belief, EWA, ME, HMMs, HMME, and NH-HMME models (Ho et al. 2002). For the time-varying EWA, we impose this constraint by constraining ρ_t to be non-decreasing in t.

Table 2 Parameter estimates for the NH-HMME model

Parameter label	Continental divide	Median action	Mixed strategy	Patent race	Pot games	p-Beauty contest
λ_R	20.29	6.44	0.75	0.22	0.16	0.72
	(9.05, 68.26)	(3.21, 14.06)	(0.31, 1.92)	(0.11, 0.53)	(0.10 , 0.26)	(0.41, 1.25)
ϕ_R	0.26	0.74	0.99	0.85	0.24	0.65
	(0.08, 0.51)	(0.35, 0.96)	(0.99, 1.00)	(0.71, 0.92)	(0.07, 0.51)	(0.50, 0.77)
λ_B	42.66	22.77	1.36	0.81	2.08	1.63
	(25.54, 73.65)	(10.52, 47.53)	(0.41, 4.66)	(0.52, 1.30)	(0.71, 5.70)	(1.24, 2.15)
ϕ_B	0.84	0.31	0.82	0.73	1.00	0.51
	(0.70, 0.92)	(0.05 , 0.76)	(0.41, 0.97)	(0.51, 0.87)	(0.99, 1.00)	(0.37, 0.62)
N_B^0	5.69	0.83	1.42	1.83	14.13	1.11
	(2.81, 11.35)	(0.12, 1.93)	(0.17, 11.84)	(0.73, 4.01)	(2.11, 66.48)	(0.76, 1.55)
q_{RR}	0.63	0.96	0.81	0.73	0.87	0.47
	(0.35, 0.85)	(0.78, 0.99)	(0.38, 0.97)	(0.54, 0.85)	(0.72, 0.93)	(0.35, 0.60)
q_{BB}	0.61	0.30	0.28	0.43	0.96	0.66
	(0.41, 0.79)	(0.08, 0.69)	(0.07, 0.61)	(0.25, 0.64)	(0.86, 0.99)	(0.56, 0.76)
$\boldsymbol{\rho}_R$	−1.61	−2.76	1.81	−0.42	0.63	0.03
	(−2.88, −0.26)	(−5.06, −1.04)	(0.04, 3.38)	(−1.19, 0.43)	(−0.30, 2.07)	(−0.49, 0.52)
$\boldsymbol{\rho}_B$	1.84	3.43	1.85	0.59	−2.14	0.54
	(0.62, 3.00)	(0.99, 5.71)	(0.41, 3.51)	(−0.43, 1.51)	(−3.04, −1.18)	(0.07, 1.06)
ω_R	0.13	0.14	0.17	0.38	0.98	0.50
	(0.04, 0.29)	(0.03, 0.40)	(0.03, 0.56)	(0.17, 0.64)	(0.93, 0.99)	(0.36, 0.62)

one round to the next. Relative to the pure equilibrium games, the mixed equilibria games (e.g., mixed strategy and patent race) exhibit lower stickiness to the belief rule state. This result is consistent with the findings that the reinforcement model generally performs well in mixed equilibria games (Erev and Roth 1998). The value of ω_R indicates that for most of the games (except for pot games and p-Beauty) players have a tendency to initiate game play with belief learning. However, as the game progresses, players shift towards a reinforcement rule. We elaborate on these dynamics in the next subsection.

As seen in Table 2, several of the parameters that capture the effect of foregone payoffs on the transition matrix are significantly different from zero. Thus, less than optimal actions may lead to a change in the learning rule being used by players. We find stronger and more statistically significant effect for the effect of leaving money on the table in the previous round on players in the belief state relative to players in the reinforcement state. Indeed, in the belief state players are postulated to be more attuned to forgone payoffs.

To illustrate the effect of leaving money on the table in the previous round on the transitions between learning rules, we report in Table 3, the average transition matrix $\mathbf{Q}$, when $x_{t-1} = 0$ and when $x_{t-1} = 1$. The covariate value $x_{t-1} = 0$ represents instances when, conditioned on the opponents actions, the the best strategy was chosen in the previous round, and $x_{t-1} = 1$, represents instances when the player did not play the payoff maximizing strategy in the previous round. Each cell in Table 3 represents $q_{zz'}$, the probability of transitioning from learning rule z to learning rule z'. We find that for five out of the six games, leaving money on the table in the previous round substantially increased the player's likelihood of staying in the belief state, given that the player was already in that state in the previous round. Moreover, when using reinforcement learning within coordination games (continental divide and median action), not playing the optimal strategy in the previous round induces

Table 3 Change in mean transition matrices as a function of forgone payoff

		$\mathbf{Q}(x_{t-1} = 0)$		$\mathbf{Q}(x_{t-1} = 1)$	
		R	B	R	B
Continental divide	R	0.63	0.37	0.25	0.75
	B	0.39	0.61	0.09	0.91
Median action	R	0.96	0.04	0.61	0.39
	B	0.70	0.30	0.06	0.94
Mixed strategies	R	0.81	0.19	0.96	0.04
	B	0.73	0.27	0.29	0.71
Patent race	R	0.72	0.28	0.64	0.36
	B	0.57	0.43	0.42	0.58
Pot games	R	0.81	0.19	0.88	0.12
	B	0.08	0.92	0.35	0.65
p-Beauty contest	R	0.47	0.53	0.48	0.52
	B	0.34	0.66	0.23	0.77

Table 4 Parameter estimates EWA

Parameter label	Continental divide	Median action	Mixed strategy	Patent race	Pot games	p-Beauty contest
ϕ	0.78	0.95	0.99	0.92	0.61	0.85
	(0.58, 0.91)	(0.82, 0.99)	(0.99, 1.00)	(0.87, 0.95)	(0.28, 0.83)	(0.77, 0.91)
δ	0.88	0.92	0.31	0.58	0.14	0.62
	(0.82, 0.92)	(0.87, 0.95)	(0.14, 0.56)	(0.50, 0.67)	(0.01, 0.54)	(0.46, 0.78)
λ	6.89	10.88	0.32	0.79	0.17	0.97
	(5.02, 9.65)	(7.41, 15.99)	(0.14, 0.83)	(0.53, 1.22)	(0.07, 0.40)	(0.74, 1.31)
ρ	0.07	0.08	0.94	0.91	0.18	0.13
	(0.01, 0.25)	(0.02, 0.28)	(0.77, 0.99)	(0.83, 0.95)	(0.01, 0.72)	(0.05, 0.31)
$N(0)$	0.96	0.48	12.66	2.02	0.13	1.13
	(0.63, 1.20)	(0.07, 1.00)	(3.15, 54.76)	(0.45, 5.89)	(0.01, 1.14)	(1.03, 1.43)

a shift towards belief learning. For most games, the effect of the forgone-payoff covariate on players in the reinforcement state is relatively small. Notice that for the p-Beauty game, with 101 strategies to chose from, it is difficult for players to evaluate whether money was left on the table in the previous round. Furthermore, in this game players were not informed of the winning startegy at each round (Ho et al. 1998). Accordingly, the effect of x_{t-1} on the transition probabilities for the p-Beauty game in Table 3 is fairly small.

It is instructive to compare our NH-HMME results to those of the EWA model. Table 4 reports the posterior means and the 95% central posterior intervals of the population mean for the EWA model's parameters. First, we note that our parameters are somewhat different from the EWA estimates reported in the previous literature because of our richer specification of cross-player heterogeneity.[10] For example, our δ estimates are generally higher than those reported in the literature. This result is consistent with evidence of over-estimation of reinforcement behavior in the absence of heterogeneity (Wilcox 2006). More important is the contrast between the EWA and NH-HMME estimates. Of particular interest is the δ parameter which captures the weight given to forgone payoffs relative to actual payoffs and thus reflects the relative emphasis on belief learning in comparison to reinforcement learning behavior. It is interesting to note that the EWA results suggest that players in both the continental divide and median action games mainly use belief learning rules (high δ's), whereas, in the mixed strategy and pot games, players mainly use reinforcement learning.

The NH-HMME model provides different insights. While indeed in both the continental divide and median action games, player's are likely to start the game in the belief state (ω_R is small), only in the continental divide game are players more likely to stick to the belief state (q_{BB} is much larger for continental divide game relative to the median action game).

[10]Estimating our EWA models without heterogeneity resulted in estimates that are very similar to those reported in the literature.

The NH-HMME model allows us to understand such differences in dynamics in learning rules given its ability to uncover the latent state at each round of play. We discuss such dynamics in rule learning next.

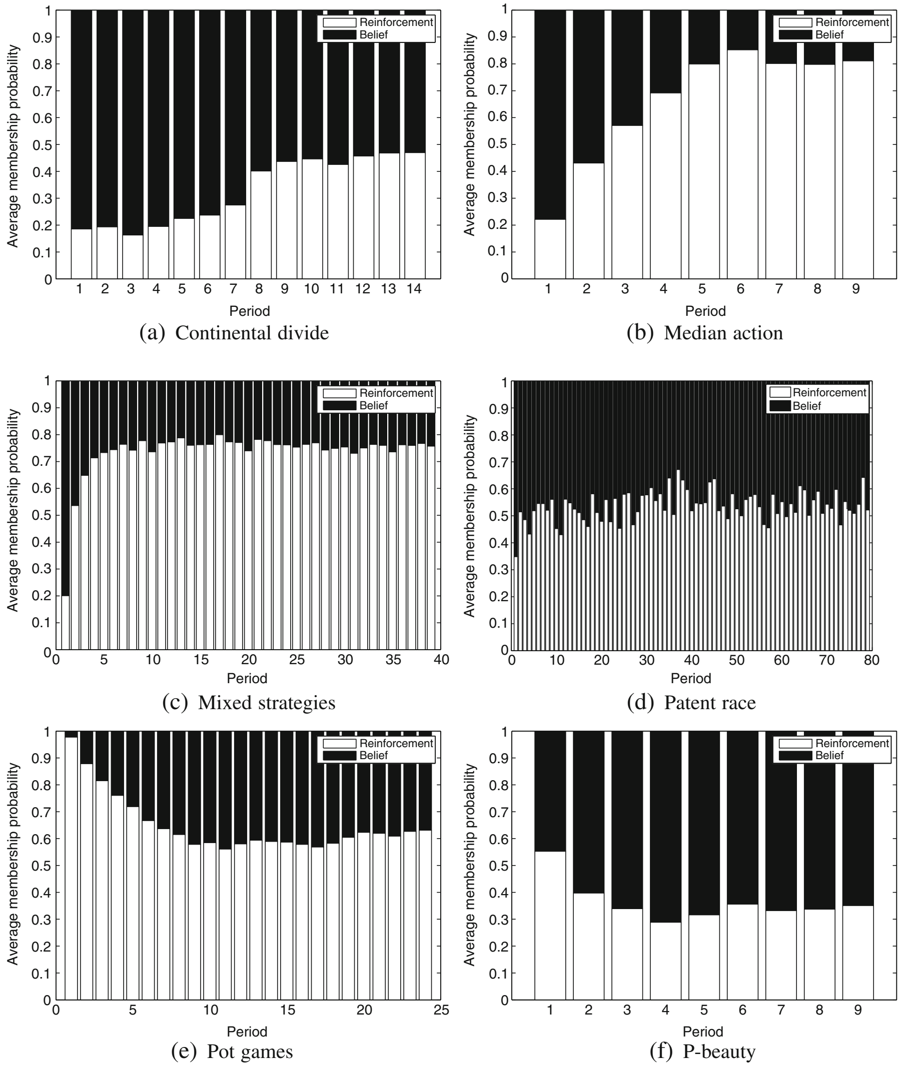
4.4.3 Learning rule dynamics

In this section we analyze in greater detail the dynamics in the learning behavior of players over the course of a game. We use Eq. 10 to compute the smoothed estimates of the posterior probability of using each learning rule at any round of the game. Figure 2 summarizes these probabilities by averaging across all players. It should be noted that these are average dynamics across players and that specific individuals may exhibit learning paths that diverge substantially from the aggregate.

As can be seen from Fig. 2, there are substantial differences in the rule dynamics across games. On average, the continental divide, median action, mixed strategy and pot games exhibit stronger dynamics compared to the patent race and *p*-Beauty games. Moreover, on average, players in games such as mixed strategies and p-beauty seem to "stabilize" in the use of the two underlying learning rules after a few rounds in the game. In contrast, in games such as median action and pot games, the player population needs more rounds to stabilize. The median action game provides an example of interesting dynamics in the use of learning rules over the course of the game. The NH-HMME model suggests that even though players in this game primarily use the reinforcement learning rule, they are much more likely to use belief learning early in the game. Owing to the strong stickiness of the reinforcement learning state (96 %) relative to the belief learning state (30 %), once players move to the reinforcement learning rule, they continue to use it subsequently (see Fig. 2b). The relative attenuation of rule dynamics in the patent race game is consistent with observed stickiness to the chosen strategy reported by Ho et al. (2007). Similarly, the drift from belief learning to reinforcement learning in the continental divide game is consistent with the pattern of shifts commonly observed in the strategy choices of this game (Camerer et al. 2002b).

For the mixed strategies and pot games, the EWA model suggests a low recognition of forgone payoffs in choosing game strategies. The parameter δ, which captures the relative emphasis on foregone payoffs compared to actual payoffs, is the lowest among all games. This reveals the emphasis on reinforcement learning in these games. The NH-HMME results, corroborate this fact but provide additional insights about the pattern of dynamics, which differs across the two games. For the mixed strategies games, players are likely to start the game in the belief state and transition over time to the reinforcement state. In contrast, for the pot games, the likelihood of starting the game in the reinforcement state is very high and players transition over time to a moderate use of the belief state.

The above discussion focussed on the average dynamics in the use of the learning rules across players. Recall, however, that our model allows



(a) Continental divide (b) Median action

(c) Mixed strategies (d) Patent race

(e) Pot games (f) P-beauty

Fig. 2 Population learning rules dynamics across games

us to estimate different parameters for each player. Figure 3 illustrates the heterogeneity in individual-level rule dynamics across games. In particular, we plot the values of q_{iRR} and q_{iBB} for each individual player in each game (at $x_{it-1} = 0$).

It is clear from the figure that the games differ with respect to the degree of heterogeneity in rule-dynamics. Specifically, the median action and pot games exhibit relatively low levels of heterogeneity in the transition parameters. Players in the mixed strategies and the p-Beauty game are quite homogenous in

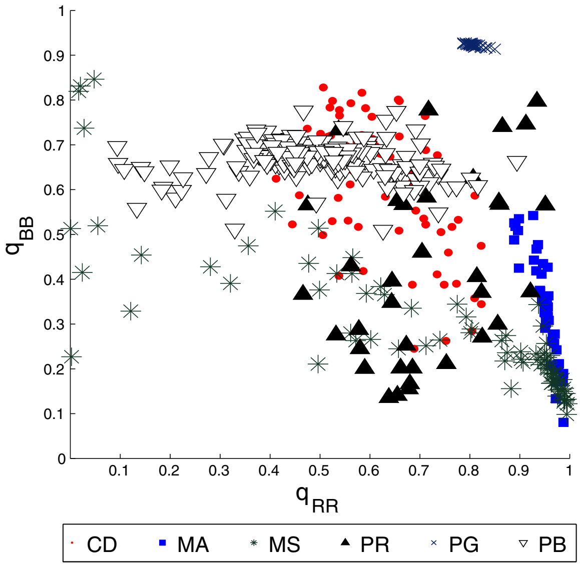


Fig. 3 Heterogeneity in dynamic learning across games: *CD*—continental divide; *MA*—median action; *MS*—mixed strategies; *PR*—patent race; *PG*—pot games; and *PB*—*p*-Beauty

their stickiness to the belief rule but are quite heterogenous in their stickiness to the reinforcement rule. The high degree of heterogeneity among players to both learning-rule states in the continental divide game may be attributed, in part, to the sensitivity of this game to the initial conditions (early strategy choices). Interestingly, the mixed equilibria games (e.g., mixed strategy and patent race) exhibit high degree of heterogeneity, possibly due to different players focusing on different equilibria. The high degree of heterogeneity in the *p*-Beauty game can be attributed to the large number of strategies in this games, leading to heterogenous outcome histories across players.

5 Conclusions and future work

Learning is fundamental to the understanding of how equilibrium arises in strategic situations. Several models of learning, such as the reinforcement and belief learning, as well as the EWA model which nests these learning processes, have been developed to describe how players choose strategies in behavioral games. These learning rules differ in the mechanisms by which attraction functions are updated from one round of play to another. In this paper, we proposed that players can transition between alternative learning rules over the course of the game. This creates an additional source of learning dynamics over and above what is captured via dynamic attraction functions. Understanding these rule dynamics can provide additional insights about the learning process and can also improve predictive performance.

We proposed a NH-HMME model in which players can transition in a Markovian manner between reinforcement and belief learning submodels over the course of the game. Experiences in the game can trigger transitions between the learning rules. We estimated our model using a hierarchical Bayesian approach to account for cross-player heterogeneity and to empirically disentangle heterogeneity from dynamics. We fit the model to six experimental games that have been previously studied in the literature. We found that, on average, across the six games studied, the proposed NH-HMME model provides better in-sample fit and out-of-sample predictions of the strategy choices.

More importantly, the proposed model sheds light on how reinforcement and belief learning rules are employed over the course of the game. The six games differ in the extent to which each rule is used and in the dynamic pattern of the usage of these rules. Such differences cannot be discerned from traditional models that yield structural parameters that capture the gestalt, but do not provide information about the shifts in the employment of learning rules. For example, the mixed strategies games, which involve mixed equilibria, exhibited a quick transition from belief to reinforcement rules. The continental divide game, on the other hand, showed a much slower drift from belief to reinforcement learning over the course of the game. We also found a high degree of heterogeneity in the choice of learning rules across players. We believe that understanding the path that decision makers take in learning from their own decisions and the decisions of their opponents can aid the study of managerial decision making (e.g., Goldfarb and Yang 2009).

There are several possible extensions to our paper. First, we examined only the transition between specific forms of reinforcement and belief learning rules. Future research could examine transitions among additional learning rules (Stahl 1999, 2001). For example, one could examine whether players are likely to transition between reciprocity and aspiration learning (Stahl and Haruvy 2002). Second, future research could apply our modeling framework to other behavioral games. This will facilitate a systematic understanding of the dimensions that underly the differences in the pattern of latent rule dynamics. We believe that the biggest opportunity is in games that involve a large number of rounds. Third, one could examine the transferability of the parameter estimates from one game to another (Haruvy and Stahl 2012). Fourth, the observation that players “stabilize” their learning behavior in some games could be further explored. A more complex NH-HMME model that allows for more flexible dynamics in the transition probabilities, can help to improve our understanding of players’ switching behavior between learning rules. Finally, one can explore richer specifications for parameter heterogeneity, such as mixture of normals or mixtures of Dirichlet process priors.

Acknowledgement The authors thank Professor Teck H. Ho for kindly providing them the data used in the paper. The second author gratefully acknowledges partial funding by the Millennium Institute on Complex Engineering Systems.

Appendix A: NH-HMME prior and full conditionals distributions

We denote by $\boldsymbol{\theta}_i = \left\{\tau_{iR}, \tau_{iB}, \boldsymbol{\rho}_{iR}, \boldsymbol{\rho}_{iB}, \widetilde{\omega}_{iR}, \lambda_{iR}, \lambda_{iB}, \phi_{iR}, \phi_{iB}, N^0_{iB}\right\}$, the set of random-effect parameters. See Eqs. 4, 7 and 8 and the attraction functions for the belief and reinforcement learning models.

A.1 Priors

The prior for the population mean $\boldsymbol{\mu}_\theta$ is assumed to be multivariate normal, given by $N(\boldsymbol{\mu}_0, \mathbf{V}_0)$ and the prior for the population precision matrix is assumed to be Wishart, $\Sigma_\theta^{-1} \backsim W(df_0, \mathbf{S}_0)$, where, the hyper-parameters are chosen so as to obtain proper, but diffuse priors.

We use the following uninformative prior hyperparameters for the two-state NH-HMME model: $\boldsymbol{\mu}_0 = [0, 0, 0, 0, 0, -0.3, -0.3, 0, 0, 0]$, $\mathbf{V}_0 = \mathbf{I}_{n\theta X n\theta}$, $df_0 = n\theta + 2$, $\mathbf{S}_0 = (df_0 - n\theta - 1)\mathbf{I}_{n\theta}$, Where, N is the number of individuals, and $\mathbf{I}_{n\theta}$ denotes an identity matrix of rank $n\theta$, which represents the dimensionality of $\boldsymbol{\mu}_\theta$.

An appropriate selection of the priors is particularly critical in this model since the parameters are transformed to an exponential scale. Moreover, we model heterogeneity using a hierarchical structure, where the variance of $\boldsymbol{\mu}_\theta$ and Σ_θ are added at each individual draw of each $\boldsymbol{\theta}_i$. This increases the variance of the transformed variables. Accordingly, we chose prior hyperparameters for $\boldsymbol{\mu}_\theta$ and Σ_θ such that their priors are diffused in the *transformed* space.

A.2 Posterior

The posterior distribution is proportional to the product, $p(\text{data}|\{\boldsymbol{\theta}_i\})$ $p(\{\boldsymbol{\theta}_i\}|\boldsymbol{\mu}_\theta, \Sigma_\theta)p(\boldsymbol{\mu}_\theta)p(\Sigma_\theta)$, where, $p(data|\{\boldsymbol{\theta}_i\})$ is the sampling density conditional on the individual-level parameters and $p(\{\boldsymbol{\theta}_i\}|\boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta) = \prod_i p(\boldsymbol{\theta}_i|\boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta)$ is the joint distribution of the individual-level parameters. The posterior is not available in closed form as the likelihood is not conjugate to the priors, and hence we use McMC methods to sample from the joint posterior.

A.3 Full conditionals

1. The full conditional distribution for the individual-level parameters $\boldsymbol{\theta}_i$ is unknown and can be written as

$$p(\boldsymbol{\theta}_i|\boldsymbol{\mu}_\theta, \Sigma_\theta, \text{data}_i) \propto p(\text{data}_i|\boldsymbol{\theta}_i)\exp\left(-\frac{1}{2}(\boldsymbol{\theta}_i - \boldsymbol{\mu}_\theta)'\Sigma_\theta^{-1}(\boldsymbol{\theta}_i - \boldsymbol{\mu}_\theta)\right).$$

   We use a Metropolis-Hastings step to draw from it. In particular, we use a Gaussian random-walk M-H where the candidate vector of parameters $\boldsymbol{\varphi}^{(c)}$ for $\boldsymbol{\theta}_i$ at iteration t is drawn from $N(\boldsymbol{\theta}^{(t-1)}, \sigma^{2(t-1)}\Delta^{(t-1)})$ and accepted using the M-H acceptance ratio. The tuning parameters $\sigma^{2(t-1)}$ and $\Delta^{(t-1)}$

are adaptively chosen to yield an acceptance rate of approximately 30 %. We use the method proposed by Atchadé (2006) to adapt these tuning parameters.

2. The full conditional for the population mean is given by a multivariate normal distribution, i.e., $\boldsymbol{\mu}_\theta \backsim N(\boldsymbol{\mu}_n, \mathbf{V}_n)$, where, $\mathbf{V}_n^{-1} = \mathbf{V}_0^{-1} + N\Sigma_\theta^{-1}$ and $\boldsymbol{\mu}_n = \mathbf{V}_n[\boldsymbol{\mu}_0\mathbf{V}_0^{-1} + N\bar{\theta}\Sigma_\theta^{-1}]$, N being the total number of players.
3. The full conditional for the population precision is a Wishart distribution, i.e., $\Sigma_\theta^{-1} \backsim W(df_1, \mathbf{S}_1)$, where, $df_1 = df_0 + N$ and $\mathbf{S}^{-1} = \sum_{i=1}^{N}(\boldsymbol{\theta}_i - \boldsymbol{\mu}_\theta)(\boldsymbol{\theta}_i - \boldsymbol{\mu}_\theta)' + \mathbf{S}_0^{-1}$.

Inference is based on making multiple draws from the above full conditional distributions. An initial set of draws are discarded as burn-in iterations to reduce the impact of atypical starting values for the parameters on final inference.

Appendix B: Simulation

We conduct a simulation exercise to study the empirical identification of the NH-HMME parameters.

Data generation We use a procedure similar to the one used by Salmon (2001). In the context of the continental divide game used in our empirical application, we simulate 70 individuals playing the game for 15 periods. In the continental divide game, individuals play in groups, in each round, each member of the group chooses an integer between 1 and 14. At the end of the round, the median of the numbers chosen by the members of the group is revealed and the payoffs are computed.

We simulate individuals using the proposed NH-HMME in which individuals transition between reinforcement and belief learning rule states. To provide a realistic test of parameter recovery, the parameter values used in the simulation correspond to the values estimated in the empirical application for that game. We then estimated five models: (i) Reinforcement, (ii) Belief, (iii) EWA, (iv) EWA time varying and (v) NH-HMME. Notice that the first two models are nested versions of the NH-HMME, and correspond to a NH-HMME with only one state. Table 5 summarizes the performance of each model regarding fit and predictive ability.

Table 5 Model selection

Learning model	LMD	MSD-in	MSD-out	HR in	HR out
Belief	−1084.8	0.050	0.050	38.1	43.9
Reinforcement	−1404.6	0.052	0.048	45.0	53.9
EWA	−1000.9	0.046	0.045	46.6	50.4
EWA time varying	−992.0	0.044	0.057	47.7	42.5
NH-HMME	**−958.9**	**0.044**	**0.044**	**49.9**	**56.1**

Table 6 Transformed parameter estimates. Data generated according to a NH-HMME

Par. label	True pars	2.5 %	50 %	97.5 %
ϕ_R	0.26	0.07	0.20	0.37
λ_R	20.29	13.64	27.94	68.39
ϕ_B	0.84	0.73	0.84	0.91
λ_B	42.66	27.67	40.10	59.04
N_B	5.69	1.42	4.34	10.15
q_{11}	0.63	0.44	0.65	0.79
q_{22}	0.61	0.54	0.69	0.81
ρ_R	−1.61	−2.83	−1.75	−0.66
ρ_B	1.84	0.55	1.39	2.33
w_R	0.13	0.08	0.17	0.30

As expected, the results in Table 5 confirm that the NH-HMME is the model that fits the data best. This model also predict best the holdout sample.

We now explore the identification of the NH-HMME parameters. Table 6 shows the parameters estimated for the NH-HMME model. The estimation procedure did very well in recovering the true parameter values that were used to generate the data. We can observe that the true values for all parameters fall within the 95 % highest posterior density (HPD) interval estimated by the NH-HMME.

Finally, we analyze membership recovery. Figure 4 shows the true and the estimated average membership probability for each round of the game. We can see that the estimated model does well in recovering the true underlying learning rules states.

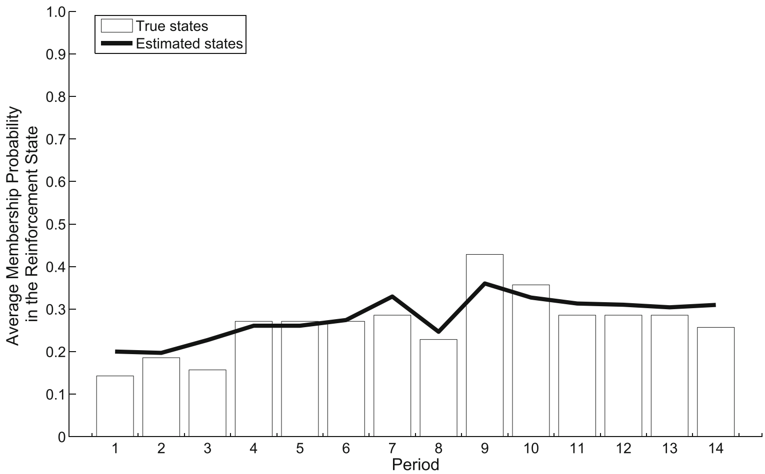


Fig. 4 Recovery of the population learning rule states

Appendix C: Fit and predictive ability for each game

Table 7 Model fit and prediction measures

	Continental divide					Median action				
	LMD	MSD		Hit rate		LMD	MSD		Hit rate	
Learning models		In	Out	In	Out		In	Out	In	Out
Reinforcement	−1370.7	0.052	0.034	45.1	66.4	−291.1	0.060	0.033	67.9	86.4
Belief	−1062.0	0.048	0.042	42.0	54.6	−232.8	0.045	0.017	74.1	94.4
EWA	−1022.3	0.046	0.034	48.1	64.6	−215.0	0.042	0.022	79.3	90.7
EWA time varying	−967.8	0.042	0.037	50.9	60.7	−180.8	0.035	0.026	80.2	89.5
ME no dynamics	−1094.1	0.049	0.041	46.0	56.4	−236.9	0.045	0.025	77.2	90.1
F-HMME	−1062.0	0.047	0.039	47.3	56.8	−222.6	0.046	0.018	78.1	92.0
B-HMME	−1101.2	0.049	0.041	45.4	55.7	−239.9	0.046	0.024	75.3	90.7
HMM 2 EWAs	−1035.5	0.046	0.032	47.1	67.1	−213.4	0.045	0.016	77.2	93.2
HMM 3 EWAs	−1114.3	0.050	0.033	42.9	66.1	−227.3	0.047	0.018	74.4	90.7
HMME	−1018.9	0.045	0.035	50.1	66.1	−210.7	0.039	0.023	80.2	90.1
NH-HMME	−1009.5	0.045	0.033	49.4	65.7	−197.3	0.038	0.022	80.2	89.5

	Continental divide					Median action				
	LMD	MSD		Hit rate		LMD	MSD		Hit rate	
Learning models		In	Out	In	Out		In	Out	In	Out
Reinforcement	−3046.6	0.148	0.149	39.5	38.3	−2200.4	0.087	0.084	62.7	64.1
Belief	−3063.5	0.149	0.152	38.2	38.0	−2317.7	0.093	0.096	58.3	54.5
EWA	−2999.3	0.146	0.149	39.9	37.6	−2062.9	0.084	0.083	62.4	63.2
EWA time varying	−3098.0	0.151	0.158	36.9	37.7	−2046.1	0.083	0.084	62.7	60.2
ME no dynamics	−3052.3	0.149	0.149	38.9	38.0	−2147.7	0.086	0.086	62.4	62.5
F-HMME	−3014.7	0.147	0.150	40.0	37.1	−2157.7	0.086	0.085	62.5	63.0
B-HMME	−3037.5	0.148	0.149	39.3	38.0	−2131.6	0.086	0.084	62.8	63.7
HMM 2 EWAs	−2976.4	0.145	0.148	39.8	37.7	−1992.1	0.081	0.083	63.8	64.0
HMM 3 EWAs	−3041.2	0.100	0.149	38.5	37.2	−2053.0	0.083	0.080	62.8	65.3
HMME	−3009.2	0.147	0.150	40.0	38.0	−2026.4	0.082	0.083	63.6	63.8
NH-HMME	−2973.5	0.144	0.150	40.4	37.9	−2019.6	0.081	0.084	63.7	63.8

	Continental divide					Median action				
	LMD	MSD		Hit rate		LMD	MSD		Hit rate	
Learning models		In	Out	In	Out		In	Out	In	Out
Reinforcement	−367.8	0.215	0.197	64.6	70.1	−5216.3	0.010	0.010	5.6	4.6
Belief	−378.8	0.218	0.219	60.6	64.9	−5049.5	0.010	0.010	6.7	5.3
EWA	−365.0	0.215	0.195	65.1	69.8	−5038.5	0.010	0.010	7.0	7.5
EWA time varying	−368.8	0.213	0.213	64.2	68.4	−4975.5	0.009	0.011	8.8	9.2
ME no dynamics	−364.7	0.216	0.192	65.8	69.4	−5065.2	0.010	0.010	5.4	6.8
F-HMME	−366.7	0.216	0.191	65.8	70.1	−5043.9	0.010	0.010	6.1	10.0
B-HMME	−366.2	0.216	0.196	64.9	70.5	−5064.9	0.010	0.010	5.4	5.4
HMM 2 EWAs	−365.5	0.217	0.194	64.8	69.1	−5053.8	0.010	0.010	6.0	7.7
HMM 3 EWAs	−363.7	0.218	0.192	65.8	70.1	−5104.4	0.010	0.010	6.0	8.2
HMME	−367.4	0.215	0.193	66.0	70.5	−5013.2	0.010	0.010	7.5	7.8
NH-HMME	−363.3	0.215	0.193	65.8	71.5	−4987.5	0.010	0.010	9.4	11.1

References

Allenby, G., Arora, N., Ginter, J. (1998). On the heterogeneity of demand. *Journal of Marketing Research, 34*(3), 384–389.

Ansari, A., & Iyengar, R. (2006). Semiparametric Thurstonian models for recurrent choices: a Bayesian analysis. *Psychometrika, 71*(4), 631–657.

Ansari, A., & Mela, C. (2003). E-customization. *Journal of Marketing Research, 40*(22), 131–145.

Amaldoss, W., Chong, J.-K., Ho, T.H. (2005). Choosing the right pond: EWA learning in games with different group sizes. Working paper, Department of Marketing, University of California, Berkeley.

Atchadé, Y.F. (2006). An adaptive version for the metropolis adjusted Langevin algorithm with a truncated drift. *Methodology and Computing in Applied Probability, 8*(2), 235–254.

Brown, G.W. (1951). Iterative solution of games by fictitious play. In T. Jallings, & C. Koopmans (Eds.), *Activity analysis of production and allocation*. New York: Wiley.

Camerer, C. (2003). *Behavioral game theory: Experiments on strategic interaction.* Princeton: Princeton University Press.

Camerer, C., & Ho, T.H. (1998). EWA learning in coordination games: probability rules, heterogeneity, and time variation. *Journal of Mathematical Psychology, 42*, 305–326.

Camerer, C., & Ho, T.H. (1999). Experience-weighted attraction learning in games. *Econometrica, 87*, 827–874.

Camerer, C., Ho, T.H., Chong, J.-K. (2002a). Sophisticated learning and strategic teaching. *Journal of Economic Theory, 104*, 137–188.

Camerer, C., Ho, T.H., Chong, J.-K. (2003). Models of thinking, and teaching in games. *The American Economic Review, 93*(2), 192–195.

Camerer, C., Ho, T.H., Chong, J.-K. (2007). Self-tuning experience weighted attraction learning in games. *Journal of Economic Theory, 133*, 177–198.

Camerer C., Hsia, D., Ho, T.H. (2002b). EWA learning in bilateral call markets. In A. Rapoport, & R. Zwick (Eds.), *Experimental business research*.

Cournot, A. (1960). *Recherches sur les Principes Mathematiques de la Theorie des Richesses.* Translated into English by N. Bacon as Researches in the Mathematical Principles of the Theory of Wealth. London: Haffner.

Crawford, V. (1995). Adaptive dynamics in coordination games. *Econometrica, 63*, 103–143.

Erev, I., & Haruvy, E. (2005). Generality, repetition, and the role of descriptive learning models. *Journal of Mathematical Psychology, 49*(5), 357–371.

Erev, I., & Haruvy, E. (2012). Learning and the economics of small decisions. In J. Kagel, & A. Roth (Eds.), *The handbook of experimental economics* (Vol. 2).

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

Erev, I., & Roth, A. (2007). Multi-agent learning and the descriptive value of simple models. *Artificial Intelligence, 171*, 423–428.

Fudenberg, D., & Levine, D. (1998). *The theory of learning in games.* Cambridge: MIT.

Gelman, A., & Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science, 7*, 457–472.

Goldfarb, A., & Yang, B. (2009). Are all managers created equal? *Journal of Marketing Research, 46*(5), 612–622.

Haruvy, E., & Erev, I. (2002). On the application and interpretation of learning models. In R. Zwick, & A. Rapoport (Eds.), *Experimental business research* (pp. 285–300). Boston: Kluwer Academic.

Haruvy, E., & Stahl, D. (2012). Learning transference between dissimilar symmetric normal-form games. *Games and Economic Behavior, 74*(1), 208–221.

Heckman, J.J. (1981). Heterogeneity and state dependence. In S. Rosen (Ed.), *Studies in labor markets*(pp. 91–139). ChicagoL University of Chicago Press.

Ho, T.H., Camerer, C., Chong, J.-K. (2002). Functional EWA: a one-parameter theory of learning in games. Working paper, University of California, Berkeley.

Ho, T.H., Camerer, C., Chong, J.-K. (2007). Self-tuning experience weighted attraction learning in games. *Journal of Economic Theory, 133*, 177–198.

Ho, T.H., Camerer, C., Weigelt, K. (1998). Iterated dominance and iterated best-response in experimental P-beauty contests. *American Economic Review, 88*, 947–969.
Ho, T.H., Wang, X., Camerer, C. (2008). Individual differences in the EWA learning with partial payoff information. *The Economic Journal, 118*, 37–59.
Hück, S., H-Normann, T., Oechssler, J. (1999). Learning in Cournot oligopoly—an experiment. *The Economic Journal, 109*, 80–95.
Jordan, M.I., & Jacobs, R.A. (1994). Hierarchical mixtures of experts and the EM algorithm. *Neural Computation, 6*(2), 181–214.
Keane, M.P. (1997). Modeling heterogeneity and state dependence in consumer choice behavior. *Journal of Business and Economic Statistics, 15*(July), 310–327.
Kunreuther, H., Silvasi, G., Bradlow, E.T., Small, D. (2009). Bayesian analysis of deterministic and stochastic prisoner's dilemma games. *Judgment and Decision Making, 4*(5), 363–384.
McDonald, I., & Zucchini, W. (1997). *Hidden Markov and other models for discrete-valued time series*. London: Chapman & Hall.
Montoya, R., Netzer, O., Jedidi, K. (2010). Dynamic allocation of pharmaceutical detailing and sampling for long-term profitability. *Marketing Science, 29*(5), 909–924.
Mookherjee, D., & Sopher, B (1994). Learning behavior in an experimental matching pennies game. *Games and Economic Behavior, 7*, 62–91.
Mookherjee, D., & Sopher, B. (1997). Learning and decision costs in experimental constant sum games. *Games and Economic Behavior, 19*(1), 97–132.
Netzer, O., Lattin, J.M., Srinivasan, V. (2008). A hidden Markov model of customer relationship dynamics. *Marketing Science, 27*(2), 185–204.
Nevo, I., & Erev, I. (2010). On surprise, change, and the effect of recent outcomes. Working paper, Technion, Haifa, Israel.
Peng, F., Jacobs, R.A., Tanner, M.A. (1996). Bayesian inference in mixtures-of-experts and hierarchical mixtures-of-experts models with an application to speech recognition. *Journal of the American Statistical Association, 91*, 953–960.
Rapoport, A., & Amaldoss, W. (2000). Mixed strategies and iterative elimination of strongly dominated strategies: an experimental investigation of states of knowledge. *Journal of Economic Behavior and Organization, 42*, 483–521.
Roth, A.E., & Erev, I. (1995). Learning in extensive-form games: experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior, 8*, 164–212.
Salmon, T. (2001). An evaluation of econometric models of adaptive learning. *Econometrica, 69*, 1597–1628.
Salmon, T. (2004). Evidence for learning to learn behavior in normal form games. *Theory and Decision, 56*(4), 367–404.
Selten, R. (1991). Evolution, learning, and economic behavior. *Journal of Risk Uncertainty, 3*, 3–24.
Stahl, D. (1999). Evidence based rules and learning in symmetric normal form games. *International Journal of Game Theory, 28*, 111–130.
Stahl, D. (2000). Rule learning in symmetric normal-form games: theory and evidence. *Games and Economic Behavior, 32*, 105–138.
Stahl, D. (2001). Population rule learning in symmetric normal-form games: theory and evidence. *Journal of Economic Behavior and Organization, 1304*, 1–17.
Stahl, D. (2003). Mixture models of individual heterogeneity. In *Encyclopedia of cognitive science*. London: MacMillan.
Stahl, D., & Haruvy, E. (2002). Aspiration-based and reciprocity-based rules in learning dynamics for symmetric normal-form games. *Journal of Mathematical Psychology, 46*(5), 531–553.
Thorndike, E.L. (1898). Animal intelligence: an experimental study of the associative processes in animals. In *Psychological review, Monograph supplements* (No. 8). New York: Macmillan.
Van Huyck, J., Battalio, R., Beil, R. (1990). Tacit coordination games, strategic uncertainty, and coordination failure. *American Economic Review, 80*, 234–248.
Van Huyck, J., Cook, J., Battalio, R. (1997). Adaptive behavior and coordination failure. *Journal of Economic Behavior and Organization, 32*, 483–503.
Van Huyck, J., Battalio, R., Rankin, F. (2007). Selection dynamics and adaptive behavior without much information. *Economic Theory, 33*(1), 53–65.
Wilcox, N. (2006). Theories of learning in games and heterogeneity bias. *Econometrica, 74*, 1271–1292.