UNIVERSITY OF Southampton
School of Social Sciences

Economics Division
University of Southampton
Southampton SO17 1BJ, UK

Discussion Papers in
Economics and Econometrics

Title Strategic Learning With Finite Automata Via The EWA-Lite Model

By Christos A. Ioannou and Julian Romero

No. 1109

This paper is revise and resubmit at the Games and Economic Behavior

This paper is available on our website
http://www.southampton.ac.uk/socsci/economics/research/papers

ISSN 0966-4246

Strategic Learning With Finite Automata Via The EWA-Lite Model

Christos A. Ioannou [*†] Julian Romero [‡]

University of Southampton Purdue University

This draft: February 12, 2012

Abstract

We modify the self-tuning Experience Weighted Attraction (EWA-lite) model of Camerer, Ho, and Chong (2007) and use it as a computer testbed to study the probable performance of a set of two-state automata in four symmetric 2×2 games. The model suggested allows for a richer specification of strategies and solves the inference problem of going from histories to beliefs about opponents' strategies, in a manner consistent with "belief-learning". The predictions are then validated with data from experiments with human subjects. Relative to the action reinforcement benchmark model, our modified EWA-lite model can better account for subject-behavior.

[*]The usual disclaimer applies.

[†]Mailing Address: Department of Economics, University of Southampton, Southampton, SO17 1BJ, United Kingdom. Email: c.ioannou@soton.ac.uk

[‡]Mailing Address: Department of Economics, Krannert School of Management, Purdue University, Lafayette, IN 47907. Email: jnromero@purdue.edu

1 Introduction

In their seminal paper, Camerer and Ho (1999) introduced a truly hybridized workhorse of strategic learning, the Experience Weighted Attraction (EWA) model. Despite its originality in combining elements of the fictitious play model and the choice reinforcement model,[1] EWA was criticized for carrying "too" many free parameters. Responding to the criticism, Camerer, Ho, and Chong (2007) replaced two of the free parameters with functions that self-tune. Appropriately labeled, EWA-lite, the self-tuning EWA is econometrically simpler than the prototype, yet still does exceptionally well in a multitude of games where strategies are stage-game actions. More specifically, Camerer, Ho, and Chong (2007) indicate that EWA-lite does as well as the EWA in predicting behavior in seven different games and fits reliably better than the quantal response equilibrium model benchmark. In fact, recently, Chmura, Goerg, and Selten (2011) note that "the good performance of the self-tuning EWA on the individual level is remarkable" (p.25).

Despite its success in predicting behavior, EWA-lite has been constrained by its inability to accommodate for repeated strategies. As Camerer and Ho (1999) acknowledge in their conclusion, the model will "have to be upgraded to cope with ... (a richer) specification of strategies... Incorporating a richer specification of strategies is important because stage-game strategies are not always the most natural candidates for the strategies that players learn about." (p.871) The literature thus far, seems to have been predominantly focussed on action-learning models.[2] Action learning models have limitations. Erev and Roth (1998) for instance, note that it will "not generally be the case that learning behavior can be analyzed in terms of stage-game actions alone" (p.872). Furthermore, McKelvey and Palfrey (2001) argue for the development of strategic learning models in which players learn not about the performance of actions, but rather of repeated strategies (p.19) (see also, Haruvy and Stahl (2002)).

Our objective in this study is to modify the EWA-lite model to, first, accommodate for a richer specification of strategies and, second, to allow for a belief-based learning rooted on players updating their beliefs on the probability distribution of the other players' strategies. Crucially, the

[1] The fictitious play model operates on the premise that agents keep track of the history of previous play by other agents and can thus form beliefs about what others will do in the future based on past observation. Consequently, agents choose a strategy that maximizes their expected payoff given the beliefs they formed. On the other hand, the choice reinforcement model assumes that strategies are "reinforced" by their previous payoff, and that the propensity to choose a strategy depends on its stock of reinforcement.

[2] A notable exception is the study of Hanaki, Sethi, Erev, and Peterhansl (2005) who apply simple reinforcement learning on a set of two-state automata.

modified model nests the model of Camerer, Ho, and Chong (2007). Central to our framework is a key assumption that "allows" an agent to be fully-aware of all possible strategies in the candidate strategy set of the other player(s). In particular, we impose *a priori* complexity constraints on the candidate set of strategies so as to limit the number of potential strategies considered. Furthermore, the strategies of the agents are implemented by a type of finite automaton, called a *Moore machine* (Moore (1956)).

According to the thought experiment, a population of agents is to play a game in *fixed* pairs. The games used here, are of complete information and perfect monitoring where after each move, only the payoff (and not the strategy) of each agent is revealed. An agent is required to choose a strategy. Strategies are chosen based on their attractions. Initially, each of the strategies in an agent's candidate set has an equal attraction and, hence, an equal probability of being chosen. Attractions are updated periodically as the payoffs resulting from strategy-choices are observed. More specifically, at each period, there is a small and constant probability that attractions will be updated and strategy-revision will occur. If strategy-revision does occur, the new strategy is chosen on the basis of the updated attractions. Over the course of this process some strategies decline in use, while others are observed with greater frequency.

The predictions of the modified EWA-lite model are then validated with experimental (human) data from four symmetric 2×2 games and compared to the predictions of a baseline. Relative to the action reinforcement benchmark model, the modified EWA-lite can better account for subject-behavior. The rest of the paper is organized as follows. In Section 2, the related literature is reviewed while in Section 3, we introduce the notation and define the type of finite automaton used. In Section 4, the methodology is explained. In Section 5, the results of the computational simulations are presented. In Section 6, the results are discussed and validated with data from experiments with human subjects. Finally, in the Conclusion, we offer direction for future research.

2 Literature Review

The study is related to several strands of literature. First, it builds on the finite automata literature. Using finite automata as the carriers of agents' strategies was first suggested by Aumann (1981) for the study of decision-making with bounded rationality. The first application originated in the work of Neyman (1985) who investigated a finitely-repeated game model in which the pure strategies available to the agents were those which could be generated by machines utilizing no

more than a certain number of states. Ben-Porath (1990) and Megiddo in collaboration with Widgerson (1985) also pursued this line of enquiry; the latter in the context of Turing machines. Additionally, several researchers have studied the effect of complexity on the set of equilibria in repeated games with finite automata. Abreu and Rubinstein (1988) for one, showed that if agents' preferences are increasing in repeated game payoffs and decreasing in the complexity of the strategies employed, then the set of Nash-equilibrium payoffs that can occur is dramatically reduced from the folk-theorem result of Fudenberg and Maskin (1986). Yet, they indicate that a wide variety of payoffs remains consistent with equilibrium behavior in the presence of complexity costs. It is noteworthy that Abreu and Rubinstein (1988), just like the other papers mentioned, define the complexity of a strategy as the number of states of the minimal automaton implementing it. On the other hand, Banks and Sundaram (1990) argue that the traditional number-of-states measure of complexity neglects some essential features such as informational requirements at a state. They propose instead, a criterion of complexity which takes into account both the size (number of states) and transitional structure of an automaton. Under this proposition, they prove that the resulting Nash equilibria of the game are now trivial: the automata recommend actions every period that are invariably stage-game Nash equilibria.

Second, the study is related to the large literature on learning. There exist many competing models to explain how individuals learn in a repeated game-setting. In belief-based models, players tend to choose strategies that have high expected payoffs given beliefs formed by observing the history of what others did. Some special cases of weighted fictitious play models are fictitious play and Cournot best-response (Cournot 1960).[3] A more general belief model, allowing idiosyncratic shocks in beliefs and time-varying weights was developed by Crawford (1995) to fit data from coordination games. Crawford and Broseta (1998) extended the model to allow ARCH error-terms and applied it to coordination with preplay auctions. Recently, Chmura, Goerg, and Selten (2011) introduced the action-sampling learning model, which is based on a fictitious play process that only considers random periods and not the entire history. Another intriguing model, also, introduced in their study is the impulse-matching learning model. An impulse essentially quantifies the regret of not choosing the best response given what the other player chose. The model begins by transforming the original payoffs (so as to incorporate loss aversion), and then calculates the probabilities of the actions based on impulse sums. An impulse sum aggregates all impulses

[3]Boylan and El-Gamal (1993) compare fictitious play and Cournot learning in coordination and dominance-solvable games; they find overwhelming relative support for fictitious play.

experienced.

Other studies concentrate only on reinforcement learning. Harley (1981) posited a reinforcement model using cumulative payoffs and simulated its behavior in several games. The Harley model was later extended by Roth and Erev (1995) to include spillover of reinforcement to neighboring strategies. Their model fits the time trends in ultimatum, public goods, and responder-competition games but converges too slowly. Finally, Hanaki, Sethi, Erev, and Peterhansl (2005) demonstrate that a simple reinforcement model of learning applied to a set of two-state automata accounts for the behavior of human subjects in the Stag-Hunt game, the Battle of the Sexes, the Prisoner's Dilemma game and the Chicken game; and does so without assuming that fairness and reciprocity are primitive concerns. These studies of belief and reinforcement learning find that each approach, evaluated separately, has some explanatory power. Reinforcement does better in constant-sum games and belief learning in coordination games.

On the other hand, Camerer and Ho (1999) introduced a truly hybridized model of learning, the EWA model, which captures strategic learning by combining elements of, both, weighted fictitious play and reinforcement learning. A limitation of this approach is the fact that agents are myopic in the sense that they do not anticipate the consequences of their strategy on the future use of alternative strategies by their opponents. In a subsequent paper, Camerer and Ho (2002) address this limitation by extending the EWA to capture sophisticated learning and strategic teaching in repeated games. The generalized model assumes there is a mixture of adaptive learners and sophisticated players. A sophisticated player rationally best-responds to his forecasts of all other behaviors. A sophisticated player can be either myopic or farsighted. A farsighted player develops multiple-period rather than single-period forecasts of others behaviors and chooses to "teach" the other players by choosing a strategy scenario that gives him the highest discounted net present value. Camerer and Ho estimate the model using data from p-beauty contests and repeated trust games with incomplete information. The generalized model is better than the adaptive EWA model in describing and predicting behavior. In contrast, the EWA-lite of Camerer, Ho, and Chong (2007) addresses criticisms that 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.

Two unresolved issues of the EWA literature have been the need to incorporate a richer specification of strategies and the inference problem of going from histories to beliefs about opponents' strategies. Incorporating a richer specification of strategies is important because stage-game strate-

gies are not always the most natural candidates for the strategies that players learn about. The open question, therefore, is what rules to specify a priori, and how a model can winnow down a very large set of possible rules as quickly as humans do. The present study addresses these two issues by modifying EWA-lite in such a way so as to allow a richer specification of strategies and a belief-based learning rooted on players updating their beliefs on the probability distribution of the other players' strategies.

3 Preliminaries

3.1 Notation

To simplify exposition we introduce first, the notation for repeated games. The stage-game is represented in normal form, with a set of players $I = \{1, ..., n\}$, an *action set* for each player, A^i, and a real-valued payoff function $g^i : A \to \mathbb{R}$ for each player. The payoff function maps every action profile $(a^i, a^{-i}) \in A$ into a payoff for i, where A denotes the cartesian product of the action spaces A^i, written as $A \equiv \underset{i=1}{\overset{I}{\times}} A^i$. Mixed strategies profiles are denoted by (α^i, α^{-i}) and payoff functions are extended to mixed strategies in the usual way. In a repeated game, the stage-game is played over a sequence of periods, $t = 1, ..., T$. Players share a common discount factor $\beta < 1$, and payoffs in the repeated game are evaluated as the average discounted value. The utility function of player i for action profile a in the repeated game is denoted as $u^i(a)$. Each player i's (pure action) minmax payoff v_p^i is given by

$$v_p^i \equiv \min_{a^{-i}} \max_{a^i} u^i(a^i, a^{-i}) \equiv \max_{a^i} u^i(a^i, \hat{a}_i^{-i}) \equiv u^i(\hat{a}_i),$$

so that $\hat{a}_i$ is an action profile that minmaxes player i. The payoff v_p^i is the lowest payoff that the other players can force on player i in the stage-game (using pure actions). The set of stage-game payoffs generated by pure action profiles is

$$\mathcal{F} \equiv \{v \in \mathbb{R}^n : \exists a \in A \; such \; that \; g(a) = v\},$$

while the set of feasible payoffs is

$$\mathcal{F}^\dagger \equiv co\mathcal{F},$$

where $co\mathcal{F}$ is the convex hull of $\mathcal{F}$. Finally, the set of strictly (pure action) individually rational and feasible payoffs is

$$\mathcal{F}^{+p} \equiv \{v \in \mathcal{F}^{\dagger} : v^i > v_p^i,\ \forall i\}.$$

The interior of the set $\mathcal{F}^{+p}$ is assumed to be non-empty. Consider infinitely-repeated games with *perfect monitoring*. In each period $t = 0, 1, ...,$ the stage-game is played with the action profile chosen in period t publicly observed at the end of that period. The *history* of play at time t is denoted by $h_t = (a_0, ..., a_{t-1}) \in A^t$, where $a_r = (a_r^1, ..., a_r^n)$ denotes the actions taken in period r, and the set of histories is given by

$$\mathcal{H} = \bigcup_{t=0}^{\infty} A^t$$

where we define the initial history to the null set $A^0 = \{\emptyset\}$. A strategy s^i for player i is a function $s^i : \mathcal{H} \to A^i$.

3.2 Finite Automata

Our motivation to use finite automata rests on the desire to reduce the computational burden as well as to reflect elements of bounded rationality and complexity.[4] A finite automaton is a mathematical model of a system with discrete inputs and outputs. The system can be in any one of a finite number of internal configurations or "states". The state of the system summarizes the information concerning past inputs that is needed to determine the behavior of the system on subsequent inputs. The specific type of finite automaton used here is a Moore machine (Moore 1956). A *Moore machine* for an adaptive agent i, M^i, in a repeated game $G = (N, \{A^i\}_{i \in I}, \{g^i\}_{i \in I})$ is a four-tuple $(Q^i,\ q_0^i,\ f^i,\ \tau^i)$ where Q^i is a finite set of internal states of which q_0^i is specified to be the initial state, $f^i : Q^i \to A^i$ is an output function that assigns an action to every state, and $\tau^i : Q^i \times A^{-i} \to Q^i$ is the transition function that assigns a state to every two-tuple of state and other agent's action. It is pertinent to note that the transition function depends only on the present state and the other agent's action. This formalization fits the natural description of a strategy as agent i's plan of action in all possible circumstances that are consistent with agent i's

[4]Bounded rationality suggests that a player may not consider all feasible strategies but instead limits himself to "less complex" strategies. The complexity of finite automata may be defined in a number of ways (a detailed exposition is provided in the Appendix).

plans. In contrast, the notion of a game-theoretic strategy for agent i requires the specification of an action for every possible history, including those that are inconsistent with agent i's plan of action.[5]

In the first period, the state is q_0^i and the automaton chooses the action $f^i(q_0^i)$. If a^{-i} is the action chosen by the other agent in the first period, then the state of agent i's automaton changes to $\tau^i(q_0^i, a^{-i})$, and in the second period agent i chooses the action dictated by f^i in that state. Then, the state changes again according to the transition function given the other agent's action. Thus, whenever the automaton is in some state q, it chooses the action $f^i(q)$ while the transition function τ^i specifies the automaton's transition from q (to a state) in response to the action taken by the other agent.

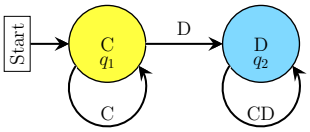


Figure 1: Grim-Trigger

$Q^i = \{q_1, q_2\}$
$q_0^i = q_1$
$f^i(q_1) = C$ and $f^i(q_2) = D$
$\tau^i(q, a^{-i}) = \begin{cases} q_1 & (q,a^{-i})=(q_1,C) \\ q_2 & otherwise \end{cases}$

For example the automaton (Q^i, q_0^i, f^i, τ^i) in Figure 1, carries out the Grim-Trigger strategy in the context of the Prisoner's Dilemma game. Thus, the strategy chooses C so long as both agents have chosen C in every period in the past, and chooses D otherwise. In the transition diagram, the vertices denote the internal states of the automaton and the arcs labeled with the action of the other agent, indicate the transition to the states.

[5]To formulate the game-theoretic strategy, one would have to construct the transition function such that $\tau^i : Q^i \times A \to Q^i$ instead of $\tau^i : Q^i \times A^{-i} \to Q^i$. The exposition of the model with the use of the game-theoretic formulation is provided in the Appendix.

4 Methodology

Players have propensities or attractions associated with each of their strategies and these attractions determine the probabilities with which strategies are chosen when players experiment. Initially, all strategies have an equal attraction and hence an equal probability of being chosen. The learning process evolves through the attractions of strategies. It is thus pertinent that players play the stage-game a number of times before evaluating their current strategy. If the probability of switching strategies is not too large, meaningful evaluations of repeated-game strategies are possible. Let us denote the strategies of player i by $s^i \in S^i$, and a strategy combination of the n players except i by $s^{-i} = (s^1, ..., s^{i-1}, s^{i+1}, ..., s^n)$. Finally player i's payoff in period t is denoted as $\pi(s^i(t), s^{-i}(t))$.

The modified EWA-lite model consists of two variables which are updated once an agent switches strategies.[6] Crucially, we will assume that players update their strategies simultaneously. The first variable is $N(t)$ and is interpreted as the number of observation-equivalents of past experience in t periods. The second variable, denoted as $A^i_j(\chi)$, indicates player i's attraction of strategy j *after* cluster χ.[7] The length of a cluster can either be determined deterministically or randomly. For simplicity, in this exposition, we assume that the cluster length is fixed. The variables $N(t)$ and $A^i_j(\chi)$ begin with some prior values, $N(0)$ and $A^i_j(0)$. These prior values can be though of as reflecting pre-game experience, either due to learning transferred from different games or due to introspection. The evolution of learning over a cluster χ with $\chi \geq 1$ is governed by the following rule:

$$A^i_j(\chi) = \frac{\phi \cdot N(t-1) \cdot A^i_j(\chi - 1) + (1-\delta) \cdot I(s^i_j, s^i(\chi)) \cdot R^i_j(\chi) + \delta \cdot E^i_j(\chi)}{\phi \cdot N(t-1) + 1}. \tag{1}$$

The rule incorporates the decay rate function and the attention function. The decay rate function $\phi(\cdot)$ weighs lagged attractions. It reflects a combination of "forgetting" and "motion detection". The latter refers to a detection of change in the environment by the player; that is, when a player senses that other players are changing, a self-tuning $\phi^i(\cdot)$ decreases so as to allocate less weight to the distant past. The core of the $\phi^i(\cdot)$ is a "surprise index", which indicates the difference between other player's most recent strategies and their strategies in the previous

[6] For more detailed information on EWA, please see Camerer and Ho (1999).

[7] The timing of the process is a matter of taste. The results hold whether the variables are updated at the start or the end of the period subject to the appropriate changes in notation.

clusters. First, define

$$H_k^i(s_k^{-i}, h(\chi)) = \begin{cases} \frac{1}{\alpha} & s_k^{-i} \text{ is consistent with } h(\chi) \\ 0 & \text{else,} \end{cases} \tag{2}$$

where $h(\chi)$ denotes the history of action profiles over the χ cluster of periods. The α in the denominator, indicates the number of other player's strategies that are consistent with the history $h(\chi)$. Thus, α secures an equal weight across all other player's strategies which satisfy the history of action profiles. If, for example, there are two strategies that satisfy the history profile, then, the two strategies receive a weight of $\frac{1}{2}$ each. In addition, σ is a cumulative history vector across the other players' strategies k,

$$\sigma_k^i(\Xi) = \sum_{\epsilon=1}^{\Xi} \frac{H_k^i(s_k^{-i}, h(\epsilon))}{\Xi}, \tag{3}$$

which records the historical frequencies of the strategy-choices of the other players over the Ξ clusters.[8] Furthermore, the immediate "history" vector element $r_k^i(\Xi)$ indicates the weights placed across the other players' strategies k in the last cluster Ξ. That is,

$$r_k^i(\Xi) = H_k^i(s_k^{-i}, h(\Xi)) = \begin{cases} \frac{1}{\alpha} & s_k^{-i} \text{ is consistent with } h(\Xi) \\ 0 & \text{else.} \end{cases} \tag{4}$$

The surprise index $S^i(\Xi)$ simply sums up the squared deviations between the cumulative history vector $\sigma_k^i(\Xi)$ and the immediate history vector $r_k^i(\Xi)$; that is,

$$S^i(\Xi) = \sum_{k=1} (\sigma_k^i(\Xi) - r_k^i(\Xi))^2. \tag{5}$$

In other words, the surprise index captures the degree of change of the most recent choice from the historical average. Note that it varies from zero (when there is strategy-persistence) to two (when the other player chose a particular strategy "forever" and suddenly switches to something brand new). Thus, when the surprise index is zero, we have a stationary environment; when it is one, we have a turbulent environment. The change-detecting decay rate of cluster Ξ is then

$$\phi^i(\Xi) = 1 - \frac{1}{2} S^i(\Xi). \tag{6}$$

On the other hand, the attention function $\delta(\cdot)$ is the weight placed on foregone payoffs. Presumably this is tied to the attention subjects pay to alternative payoffs, ex post. Subjects with

[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.

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. To capture this property, define

$$\delta_j^i(\chi) = \begin{cases} 1 & \text{if } E_j^i(\chi) \geq R_j^i(\chi) \\ 0 & \text{otherwise,} \end{cases} \tag{7}$$

where $R_j^i(\chi)$ denotes the reinforcement value of the strategy j used over the periods $t-\tilde{\tau}+1, ..., t$ in some cluster χ. It is defined as the average payoff obtained by player i over the cluster χ of $\tilde{\tau}$ periods

$$R_j^i(\chi) = \frac{1}{\tilde{\tau}} \sum_{r=t-\tilde{\tau}+1}^{t} \pi^i(r), \tag{8}$$

where $\pi^i(r)$ is the payoff obtained by player i in period r. On the other hand, the expected payoff of cluster χ is taken over beliefs according to

$$E_j^i(\chi) = \sum_{s_k^{-i} \in S^{-i}} \pi(s_j^i, s_k^{-i}) \cdot p_k\left(B^{-i}(\chi)\right). \tag{9}$$

The value placed on other player's strategy k in cluster χ is defined by

$$B_k^{-i}(\chi) = \begin{cases} 1 & s_k^{-i} \text{ is consistent with } h(\chi) \\ 0 & \text{else,} \end{cases} \tag{10}$$

where $h(\chi)$ denotes the history of action profiles in cluster χ. Finally, the probability placed on other player's strategy k in cluster χ is given by

$$p_k(B^{-i}(\chi)) = \frac{B_k^{-i}(\chi))}{\sum_l B_l^{-i}(\chi)}. \tag{11}$$

The attention function $\delta(\cdot)$ enables subjects to reinforce chosen strategies and all unchosen strategies with (weakly) better payoffs by a weight of one. In contrast, unchosen strategies with strictly worse payoffs are not reinforced. Thus, the attention function captures the idea of "learning direction" theory of Selten and Stoecker (1986), whereby subjects have a tendency to move into the direction of the strategy which was ex-post the best response. This is done by shifting the attention and, consequently, the probability towards the strategy with the highest payoff.

Attractions, on the other hand, determine probabilities of choosing strategies. To specify the choice probability of strategy j we use the logit specification. Thus, the probability of a player

i choosing strategy j, when he updates his strategy at the end of cluster χ, depends on the attractions so that[9]

$$\mathbb{P}_j^i(\chi) = \frac{e^{\gamma A_j^i(\chi)}}{\sum_k e^{\gamma A_k^i(\chi)}}. \tag{12}$$

The parameter $\gamma \geq 0$ in the logistic transformation measures the sensitivity of players to attractions. Thus, if $\gamma = 0$, all strategies are equally likely to be chosen regardless of their attractions. As γ increases, strategies with higher attractions become disproportionately more likely to be chosen. In the limiting case where $\gamma \to \infty$, the strategy with the highest attraction is chosen with probability one.

5 Results

We have limited our attention to the four symmetric 2×2 games for which results are reported in both Arifovic, McKelvey, and Pevnitskaya (2006) and Hanaki, Sethi, Erev, and Peterhansl (2005); namely, the Prisoner's Dilemma, the Battle of the Sexes, the Stag Hunt, and the Chicken. The payoff matrices are illustrated in Figure 2. To test the modified EWA-lite model in these games, we run computer simulations. Each simulation consists of a fixed pair of agents that stay matched for 1000 periods. The agents are able to choose among the set of two-state automata depicted in Figure 3. At the beginning of the simulations, each agent is endowed with initial attractions $A_j^i(0) = 1.5$ (as specified in Camerer, Ho, and Chong (2007)) for each strategy j in S^i. Players are also endowed with initial experience $N^i(0) = 1$. Players update their attractions at the end of each cluster which consists of 20 periods[10] and then simultaneously update their strategies based

[9] To avoid the bias of the estimated parameter γ, the probabilities of the attractions could be calculated as

$$\mathbb{P}_j^i(\chi) = \frac{A_j^i(\chi)}{\sum_k A_k^i(\chi)}.$$

[10] Whenever players use finite-state automata, a cycle of action-pairs is eventually attained, though it may not necessarily start at period 1. In fact, in this exercise, our upper bound of two states ensures that the cycle will be attained by period 5. One may then ask why not fix the cluster length at 5 periods. The mere fact that a cycle may not start at period 1 can bias the payoffs in favor of the automaton that did better in the first period. Allowing for a cluster length of 20 periods, discounts the payoff earned in the first period and puts more weight to the payoff earned over the cycle.

	A	B
A	8,8	1,9
B	9,1	2,2

(a) Prisoners' Dilemma

	A	B
A	3,3	18,6
B	6,18	3,3

(b) Battle of the Sexes

	A	B
A	3,3	3,1
B	1,3	6,6

(c) Stag Hunt

	A	B
A	5,5	2,6
B	6,2	1,1

(d) Chicken

Figure 2: Payoff Tables

on their new attractions. All simulations presented here use an intensity parameter $\gamma = 5$ in the logit specification.[11] The results displayed in the plots are averages taken over 500 simulated pairs.

5.1 Simulations

Figure 4 displays the results of the simulations in the Prisoners' Dillemma payoff matrix (see Figure 2(a)). The cooperative action is denoted with letter A, whereas the action of defection is denoted with letter B. Each player's dominant strategy is to play B. Figure 4(a) shows the frequency of automaton pairs played over the course of the repeated game. Thus, the larger the area of the bubble, the bigger the frequency of play. The most common outcomes are for both players to play Automaton 10, or, for both players to play Automaton 12. Automaton 10 is also referred to as the "Win-Stay, Lose-Shift" strategy, and has been found to outperform "Tit-for-Tat" in the Prisoner's Dilemma in the evolutionary simulations of Nowak and Sigmund (1995). Automaton 12 is the "Grim-Trigger" strategy. Figure 4(b) shows the progressions of probabilities

[11] We have experimented with $\gamma \in \{1, 2, 3, 5, 10\}$. Our results are insensitive to reasonable changes in these values.

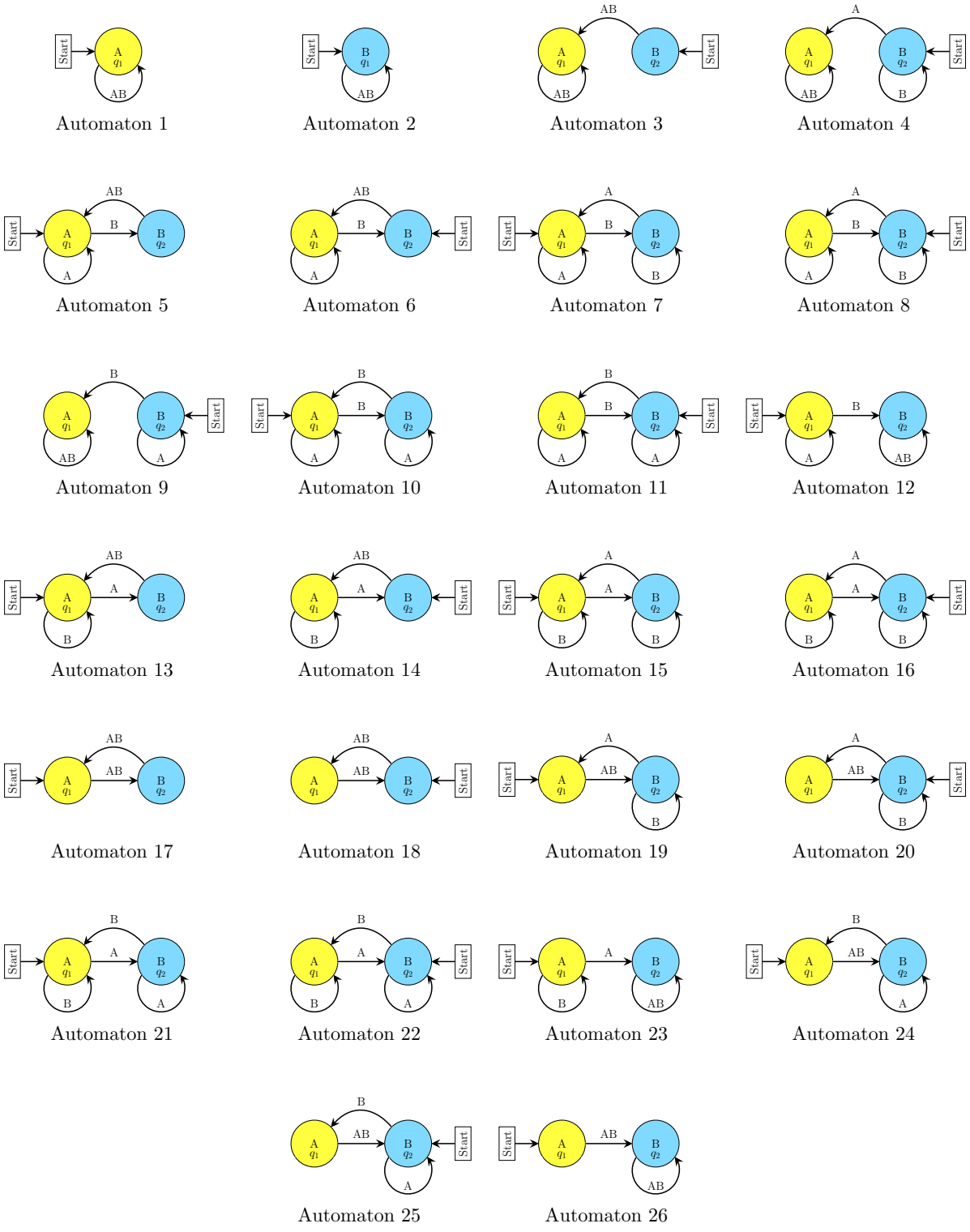
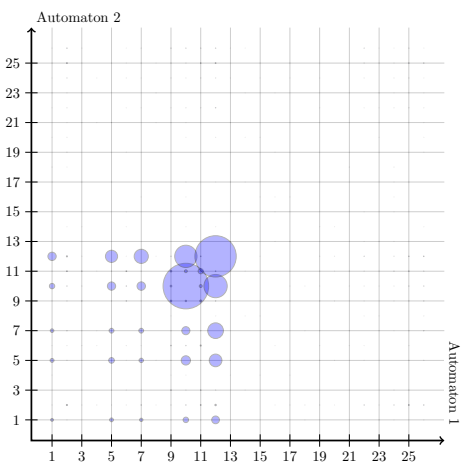
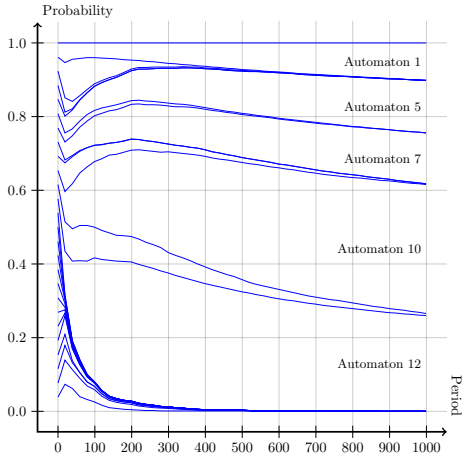


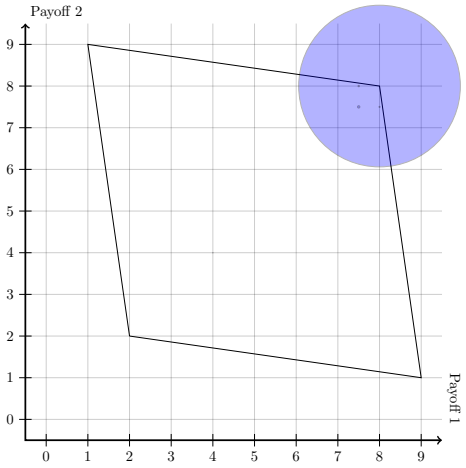
Figure 3: Two-State Automata



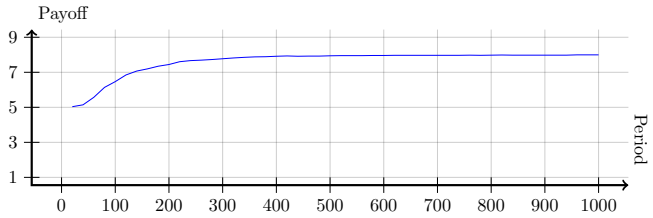
(a) Automaton Frequency



(b) Attraction Probability Progression



(c) Payoff Frequency

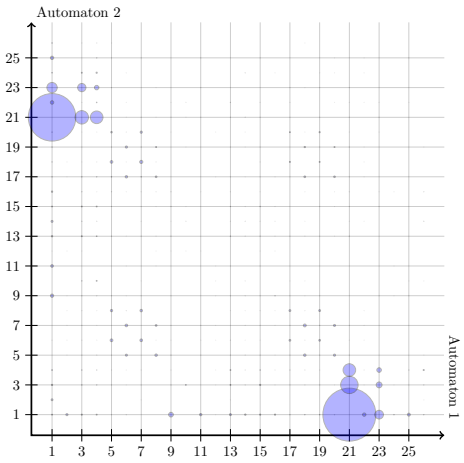


(d) Average Payoff Progression

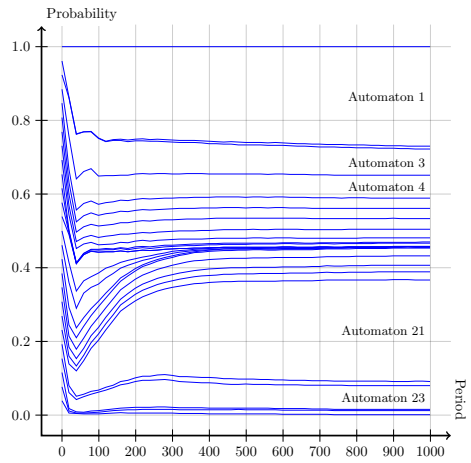
Figure 4: Prisoners' Dilemma

(determined from attractions). That is, the difference between two successive curves indicates the likelihood of an automaton being chosen. This plot suggests that in the Prisoner's Dilemma, towards the end of the 1000 periods, only five strategies are being chosen: Automata $1, 5, 7, 10$, and 12. It is important to note that in any pair-combination between these five automata the cooperative outcome (A, A) is sustained, which rewards each player with a payoff of eight. Figure 4(c) shows the set of feasible repeated game payoffs, and the frequency of each payoff combination over the final 200 periods of the 1000 period interaction. The area of the bubbe denotes the frequency of play. This plot shows that essentially all players are cooperating over the final 200 periods of the interaction. Finally, Figure 4(d) shows the progression of payoffs over the course of the interaction. The average payoff in the beginning is five, which is the payoff that would be obtained if both players are randomizing. Payoffs then quickly increase towards eight, which is the cooperative payoff outcome.

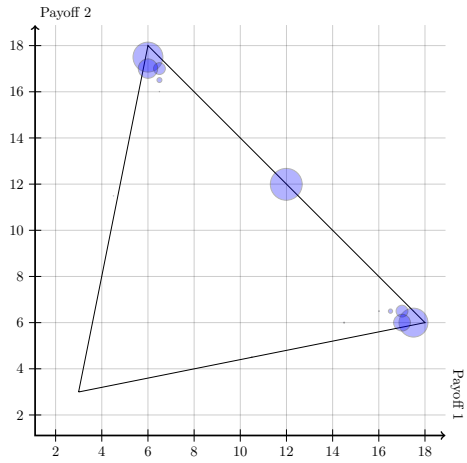
Figure 5 shows the results of the simulations in the Battle of the Sexes (see Figure 2(b)). In this game, there are two pure-strategy equilibria: (A, B) and (B, A). Each player prefers the equilibrium in which he plays A and their opponent plays B. Figure 5(b) suggests that the



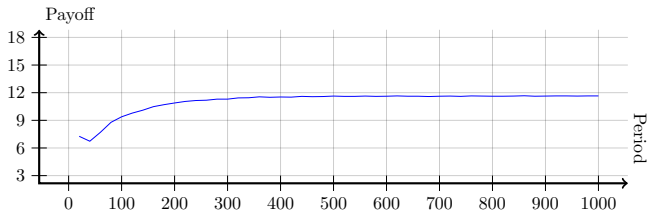
(a) Automaton Frequency



(b) Attraction Probability Progression



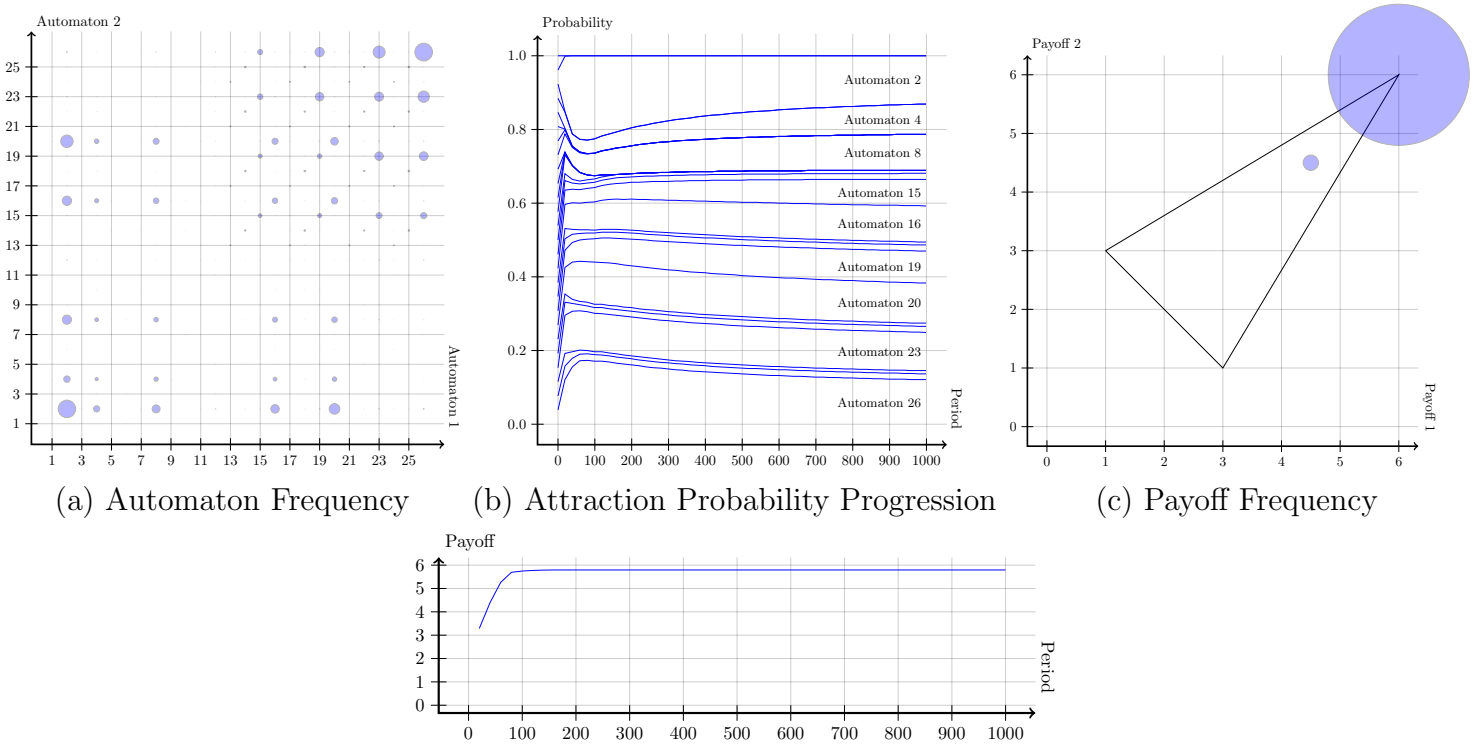
(c) Payoff Frequency



(d) Average Payoff Progression

Figure 5: Battle of the Sexes

two strategies that are most likely to be played are Automaton 1 (which always plays A) and Automaton 21. Automaton 21 switches actions only if the opponent played the same action in the previous period; otherwise, continues with the same action. This automaton is beneficial because players always want to play a different action than their opponent. Figure 5(a) suggests that when one player is playing Automaton 1 the other player is playing Automaton 21. When the players play this pair of automata, it will lead to one of the pure strategy equilibrium yielding a payoff of either $(6, 18)$ or $(18, 6)$. Looking at Figure 5(c), we see mass points at each of these payoffs. In addition, we also observe a mass point at $(12, 12)$. This latter mass point corresponds to the situation where players are alternating between the two pure strategy equilibria. This behavior has been observed experimentally (McKelvey and Palfrey (2001)), and would be impossible to obtain using a model that only allows action-learning. Please notice that in spite of observing the $(12, 12)$ in Figure 5(c), there is no corresponding mass point identified in Figure 5(a) because there are many combinations of automata that lead to alternations. For example, if one player plays Automaton 5 and the other player plays Automaton 6, there would be alternating between the pure strategy equilibria. All in all, there are 32 combinations of automata that lead to these



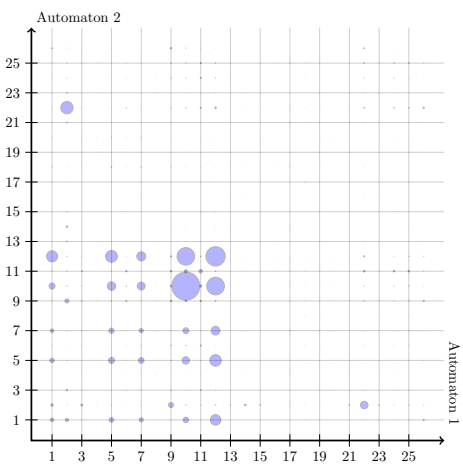
(a) Automaton Frequency (b) Attraction Probability Progression (c) Payoff Frequency

(d) Average Payoff Progression

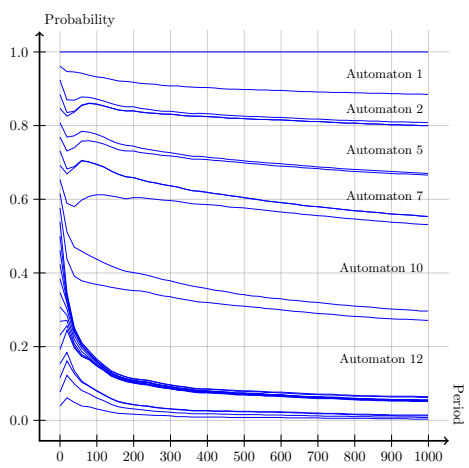
Figure 6: Stag Hunt

alternations. So even though no single automaton that leads to alternations can be identified, the combined impact of all of these combinations leads to a significant amount of alternations represented by the mass at $(12, 12)$ in Figure 5(c). Figure 5(d) shows that the average payoff per player converges to 12, which is the average payoff per player in each of the three mass points of Figure 5(c).

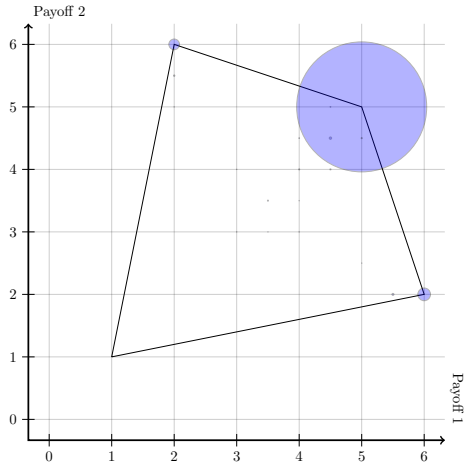
Figure 6 shows the results of the simulations in the Stag-Hunt game (see Figure 2(c)). In this game there are two pure-strategy Nash equilibria, (A, A) and (B, B); however, (B, B) is the Pareto dominant equilibrium. Figure 6(a) and Figure (b) suggest there is weak convergence to a small set of automata. Figure 6(c), on the other hand, suggests that there is convergence to a specific payoff combination; the Pareto dominant Nash equilibrium payoff of $(6, 6)$. Unlike the Prisoners' Dilemma in which agents play strategies that punish defectors, the stag-hunt game is a coordination game with aligned interests. Therefore, there are many combinations of automata that lead to coordination on the Pareto dominant equilibrium which explains the weak convergence to a small set of automata. Figure 6(d) suggests that average payoff converges to six quickly.



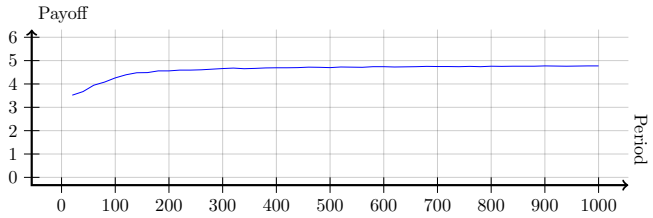
(a) Automaton Frequency



(b) Attraction Probability Progression



(c) Payoff Frequency



(d) Average Payoff Progression

Figure 7: Chicken

Figure 7 shows the results of the simulations in the Chicken game (see Figure 2(d)). In this game there are two pure-strategy Nash equilibria: (A, B) and (B, A). Recall that in the chicken game, the cooperative outcome of (A, A), yields higher payoffs for each of the players than the average payoff when alternating between the pure-strategy Nash equilibria. The results in Figure 7(a) and Figure 7(c) look similar to those in the Prisoners' Dillemma, and show that play is converging to a small set of automata that lead to the cooperative outcome of (A, A). This observation is confirmed in Figure 7(b) which displays that a large percentage of simulations end with payoffs corresponding to the cooperative outcome. However, there are also a small number of simulations which end at the two pure-strategy Nash equilibrium payoffs.

6 Discussion

One of the benefits of adding the belief component on top of the reinforcement component is that it increases the speed of convergence. In order for a given strategy's attraction to be updated in a reinforcement learning model, the strategy must be played first. When beliefs are added to

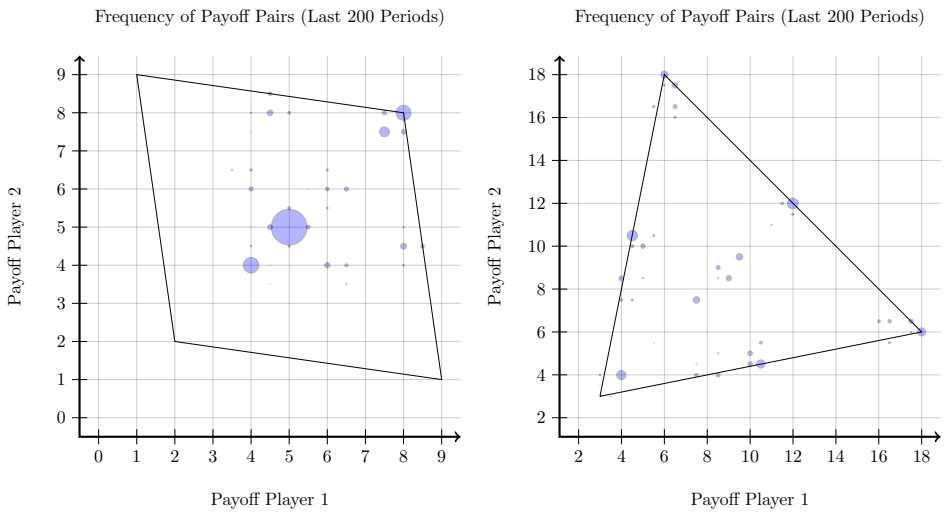
the model, the attractions for every strategy are updated at the end of every cluster. Therefore, the attraction on a strong strategy can start to increase with the first attraction-update; in contrast, with only the reinforcement component, a strong strategy remains unaffected in terms of attraction-weights until it gets selected. The timelines of attraction probabilities and average payoffs in Figures 4-7 parts (b) and (d) suggest that the convergence in these simulations is indeed relatively fast. These timelines show that the simulations do not change drastically after the first 200 or 300 periods. Since the cluster length is 20 periods, this suggests that the simulations are converging within the first 10-15 attraction updates.

A natural benchmark to compare the modified EWA-lite model's predictions is the basic reinforcement learning model. In the specific reinforcement learning model players simultaneously update there strategies after clusters of fixed length. In addition, the players' attractions to strategies are updated using the following rule:

$$A_j^i(\chi) = \begin{cases} \omega A_j^i(\chi-1) + (1-\omega) R^i(\chi-1) & \text{if strategy } j \text{ is played} \\ A_j^i(\chi-1) & \text{else} \end{cases}$$

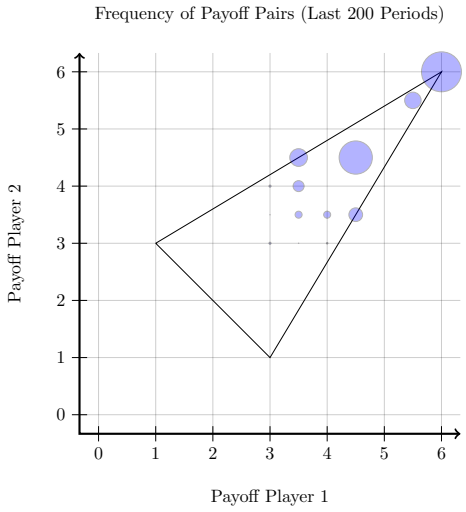
where $R^i(\chi-1)$ is the reinforcement payoff for the $(\chi-1)^{\text{st}}$ cluster. The agents choose strategies using a logistic function similar to that in (12) with parameter γ. When simulations are run with this rule, the attractions are updated much slower than the modified EWA-lite model. The simulations of the four games studied here are provided in Figure 8. Please notice that even after 10,000 periods, the reinforcement learning model is unable to provide sharp predictions.

Ultimately, one would like to validate the predictions of the modified EWA-lite model with evidence from experiments with human data. Arifovic, McKelvey, and Pevnitskaya (2006) provide experimental data on the four games reported in Section 5. The data consists of 24 pairs of subjects playing a game with a fixed opponent for 50 periods. Some common trends emerge from the data that are typically difficult to come by with less sophisticated models such as Fictitious Play. Some of these trends include convergence to the cooperative outcome in the Prisoners' Dilemma game as well as alternations between the two pure-strategy Nash equilibrium in the Battle of the Sexes game. Both of these behaviors emerge from the simulations with the modified self-tuning EWA. Essentially all of the agents playing the Prisoner's Dilemma attain the cooperative outcome (A, A) as displayed in Figure 4(c). In the Battle of the Sexes, a significant amount of the agents converge to strategies that alternate between the equilibria, while the remainder converge to one of the pure-strategy equilibria (see Figure 5(c)). Action learning models have a difficult time obtaining convergence to strategies that alternate between the pure-strategy equilibria in the Battle of the

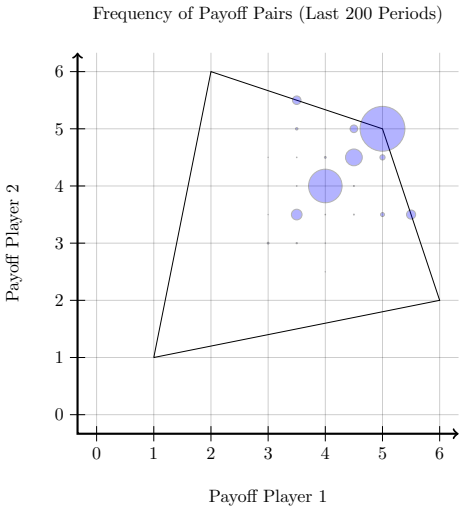


(a) Prisoners' Dilemma

(b) Battle of the Sexes



(c) Stag Hunt



(d) Chicken

Figure 8: The baseline is the reinforcement learning model: the simulations use a $\lambda = 5$ for 10,000 periods and averaged over 500 simulations.

Sexes. The simulations also match the data from the Stag Hunt and Chicken games. Overall, the simulations match the data from experimental outcomes quite well, however further examination of this problem is essential, which would likely include collecting more data on these repeated interactions.

7 Conclusion

The enterprise of finding out what strategies subjects actually choose has not progressed to the point that one would hope. As a first step in that direction, we propose a modification of the EWA-lite model to accommodate for a richer specification of strategies, in a manner consistent with belief-learning. Crucially, the modified model nests the model of Camerer, Ho, and Chong (2007). Our framework makes no a priori assumptions on fairness or reciprocity as primitive concerns, but follows an entirely orthodox process of learning in which material payoffs are the driving force. The predictions of the modified model are validated with data from experiments with human subjects across four symmetric 2×2 games: the Prisoner's Dilemma, the Battle of the Sexes, the Stag Hunt, and the Chicken. Relative to the action reinforcement benchmark model, the modified EWA-lite can better account for subject-behavior.

Ultimately, one would like to test the strategy-predictions of this study in the field, or more likely, in an experimental lab. After all, experiments contribute diagnostic evidence on the generalizability of results. Thus, one could assume a priori that the subjects can choose an automaton out of a list of possible two-state automata. Their automaton-choice could be revised periodically after being informed on the payoffs and the action profiles in the last cluster of periods. In this way, evidence from the lab could be reconciled with the predictions of the specific study.

Another important direction for further research would be to study the feasibility of our approach in settings of greater complexity, with a larger set of players and strategies. Our proposed framework accommodates for such flexibility as it allows admissibility of strategies of greater complexity. A potential empirical extension is an analysis of the goodness-of-fit of the model to the large and varied experimental data that is available. This would require estimation of the model parameters and out-of-sample comparisons with other learning models.

In addition, one would like to test the susceptibility of the results to small amounts of errors. In this study, it was assumed that the agents' strategies were implemented by error-free automata. Agents, in real life, engage in actions that are constrained by the limitations of human nature and

the surrounding environment. Thus, oftentimes agents suffer from a measure of uncertainty about their own as well as their colleagues' actions. In large and complex firms for example, division chiefs are often physically removed from each other and are consequently unable to observe each other's behavior directly. Moreover, division chiefs are prone to errors in the implementation of their own actions (along the lines of Selten's trembling hand). Due to these disturbances, the decision-makers may occasionally draw incorrect inferences about their peers' actions.

Furthermore, it would be interesting to examine whether the results are robust to the symmetry of the payoffs. One of the basic features of the games is the requirement that the values assigned to the game are the same for both agents. Not uncommon however, are social transactions where not only is each agent's outcome dependent upon the choices of the other, but also where the resources and therefore possible rewards, of one agent exceed those of the other. A social interaction characterized by a disparity in resources and potentially larger rewards for one of the two participants would in all likelihood call into play questions of inequality. Thus, one could run two co-evolving populations with asymmetric payoffs, to see how inequality comes into play and in particular, how the asymmetry in payoffs affects "cooperative" behavior.

References

Abreu, Dilip, and Ariel Rubinstein. "The Structure of Nash Equilibrium in Repeated Games with Finite Automata." *Econometrica* 56: (1988) 1259–82.

Arifovic, Jasmina, Richard McKelvey, and Svetlana Pevnitskaya. "An Initial Implementation of the Turing Tournament to Learning in Repeated Two Person Games." *Games and Economic Behavior* 57: (2006) 93–122.

Aumann, Robert. "Survey of Repeated Games." In *Essays in Game Theory and Mathematical Economics in Honor of Oscar Morgestern*. Manheim: Bibliographisches Institut, 1981.

Banks, Jeffrey S., and Rangarajan K. Sundaram. "Repeated Games, Finite Automata, and Complexity." *Games and Economic Behavior* 2: (1990) 97–117.

Ben-Porath, Elchanan. "The Complexity of Computing Best Response Automata in Repeated Games with Mixed Strategies." *Games and Economic Behavior* 2: (1990) 2–12.

Boylan, Richard T., and Mahmoud A. El-Gamal. "Fictitious Play: A Statistical Study of Multiple Economic Experiments." *Games and Economic Behavior* 5: (1993) 205–22.

Camerer, Colin F., and Teck-Hua Ho. "Experience Weighted Attraction Learning in Normal Form Games." *Econometrica* 67: (1999) 827–63.

———. "Sophisticated EWA Learning and Strategic Teaching in Repeated Games." *Journal of Economic Theory* 104, 1: (2002) 137–88.

Camerer, Colin F., Teck-Hua Ho, and Juin-Kuan Chong. "Self-tuning Experience Weighted Attraction Learning in Games." *Journal of Economic Theory* 133: (2007) 177–98.

Chmura, Thorste, Sebastian Goerg, and Reinhard Selten. "Learning in Experimental 2X2 Games.", 2011. Mimeo.

Crawford, Vincent, and Bruno Broseta. "What Price Coordination? The Efficiency-Enhancing Effect of Auctioning the Right to Play." *American Economic Review* 88: (1998) 198–225.

Crawford, Vincent P. "Adaptive Dynamics in Coordination Games." *Econometrica* 63: (1995) 103–43.

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

Fudenberg, Drew, and Eric Maskin. "The Folk Theorem in Repeated Games with Discounting and with Incomplete Information." *Econometrica* 54: (1986) 533–554.

Hanaki, Nobuyuki, Rajiv Sethi, Ido Erev, and Alexander Peterhansl. "Learning Strategies." *Journal of Economic Behavior and Organization* 56: (2005) 523–42.

Harley, Calvin B. "Learning the Evolutionary Stable Strategies." *Journal of Theoretical Biology* 89: (1981) 611–33.

Haruvy, Ernan, and Dale O Stahl. "Aspiration-based and Reciprocity-based Rules in Learning Dynamics for Normal-Form Games." *Journal of Mathematical Psychology* 46: (2002) 531–53.

McKelvey, Richard, and Thomas R. Palfrey. "Playing in the Dark: Information, Learning, and Coordination in Repeated Games.", 2001. Mimeo.

Moore, Edward F. "Gedanken Experiments on Sequential Machines." *Annals of Mathematical Studies* 34: (1956) 129–53.

Neyman, Abraham. "Bounded Complexity Justifies Cooperation in the Finitely Repeated Prisoner's Dilemma." *Economic Letters* 19: (1985) 227–229.

Nowak, Martin A., and Karl Sigmund. "A Strategy of Win-Stay, Lose-Shift that Outperforms Tit-for-Tat in Prisoner's Dilemma." *Nature* 364: (1995) 56–8.

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

Selten, Reinhard, and Rolf Stoecker. "End Behavior in Sequences of Finite Prisoner's Dilemma Supergames." *Journal of Economic Behavior and Organization* 7: (1986) 47–70.

Appendix

Remark. The modified self-tuning EWA model nests the self-tuning EWA model of Camerer, Ho, and Chong (2007).

Proof. This proof examines a special case of the modified EWA model. Assume:

1. $S^1 = S^2 = \{\text{Automaton 1}, \text{Automaton 2}\}$ in Figure 3, and
2. $\nu = 1$, so that all clusters have a length of one.

Given these assumptions, the modified self-tuning EWA model is equivalent to the self-tuning EWA model of Camerer, Ho, and Chong (2007). To prove this, we show that in this special case the terms ϕ, δ, R_j^i, and E_j^i from the modified self-tuning EWA model are all equivalent to the corresponding terms in the action-based self-tuning EWA model presented in Camerer, Ho, and Chong (2007).

Since the clusters are assumed to all have length one, they are referred to as periods and denoted with t rather than χ. Let $a^i(t)$ be the action chosen by player i in period t, and $a(t)$ be the action profile in period t. Notice that since there are only two strategies, the strategy can be inferred from a.

Let us first derive the decay function $\phi(\cdot)$. Since there are only two strategies, we can rewrite the history vector from (2) as

$$H_k^i(s_k^{-i}, a(t)) = \begin{cases} 1 & s_k^{-i} \text{ is consistent with } a(t) \\ 0 & \text{else} \end{cases} = \begin{cases} 1 & s_k^{-i} = s^{-i}(t) \\ 0 & \text{else} \end{cases}.$$

The cumulative history vector element $\sigma_k^i(t)$ across the other players' strategies k at time t is therefore given by

$$\sigma_k^i(t) = \sum_{\epsilon=1}^{t} \frac{H_k^i(s_k^{-i}, a(\epsilon))}{t} = \sum_{\epsilon=1}^{t} \frac{I\left(s_k^{-i}, s^{-i}(t)\right)}{t}.$$

The most recent history vector element $r_k^i(t)$ is thus $r_k^i(t) = H_k^i(s_k^{-i}, a^{-i}(t))$. The surprise index $S^i(t)$ simply sums up the squared deviations between the cumulative history vector and the immediate history vector; that is,

$$S^i(t) = \sum_{k=1} (\sigma_k^i(t) - r_k^i(t))^2.$$

Finally, the change-detector function is

$$\phi^i(t) = 1 - \frac{1}{2} S^i(t). \tag{13}$$

In this special case, this corresponding term is equivalent to the term in equation (3) in Camerer, Ho, and Chong (2007).

Let us derive next the attention function $\delta(\cdot)$. In this case, since the cluster length is one, we can rewrite (8) as

$$R_j^i(t) = \frac{1}{\tilde{\tau}} \sum_{r=t-\tilde{\tau}+1}^{t} \pi^i(r) = \sum_{r=t}^{t} \pi^i(r) = \pi^i(t). \tag{14}$$

Next, we show that $E_j^i(t) = \pi(s_j^i, s^{-i}(t))$. Since there are only two strategies in S^i, the beliefs for the previous period are unambiguous, hence we can rewrite (10) as

$$B_k^{-i}(t) = \begin{cases} 1 & s_k^{-i} \text{ is consistent with } a(t) \\ 0 & \text{else} \end{cases} = \begin{cases} 1 & s_k^{-i} = s^{-i}(t) \\ 0 & \text{else} \end{cases}.$$

This implies that,

$$p_k(B^{-i}(t)) = \frac{B_k^{-i}(t))}{\sum_l B_l^{-i}(t)} = \begin{cases} 1 & s_k^{-i} = s^{-i}(t) \\ 0 & \text{else} \end{cases}$$

hence, we get that

$$E_j^i(t) = \sum_{s_k^{-i} \in S^{-i}} \pi(s_j^i, s_k^{-i}) \cdot p_k\left(B^{-i}(t)\right) = \pi(s_j^i, s^{-i}(t)). \tag{15}$$

From equations (14) and (15) we get that the attention function is,

$$\delta_j^i(t) = \begin{cases} 1 & \text{if } \pi^i(s_j^i, s^{-i}(t)) \geq \pi^i(t) \\ 0 & \text{else} \end{cases} \tag{16}$$

which is equivalent to the term in equation (4) in Camerer, Ho, and Chong (2007).

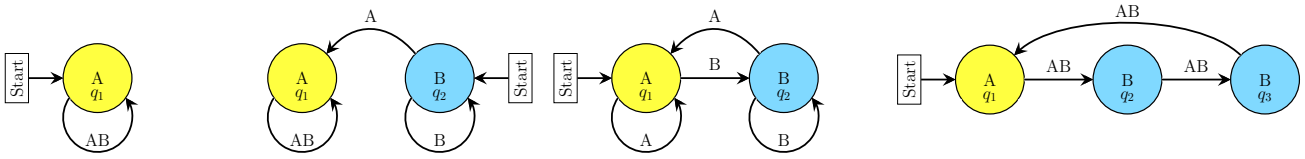
Finally, combining equations (1), (13), (14), (15), and (16), we get

$$\begin{aligned} A_j^i(t) &= \frac{\phi \cdot N(t-1) \cdot A_j^i(t-\tilde{\tau}) + \delta \cdot E_j^i + (1-\delta) \cdot I(s_j^i, s^i(t)) \cdot R_j^i(t)}{\phi \cdot N(t-1) + 1} \\ &= \frac{\phi \cdot N(t-1) \cdot A_j^i(t-1) + \delta \cdot \pi(s_j^i, s^{-i}(t)) + (1-\delta) \cdot I(s_j^i, s^i(t)) \cdot \pi^i(t)}{\phi \cdot N(t-1) + 1} \\ &= \frac{\phi \cdot N(t-1) \cdot A_j^i(t-1) + \left(\delta + (1-\delta) I(s_j^i, s^i(t))\right) \cdot \pi(s_j^i, s^{-i}(t))}{\phi \cdot N(t-1) + 1}. \end{aligned}$$

This is the equation of the self-tuning EWA model in (Camerer, Ho, and Chong (2007)). Thus, the self-tuning EWA model is nested within this model, if $\nu = 1$, and $S^i = \{\text{Automaton 1}, \text{Automaton 2}\}$. Please notice that the Averaged Choice Reinforcement model as well as the Weighted Fictitious Play model are both, also, special cases of the modified self-tuning EWA model. □

Complexity

In the exercise studied we considered only two-state finite automata in order to limit, first, the computational burden of the simulations and, second, the automaton-complexity in accord with bounded rationality. Abreu and Rubinstein (1988) define the complexity of a strategy as the number of states of the minimal automaton implementing it. Despite its natural appeal, the latter measure is insufficiently sensitive to some essential features of complexity such as the monitoring of an opponent's actions. In particular, it is possible under this measure that an informationally demanding strategy that requires a fine monitoring of the opponent's action will be awarded the same degree of complexity as one that requires little or no monitoring of the opponent's action. Since the extent of monitoring required is certainly one aspect of complexity involved in implementing a strategy, greater complexity should be assigned to strategies requiring more monitoring.



(a) Automaton 1 (b) Automaton 4 (c) Automaton 7 (d) Three-State Automaton

Figure 9: Automaton-Complexity

In this study we define complexity in terms of the number of state-action pairs (q^i, a^{-i}) that require distinct transitions. This is easily seen to be the same as the number of transitions $R(M^i)$. Thus, under this measure of complexity, $M^i \succ^c M^j$ if and only if $R(M^i) > R(M^j)$. Furthermore, it is important to notice that this measure completely orders all automata with respect to their complexity.[12] Consider the finite automata in Figure 9. Our criterion ranks Automaton 7 as the most complex of the four depicted, despite having fewer states than the three-state Automaton.

Our choice to place the upper bound at two states admits a total of 26 automata (see Figure 3). Increasing the upper bound to three states would shoot up the number of admissible automata to 1024. We feel that such a number might be significantly big and thus unrealistic. Our complexity definition as described above is flexible enough to admit a much bigger number of automata (albeit less than 1024), without precluding automata with three states as long as the transitional capacity is at or below some upper bound choice set a priori.

[12] Alternatively, one could use the measure of complexity suggested by Banks and Sundaram (1990) where an automaton M^i is more complex than M^j if M^i is at least as complex as M^j along one of the twin dimensions of transitional complexity and size, and strictly more complex along the other. Such a criterion involves vector comparisons and is consequently not "complete" (i.e., does not permit a comparison between all machines).