KRANNERT SCHOOL OF MANAGEMENT

Purdue University
West Lafayette, Indiana

Strategic Learning With Finite Automata Via The EWA-Lite Model

By

Christos A. Ioannou
Julian Romero

Paper No. 1269
Date: April, 2012

Institute for Research in the Behavioral, Economic, and Management Sciences

Strategic Learning With Finite Automata Via The EWA-Lite Model

Christos A. Ioannou *† Julian Romero ‡

University of Southampton Purdue University

This draft: May 16, 2012

Abstract

A drawback of most adaptive learning models is that they only incorporate learning of stage-game actions rather than repeated-game strategies. Yet, evidence from the laboratory suggests that, in many games, subjects use repeated-game strategies. 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 likely 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 and compared to the predictions of the action-EWA-lite benchmark model. Relative to the benchmark, our modified EWA-lite model can better account for subject behavior.

*Our paper benefited greatly from discussions with Antonella Ianni, Miltiadis Makris and Spyros Galanis. 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 hybridized workhorse of adaptive 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 EWA in predicting behavior in seven different games and fits reliably better than the Quantal Response Equilibrium (QRE) model benchmark of McKelvey and Palfrey (1995). 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-game 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. In fact, the authors note that "(i)ncorporating 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) A natural next step is to develop a strategic learning model in which players learn about the performance of repeated-game strategies (see McKelvey and Palfrey (2001) and Haruvy and Stahl (2002)) rather than about the performance of stage-game actions. This is the goal of this paper. First, we modify the EWA-lite model to accommodate for a richer specification of strategies. Second, we solve the inference problem of going from histories to beliefs about opponents' strategies in a manner consistent with "belief-learning". Crucially, the modified EWA-lite model nests the model of Camerer, Ho, and Chong (2007).

Central to our framework is a key assumption that "allows" a player 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. In addition, the strategies of the players are implemented by a type of finite automaton, called a *Moore machine* (Moore (1956)). According to the thought experiment, a *fixed* pair of players is to play an infinitely-repeated game with perfect monitoring and complete information. A player is required to choose a strategy where strategies are chosen based on their attractions. Initially, each of the strategies in a player's candidate set has an equal

[1] The fictitious play model operates on the premise that players keep track of the history of previous play by other players and can thus form beliefs about what others will do in the future based on past observation. Consequently, players choose a strategy that maximizes their expected payoff given the beliefs they have 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.

attraction and, hence, an equal probability of being chosen. The attractions of the strategies are updated periodically as the payoffs resulting from strategy choices are observed. 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 used with greater frequency.

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 belief-based models are fictitious play, weighted fictitious play and Cournot best-response (Cournot 1960).[2] 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 experienced.

Other studies concentrate only on reinforcement learning. Harley (1981) for example, applied a reinforcement model using cumulative payoffs and simulated its behavior in several games. Roth and Erev (1995) extended the Harley model to incorporate spillover of reinforcement to neighboring strategies. Their model fits the time trends in ultimatum, public goods, and responder-competition games but the convergence is relatively slow. Finally, Hanaki, Sethi, Erev, and Peterhansl (2005) demonstrate that a simple reinforcement model of learning applied to a population of agents using a set of two-state automata accounts for the behavior of human subjects in the Stag-Hunt game, the Battle of the Sexes game, 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 hybridized model of learning, the EWA model, which captures adaptive learning by combining elements of, both, weighted fictitious play and reinforcement learning. A limitation of this approach is the fact that agents are not forward looking 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

[2]Boylan and El-Gamal (1993) propose a methodology to assess the empirical justifiability of fictitious play and Cournot learning in coordination and dominance-solvable games; in the experiments analyzed, they find overwhelming relative support for fictitious play.

subsequent paper, Camerer, Ho, and Chong (2002) address this limitation by extending the EWA model to capture sophisticated learning and "strategic teaching" in repeated games. 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' repeated-game strategies. The present study addresses these two issues by modifying EWA-lite in such a way so as to allow for 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. The predictions of the modified EWA-lite model are then validated with experimental (human) data from four symmetric 2×2 games; namely, the Prisoner's Dilemma game, the Battle of the Sexes game, the Stag-Hunt game, and the Chicken game. The dataset used was collected by Mathevet and Romero (2012). Finally, we compare our predictions to a baseline. Relative to the action-EWA-lite benchmark, the modified EWA-lite can better account for subject behavior. More specifically, trends observed in the experimental lab, such as cooperation in the Prisoner's Dilemma game, alternation across the two pure-strategy Nash equlibria in the Battle of the Sexes game, and mutual conciliation in the Chicken game (evading the two pure-strategy Nash equilibria), are predicted by the modified EWA-lite model, but not the action-EWA-lite model. The rest of the paper is organized as follows. In Section 2, we define the type of finite automaton used. In Section 3, the methodology is explained. In Section 4, the results of the computational simulations are presented. In Section 5, the results are discussed and validated with data from experiments with human subjects. Finally, in the Conclusion, we offer direction for future research.

2 Finite Automata

To simplify exposition we start with some notation. The stage game is represented in standard strategic (normal) form. The set of players is denoted by $I = \{1, ..., n\}$. Each player $i \in I$ has an *action set* denoted by A^i. An *action profile* $a = (a^i, a^{-i})$ consists of the action of player i, and the actions of the other players denoted by $a^{-i} = (a^1, ..., a^{i-1}, a^{i+1}, ..., a^n) \in A^{-i}$. In addition, each player i has a real-valued payoff function $g^i : A \to \mathbb{R}$ which maps every action profile $a \in A$ into a payoff for i, where A denotes the cartesian product of the action spaces A^i, written as $A \equiv \mathop{\times}\limits_{i=1}^{I} A^i$. In the infinitely-repeated game with *perfect monitoring*, the stage game in each period $t = 0, 1, ...$ 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. 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* $s^i \in S^i$ for player i is then a function $s^i : \mathcal{H} \to A^i$, where the strategy space of i consists of K^i discrete strategies; that is, $S^i = \{s^i_1, s^i_2, ..., s^i_{K^i}\}$. Furthermore, denote a strategy combination of the n players except i by $s^{-i} = (s^1, ..., s^{i-1}, s^{i+1}, ..., s^n)$. Each player i has a payoff function $\pi^i : S \to \mathbb{R}$ which maps every strategy profile $s = (s^i(t), s^{-i}(t)) \in S$ into a payoff for i, where S denotes the cartesian product of the strategy spaces S^i. Finally, player i's payoff in period t is denoted as $\pi^i(s^i(t), s^{-i}(t))$.

Our motivation to use finite automata stems from our desire to reduce the computational burden as well as to reflect elements of bounded rationality and complexity.[3] Using finite automata as the carriers of agents' strategies was first suggested by Aumann (1981).[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 player i, M^i, in a repeated game $G = (I, \{A^i\}_{i\in I}, \{g^i\}_{i\in I})$ is a four-tuple (Q^i, q^i_0, f^i, τ^i) where Q^i is a finite set of internal states of which q^i_0 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 player's action. It is pertinent to note that the transition function depends only on the present state and the other player's action. This formalization fits the natural description of a strategy as i's plan of action in all possible circumstances that are consistent with i's plans. In contrast, the notion of a game-theoretic strategy for i requires the specification of an action for every possible history, including those that are inconsistent with i's plan of action.[5]

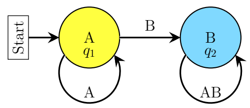
In the first period, the state is q^i_0 and the automaton chooses the action $f^i(q^i_0)$. If a^{-i} is the action chosen by the other player in the first period, then the state of i's automaton changes to $\tau^i(q^i_0, a^{-i})$, and in the second period i chooses the action dictated by f^i in that state. Then, the state changes again according to the transition function given the other agent's action. Thus,

[3] 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).

[4] 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.

[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 $\tau^i : Q^i \times A^{-i} \to Q^i$.

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 player.



$$Q^i = \{q_1, q_2\}$$
$$q_0^i = q_1$$
$$f^i(q) = \begin{cases} A & \text{if } q = q_1 \\ B & \text{if } q = q_2 \end{cases}$$
$$\tau^i\left(q, a^{-i}\right) = \begin{cases} q_1 & \text{if } \left(q, a^{-i}\right) = (q_1, A) \\ q_2 & \text{otherwise} \end{cases}$$

Figure 1: Grim-Trigger Automaton

For example, the automaton (Q^i, q_0^i, f^i, τ^i) in Figure 1 carries out the Grim-Trigger strategy. Thus, the strategy chooses A so long as both players have chosen A in every period in the past and chooses B otherwise. In the transition diagram, a vertex denotes the internal state of the automaton with the prescribed action of the agent, and the arcs labeled with the action of the other player indicate the transition to the states.

3 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. Unlike actions, repeated-game strategies require several periods to be evaluated. It is thus pertinent that players play the stage game a number of times before assessing their current strategy.

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 $N(t)$, which is interpreted as the number of observation-equivalents of past experience in period t. The second variable, denoted as $A_j^i(\chi)$, indicates player i's attraction to strategy j *after* the χ^{th} cluster of periods.[7] The length of a cluster can either be determined deterministically or randomly, and is denoted by T_χ. In this exposition, we assume that the cluster length is fixed, and denoted by T. The variables $N(t)$ and $A_j^i(\chi)$ begin with some prior values, $N(0)$ and $A_j^i(0)$. These prior values can be thought of as reflecting pre-game experience, either

[6]For more detailed information on EWA-lite, see Camerer, Ho, and Chong (2007).

[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.

due to learning transferred from different games or due to pre-play analysis. In addition, note that we use an indicator function $I(x, y)$ that equals 1 if $x = y$ and 0 otherwise. The evolution of learning over the χ^{th} cluster with $\chi \geq 1$ is governed by the following rule:

$$A_j^i(\chi) = \frac{\phi^i(\chi) \cdot N(t-1) \cdot A_j^i(\chi - 1) + (1 - \delta_j^i(\chi)) \cdot I(s_j^i, s^i(\chi)) \cdot R_j^i(\chi) + \delta_j^i(\chi) \cdot E_j^i(\chi)}{\phi^i(\chi) \cdot N(t-1) + 1}. \quad (1)$$

We outline next the differences between the modified EWA-lite model and the action-EWA-lite model. First, the reinforcement payoff in the modified EWA-lite model, $R_j^i(\chi)$, is defined as the average payoff obtained by player i over the χ^{th} cluster,

$$R_j^i(\chi) = \frac{\pi^i(s_i(\chi), s^{-i}(\chi))}{T}.$$

Second, the foregone payoffs in the modified EWA-lite model are not as simple as in the case of the action-EWA-lite model. In the latter model, the foregone payoffs are easy to calculate since an agent can infer the opponent's action in each period. In the modified EWA-lite model, agents are not able to observe the repeated-game strategy of their opponent, but, rather, only the sequence of action profiles in the cluster. Therefore, the forgone payoffs, $E_j^i(\chi)$, consist of the expectation taken over the possible repeated-game strategies that could have generated the specific sequence of action profiles in the cluster. Formally, a repeated-game strategy profile s is *consistent* with a sequence of action profiles h, if the sequence h occurs when players play s. Let $B : S \times \mathcal{H} \to \{0, 1\}$ be an indicator function which is 1 if the strategy profile is consistent with the history h and 0 otherwise,

$$B(s, h) = \begin{cases} 1 & \text{if } s \text{ is consistent with } h \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

Next, define belief function $p^i : S^i \times S^{-i} \times \mathcal{H} \to [0, 1]$ as

$$p(s^i, s^{-i}, h) = \frac{B((s^i, s^{-i}), h)}{\sum_{r \in S^{-i}} B((s^i, r), h)},$$ [8]

which can be interpreted as player i's belief that the other players played s^{-i} when player i played s^i and the history was h. Finally, let $h(\chi)$ denote the sequence of action profiles played in cluster χ. Then, the expected forgone payoff for player i of repeated-game strategy j over the χ^{th} cluster, is given by

$$E_j^i(\chi) = \sum_{s^{-i} \in S^{-i}} \pi^i(s_j^i, s^{-i}) \cdot p(s_j^i, s^{-i}, h(\chi)).$$

[8] We choose this form to maintain the one-parameter framework of Camerer, Ho, and Chong (2007). Alternatively, a logit specification could be used to calculate the probabilities.

Notice that an agent i puts an equal weight across all other agent's strategies, s^{-i}, which, when combined with s^i, are consistent with the history of action profiles in the cluster. For example, if both players use, in the χ^{th} cluster, the "Grim-Trigger" automaton as defined in Figure 1, then $h(\chi)$ consists of an all (A, A) sequence. Yet, the strategy profile where both players play "Grim-Trigger", as well as the strategy profile where player 1 plays "Grim-Trigger" and player 2 plays "All-A" are both consistent with $h(\chi)$ and therefore, each of the two strategies of the other agent is weighted equally.

In the original EWA model of Camerer and Ho (1999), the attraction function consisted of the exogenous parameters δ and ϕ. In the action-EWA-lite model, these parameters are changed from exogenous parameters to self-tuning functions $\delta(\cdot)$ and $\phi(\cdot)$, referred to as the attention function and the decay rate function, respectively. The attention function $\delta(\cdot)$ determines the weight placed on foregone payoffs. The idea is that players are more likely to focus on strategies that would have given them a higher payoff than the strategy actually played. This property is represented with the following function,

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

The attention function 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, players are paying attention to the better strategies whose attractions are increasing. This is in line with "learning direction" theory of Selten and Stoecker (1986), whereby subjects move towards the ex-post best response.

On the other hand, 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 the other player is 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 the other player's most recent strategy and the strategies chosen in the previous clusters. First, define the cumulative belief function $\sigma : S^{-i} \times \mathbb{N} \to [0, 1]$,

$$\sigma(s^{-i}, \chi) = \frac{1}{\chi} \sum_{j=1}^{\chi} p(s^i(j), s^{-i}, h(j)),$$

which records the historical frequencies of the beliefs of the other player's strategies over the χ clusters. The surprise index $\mathcal{S}^i(\chi)$ simply sums up the squared deviations between each cumulative belief $\sigma(s^{-i}, \chi)$ and the immediate belief $p(s^i, s^{-i}, h(\chi))$; that is,

$$\mathcal{S}^i(\chi) = \sum_{s^{-i} \in S^{-i}} (\sigma(s^{-i}, \chi) - p(s^i, s^{-i}, h(\chi))^2.$$

Thus, the surprise index captures the degree of change of the most recent beliefs from the historical average of beliefs. Note that it varies from zero (when there is belief persistence) to two (when a player is certain that the opponent just switched to a brand new strategy after playing a specific strategy from the beginning). The change-detecting decay rate of the χ^{th} cluster is then

$$\phi^i(\chi) = 1 - \frac{1}{2}\mathcal{S}^i(\chi).$$

Therefore, when player i's beliefs are not changing, then $\phi^i(\chi) = 1$; that is, the player puts high weight on previous attractions as indicated in (1). Alternatively, when player i's beliefs are changing, then $\phi^i(\chi) = 0$, that is, the player puts no weight on previous attractions.

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]

$$\mathbb{P}^i_j(\chi) = \frac{e^{\lambda A^i_j(\chi)}}{\sum_k e^{\lambda A^i_k(\chi)}}.$$

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

[9]An approach with an appeal to modelers desiring a highly-parsimonious model would be to calculate the probabilities of attractions with the equation

$$\mathbb{P}^i_j(\chi) = \frac{A^i_j(\chi)}{\sum_k A^i_k(\chi)}.$$

4 Results

We have limited our attention to the four symmetric 2×2 games; namely, the Prisoner's Dilemma game, the Battle of the Sexes game, the Stag-Hunt game, and the Chicken game. The payoff matrices are illustrated in Figure 2. To test the predictions of 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 1,000 periods. The agents are able to choose among the set of one-state and 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 on their new attractions. All simulations presented here use an intensity parameter $\lambda = 5$ in the logit specification.[11] The results displayed in the plots are averages taken over 500 simulated pairs.

4.1 Simulations

Figure 4 displays the results of the simulations in the Prisoner's Dilemma game (see Figure 2(a)). The cooperative action is denoted with the letter A, whereas the action of defection is denoted with the 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 outcome is for both players to play Automaton 12, the "Grim-Trigger" strategy. Figure 4(b) shows the progressions of probabilities as determined by the 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 1,000 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 3. Figure 4(c) shows the set of feasible repeated-game payoffs and the frequency of each payoff combination

[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. For expositional purposes, consider the action profiles of the first five periods of a pair consisting of the Grim-Trigger automaton and the Always-Defect automaton. The ability to earn the "temptation" payoff in the first period would bias favorably the Always-Defect automaton over the Grim-Trigger one even though a cycle would consist of an all-defections sequence. Thus, allowing for a cluster length of 20 periods, discounts the payoff earned in the first period over the payoff earned in the entire cluster length.

[11] We have experimented with $\lambda \in \{1, 2, 3, 5, 10, 20, 50\}$. Our results are relative insensitive to changes in these values.

	A	B
A	3,3	1,4
B	4,1	2,2

(a) Prisoner's Dilemma Game

	A	B
A	1,1	2,4
B	4,2	1,1

(b) Battle of the Sexes Game

	A	B
A	3,3	0,2
B	2,0	1,1

(c) Stag-Hunt Game

	A	B
A	3,3	1,4
B	4,1	0,0

(d) Chicken Game

Figure 2: Payoff Matrices

over the final 200 periods of the 1,000 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 around 2.5, which is the payoff that would be obtained if both players were randomizing. Payoffs then increase towards 3, 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). Thus, each player prefers the equilibrium in which he plays A and his opponent plays B, or vice versa. The frequency plot in Figure 5(a) covers a large number of automata although Automaton 21 shows up more frequently. Automaton 21 switches actions only if the opponent played the same action in the previous period; otherwise, continues with the same action. The large number of automata selected can be attributed to the large number of action-combinations that can produce the two pure-strategy equilibira indicated above. In addition, Figure 5(b) suggests that the two strategies that are most likely to be played are Automaton 1 (which always plays A) and Automaton 21. Looking at Figure 5(c), we see mass points at payoffs $(2, 4)$ and $(4, 2)$ which correspond to the two pure-strategy equilibria. In addition, we also observe a mass point at $(3, 3)$. This latter mass

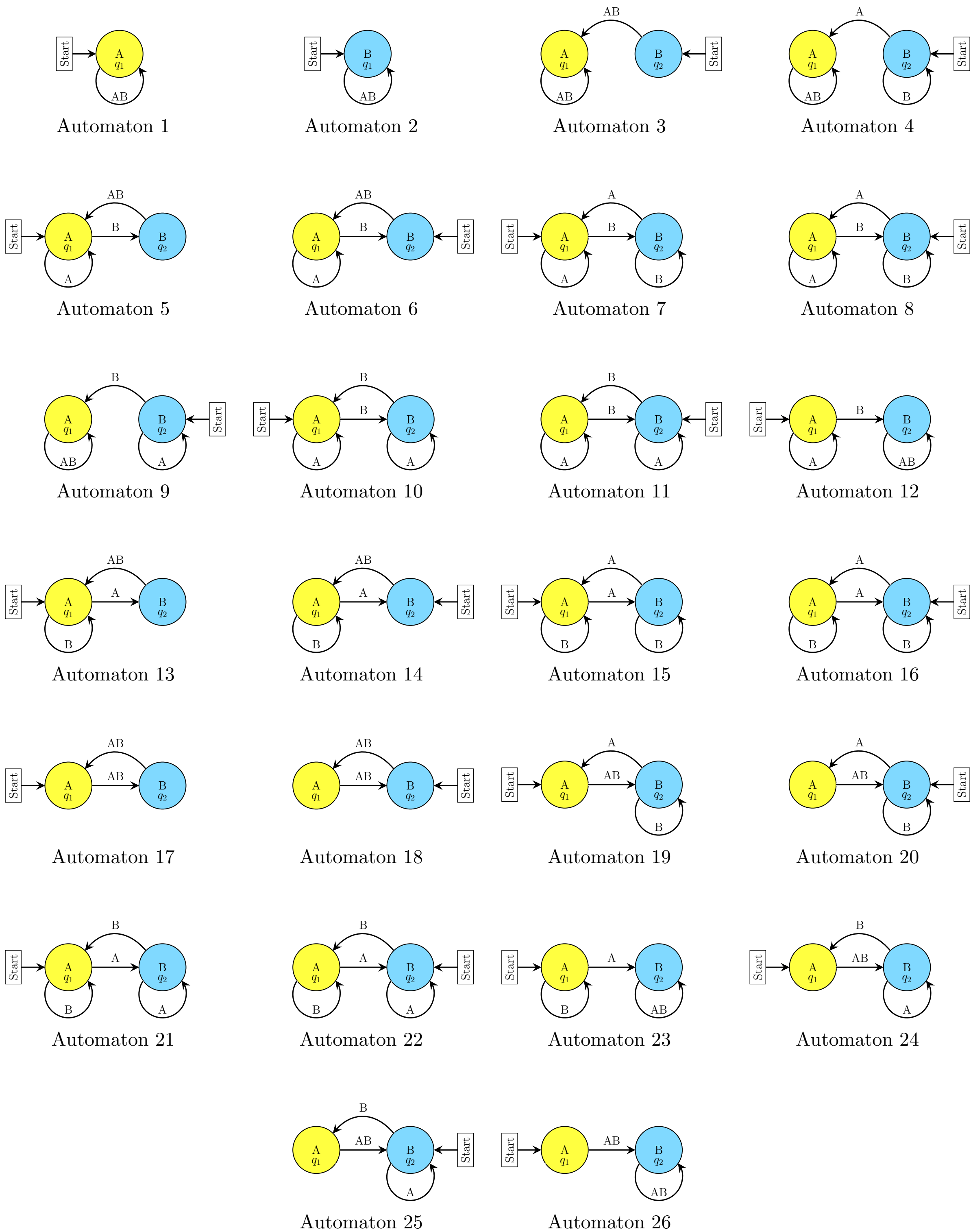
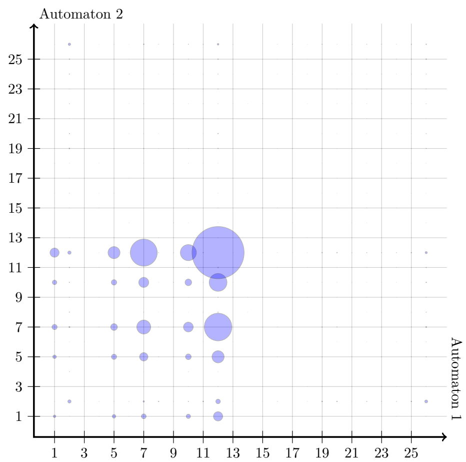
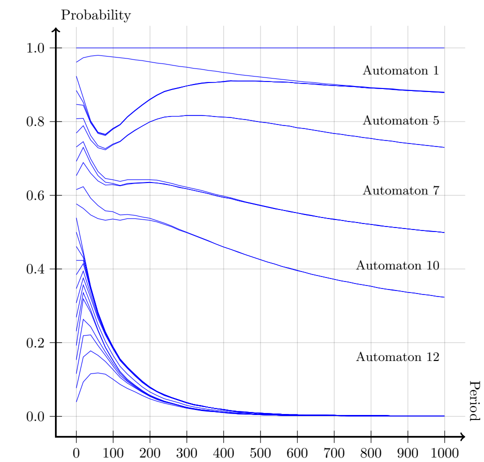


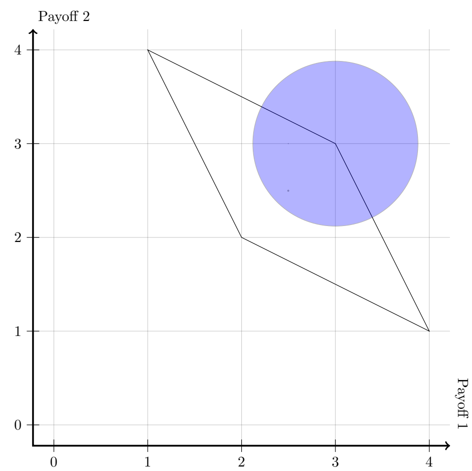
Figure 3: Two-State Automata



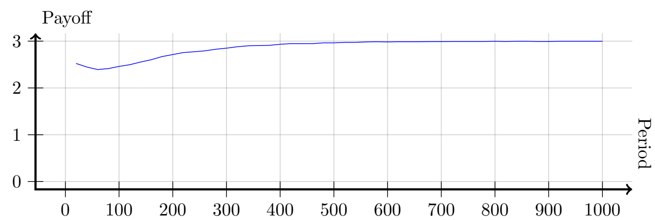
(a) Automaton Frequency



(b) Attraction Probability Progression



(c) Payoff Frequency

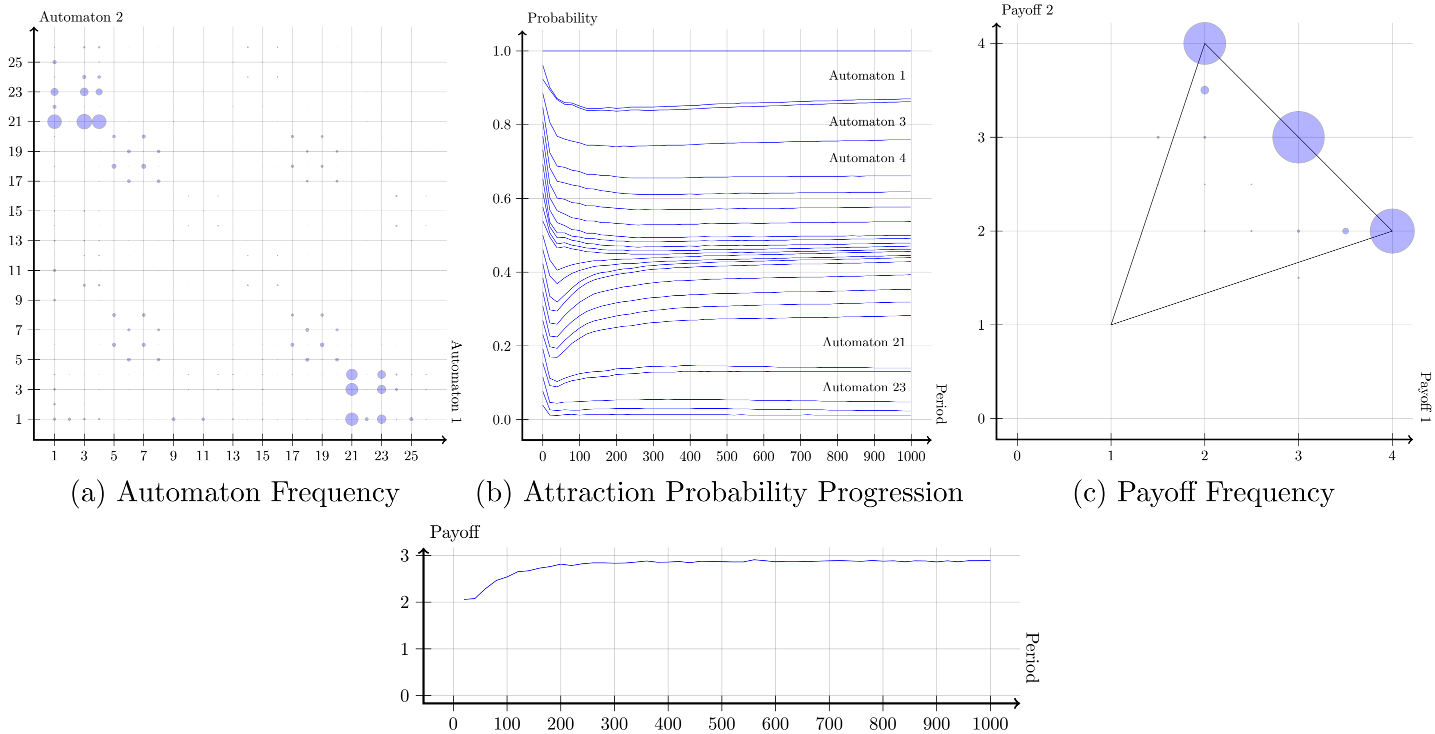


(d) Average Payoff Progression

Figure 4: Prisoner's Dilemma Game. Figures 4-7 all follow the same structure. Each pair of agents is matched for 1,000 periods. There are 500 such pairs in the population. The automaton frequency plot shows the frequency of play across 26 automata over 1,000 periods. In the probability progression plot, the difference between two successive curves indicates the likelihood of an automaton being chosen as the game evolves based on attractions. The frequency plot shows the frequency of each payoff combination over the final 200 periods and the set of feasible payoffs. The payoff progression plot indicates the average payoff of the 500 pairs as the game evolves. The intensity parameter λ is taken to equal 5 in all simulations.

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 Mathevet and Romero (2012)) and would be impossible to obtain using a model that only allows action-learning. Notice that in spite of observing the $(3,3)$ 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 alternations between the pure strategy equilibria. All in all, there are 32 combinations of automata that lead to these 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 $(3,3)$ in Figure 5(c). Figure 5(d) shows that the average payoff per player converges very close to 3.

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, (A,A) is



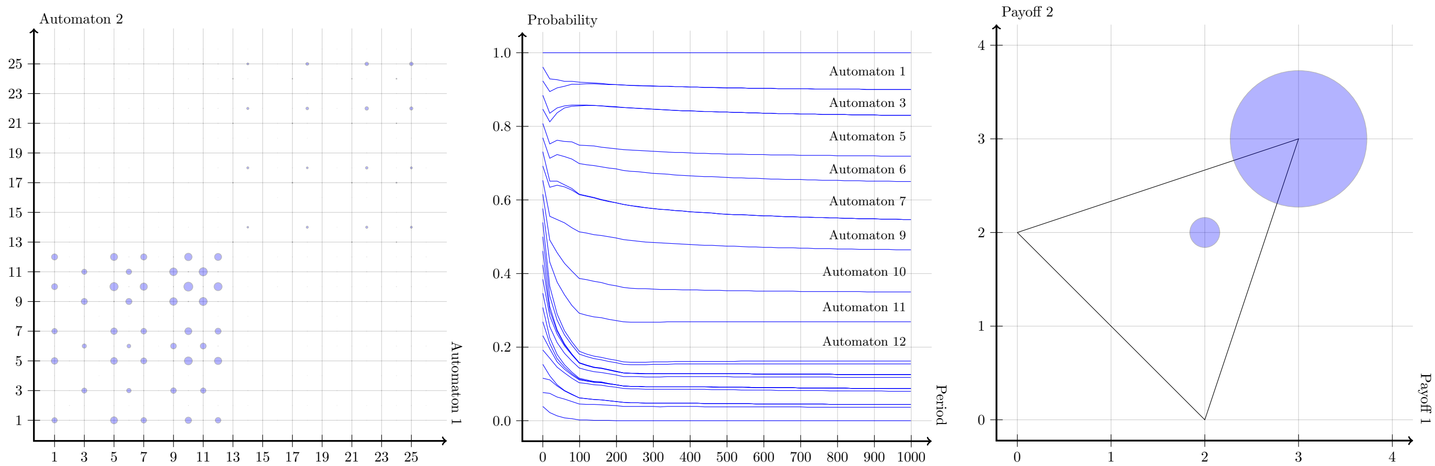
(a) Automaton Frequency (b) Attraction Probability Progression (c) Payoff Frequency

(d) Average Payoff Progression

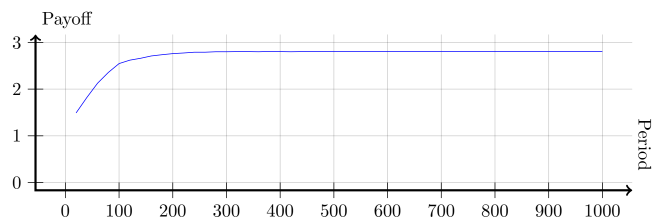
Figure 5: Battle of the Sexes Game

the Pareto dominant equilibrium. Figures 6(a) and (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 $(3, 3)$. The payoff combination (2,2) also arises in a much smaller scale than $(3, 3)$. The former payoff combination results from an alternation of the two pure-strategy Nash equilibria or $(2, 0)$ and $(0, 2)$. Unlike the Prisoner's 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 which explains the weak convergence to a small set of automata and that the average payoff converges to just less than 3 as shown in Figure 6(d).

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 Figures 7(a) and (b) look similar to those in the Prisoner's Dillemma where game-play is converging to a small set of automata with a cooperative outcome of (A, A). This observation is confirmed in Figure 7(c) which displays that a large percentage of simulations end with payoffs corresponding to the cooperative outcome. Finally, Figure 7(d) suggests that average payoff converges to just

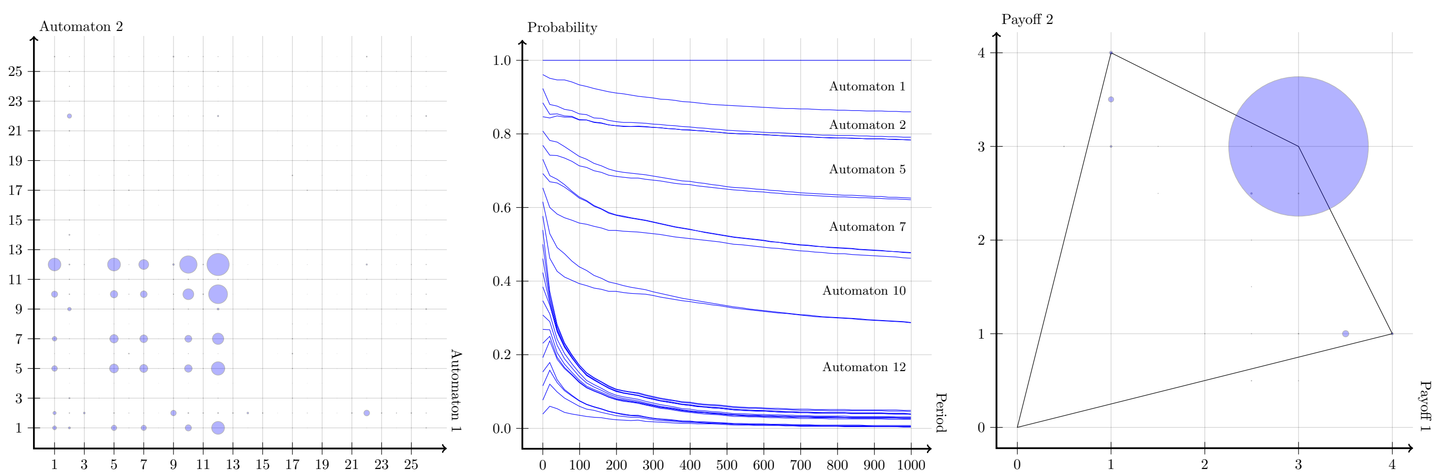


(a) Automaton Frequency (b) Attraction Probability Progression (c) Payoff Frequency

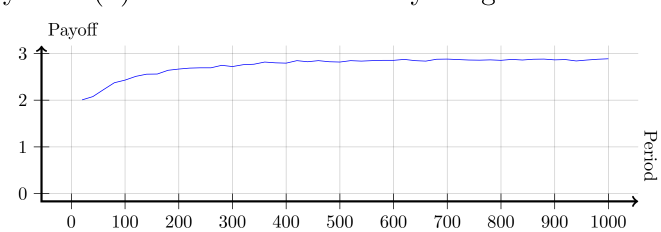


(d) Average Payoff Progression

Figure 6: Stag-Hunt Game



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



(d) Average Payoff Progression

Figure 7: Chicken Game

below 3.

5 Discussion

In Figure 8, we use the action-EWA-lite as a benchmark model and the modified EWA-lite (with one and two state-automata) to compare their predictions to the evidence from experiments with human data. Mathevet and Romero (2012) provide experimental data on the four games reported in Section 4. The data consists of 27 pairs of subjects playing a game with a fixed opponent. Each pair plays 30 periods with certainty, after which the probability of playing an additional period is 0.9. Some common trends emerge from the data such as convergence to the cooperative outcome in the Prisoner's Dilemma game, alternations between the two pure-strategy Nash equilibrium in the Battle of the Sexes game, convergence to the payoff-dominant Nash equilibrium in the Stag-Hunt game, and mutual conciliation in the Chicken game. The simulations of the modified EWA-lite model match the experimental data quite well. On the contrary, the action-EWA-lite benchmark model does relatively well in the Stag-Hunt game, but fails to predict the most likely outcomes in the other three games.

As Figure 8 shows, the extension from actions to even a simple class of repeated-game strategies improves predictions in the games studied here. It should be noted that the richer specification of strategies improves predictions in two distinct ways. First, it allows for convergence to non-trivial sequences such as alternations in the Battle of the Sexes game. Convergence to any non-trivial sequence with an action-learning model is difficult as it requires probabilities of actions to drastically change period to period. Second, the richer set of strategies allows more sophisticated strategic behavior such as punishments which leads to behavior such as cooperation in the Prisoner's Dilemma game or mutual conciliation in the Chicken game. Incorporating these additional strategic behaviors in an action-learning model is not straightforward.

Hanaki, Sethi, Erev, and Peterhansl (2005) run simulations with a reinforcement learning model over a population of agents using two-state automata. The population of 1,000 agents starts by playing a 'pre-experimental' phase consisting of at least 20,000 periods. The players are randomly matched until the population average attractions have converged. Once the population attractions have converged, players are randomly selected from the population to play in a fixed pair. In contrast, by adding a belief-based component to reinforcement learning, the modified EWA-lite model provides relatively quick convergence with fixed pairs while dispensing of the 'pre-experimental' phase. In particular, 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,

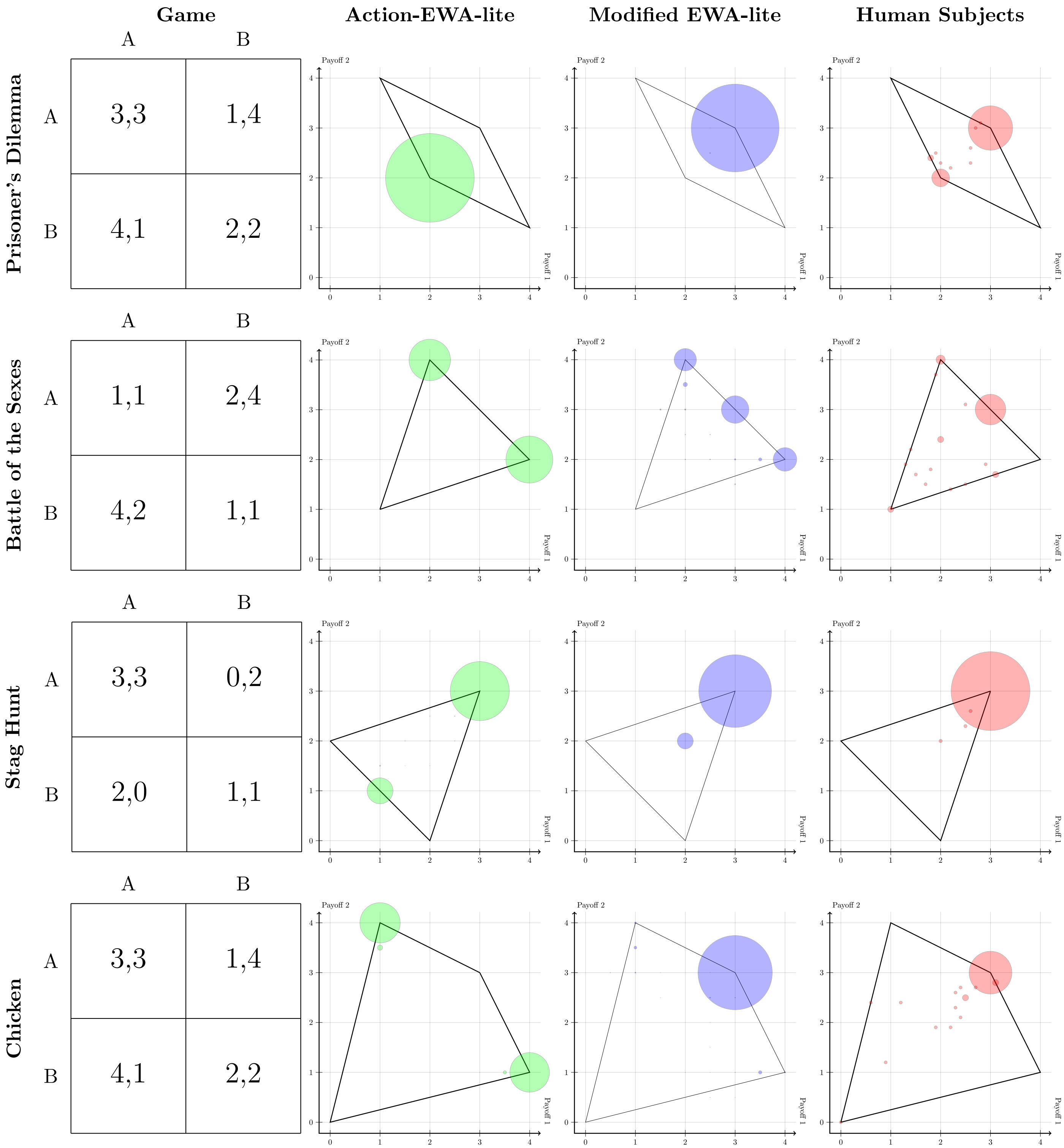


Figure 8: We compare simulations of the action-EWA-lite benchmark model and the modified EWA-lite model with human data from the experiments of Mathevet and Romero (2012). The simulations assume a $\lambda = 5$ and the payoffs are averaged over the last 200 periods (periods $800 - 1,000$). On the other hand, the experimental results with human subjects consist of averaging the last 10 periods of fixed pairs in repeated games.

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.

Next, we identify the Nash equilibria automata picked by the modified EWA-lite model.[12] In the Prisoner's Dilemma game, with the exception of Automaton 1,"Always-Cooperate", all other automaton-pairs picked by the modified EWA-lite model constitute Nash equilibria. In the Battle of the Sexes, all automaton-pairs that lead to alternations are Nash equilibria. [13] However, Automaton 21, which is played with relatively high frequency as shown in 5(a), is not a Nash equilibrium when matched with any other automaton. In the Stag-Hunt game, all automaton-pairs picked by the model are Nash equilibria. Finally, in the Chicken game all automaton-pairs that lead to the cooperative outcome (A, A) are Nash equilibria. However, Automaton 2 and Automaton 22 also occur with positive frequency. Though this pair is not a Nash equilibrium, it consists of stage-game Nash equilibrium play with the exception of the first period.

6 Conclusion

The enterprise of finding out what strategies subjects actually choose has not progressed to the degree 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). Similar to the framework in Hanaki, Sethi, Erev, and Peterhansl (2005), we make no *a priori* assumptions on fairness or reciprocity as primitive concerns, allowing monetary payoffs to be the sole driving force of learning. 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 game, the Battle of the Sexes game, the Stag-Hunt game, and the Chicken game. Relative to the action-EWA-lite benchmark model, the modified EWA-lite can better account for subject behavior.

The games examined here have but a few layers of complexity. The model seems to work adequately well with such games predicting the main patterns observed in the laboratory experiments. Yet, an interesting direction would be to also allow for sophisticated players. As it stands,

[12] The matrices with the best-responses for each of the 26 automata in the four games considered are provided in the Appendix.

[13] In the Battle of the Sexes game, players always prefer to play the opposite action to that of their opponent.

the proposed model is myopic in the sense that players are assumed to make decisions in the current period of play without taking into account how their current choice could affect their future payoffs. A natural extension of our work, would then be to endow players with sophistication so as to engage in a form of strategic "teaching" as described in Camerer, Ho, and Chong (2002).

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, often 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, they 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.

Lastly, 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 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.

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.

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.

Camerer, Colin F., Teck-Hua Ho, and Juin-Kuan Chong. "Sophisticated EWA Learning and Strategic Teaching in Repeated Games." *Journal of Economic Theory* 104: (2002) 137–88.

———. "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.

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.

Mathevet, Laurent, and Julian Romero. "Predictive Repeated Game Theory: Measures and Experiments.", 2012.

McKelvey, Richard, and Thomas R. Palfrey. “Quantal Response Equilibria in Normal Form Games.” *Games and Economic Behavior* 10: (1995) 6–38.

———. “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.

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 EWA-lite model nests the action-EWA-lite 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. $T = 1$, so that all clusters have a length of one.

Given these assumptions, the modified EWA-lite model is equivalent to the action-EWA-lite model of Camerer, Ho, and Chong (2007). To prove this, we show that in this special case the terms ϕ, δ, R^i_j, and E^i_j from the modified EWA-lite model are all equivalent to the corresponding terms in the action-EWA-lite 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.

First, the reinforcement payoff in this case is,

$$R^i_j\,(t) = \frac{\pi^i\left(a^i(t), a^{-i}(t)\right)}{T} = \pi^i(t). \tag{3}$$

Next, determine the forgone payoff term $E^i_j(t)$. Since there are only two strategies in S^i, the beliefs for the previous period are unambiguous, hence we can rewrite (2) as

$$B\left(\left(s^i, s^{-i}\right), a(t)\right) = \begin{cases} 1 & \text{if } \left(s^i, s^{-i}\right) \text{ is consistent with } a(t) \\ 0 & \text{otherwise} \end{cases} = I\left(s^{-i}, s^{-i}(t)\right),$$

where $I(x, y)$ equals 1 if $x = y$ and 0 otherwise. Therefore,

$$p(s^i, s^{-i}, a(t)) = \frac{B((s^i, s^{-i}), a(t))}{\sum\limits_{r \in S^{-i}} B((s^i, r), a(t))} = I\left(s^{-i}, s^{-i}(t)\right).$$

Hence, we get that

$$E^i_j\,(t) = \sum_{s^{-i} \in S^{-i}} \pi^i(s^i_j, s^{-i}) \cdot p(s^i_j, s^{-i}, a(t)) = \pi^i(s^i_j, s^{-i}(t)). \tag{4}$$

Therefore, $R^i_j(t) = E^i_j(t) = \pi^i(s^i_j, s^{-i}(t))$.
From equations (3) and (4) we get that the attention function is,

$$\delta^i_j(t) = \begin{cases} 1 & \text{if } \pi^i(s^i_j, s^{-i}(t)) \geq \pi^i(t) \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

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

Next, to determine the decay function $\phi(\cdot)$, notice that,

$$\sigma(s^{-i}, t) = \frac{1}{t}\sum_{j=1}^{t} p(s^i(j), s^{-i}, h(j)) = \frac{1}{t}\sum_{j=1}^{t} I\left(s^{-i}, s^{-i}(t)\right),$$

which is simply the frequency with which s^{-i} has been played in the t periods. This is the same as the vector elements $h_i^k(t)$ from Camerer, Ho, and Chong (2007). Also, notice that $p\left(s^i, s^{-i}, a(t)\right) = I\left(s^{-i}, s^{-i}(t)\right)$ is the same as the term $r_i^k(t)$ from Camerer, Ho, and Chong (2007). Therefore, the surprise index is also the same,

$$\mathcal{S}^i(t) = \sum_{s^{-i}\in S^{-i}} (\sigma(s^{-i}, t) - p(s^i, s^{-i}, a(t))^2.$$

Finally, the change-detector function is

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

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

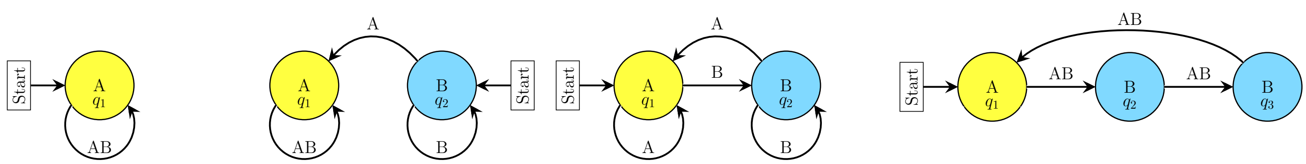
Finally, combining equations (1), (3), (4), (5), and (6), we get

$$\begin{aligned} A_j^i(t) &= \frac{\phi^i(\chi)\cdot N(t-1)\cdot A_j^i(\chi-1) + \delta_j^i(\chi)\cdot E_j^i(\chi) + (1-\delta_j^i(\chi))\cdot I(s_j^i, s^i(t))\cdot R_j^i(\chi)}{\phi^i(\chi)\cdot N(t-1)+1} \\ &= \frac{\phi^i(t)\cdot N(t-1)\cdot A_j^i(t-1) + [\delta_j^i(t) + (1-\delta_j^i(t))I(s_j^i, s^i(t))]\cdot \pi^i(s_j^i, s^{-i}(t))}{\phi^i(t)\cdot N(t-1)+1}. \end{aligned}$$

This is the equation of the action-EWA-lite model in Camerer, Ho, and Chong (2007). Thus, the action-EWA-lite model is nested within this model, if $T = 1$, and $S^i = \{\text{Automaton } 1, \text{Automaton } 2\}$. Notice that the Averaged Choice Reinforcement model as well as the Weighted Fictitious Play model are both special cases of the action-EWA-lite model and, by the above, also special cases of the modified EWA-lite 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.[14] 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.

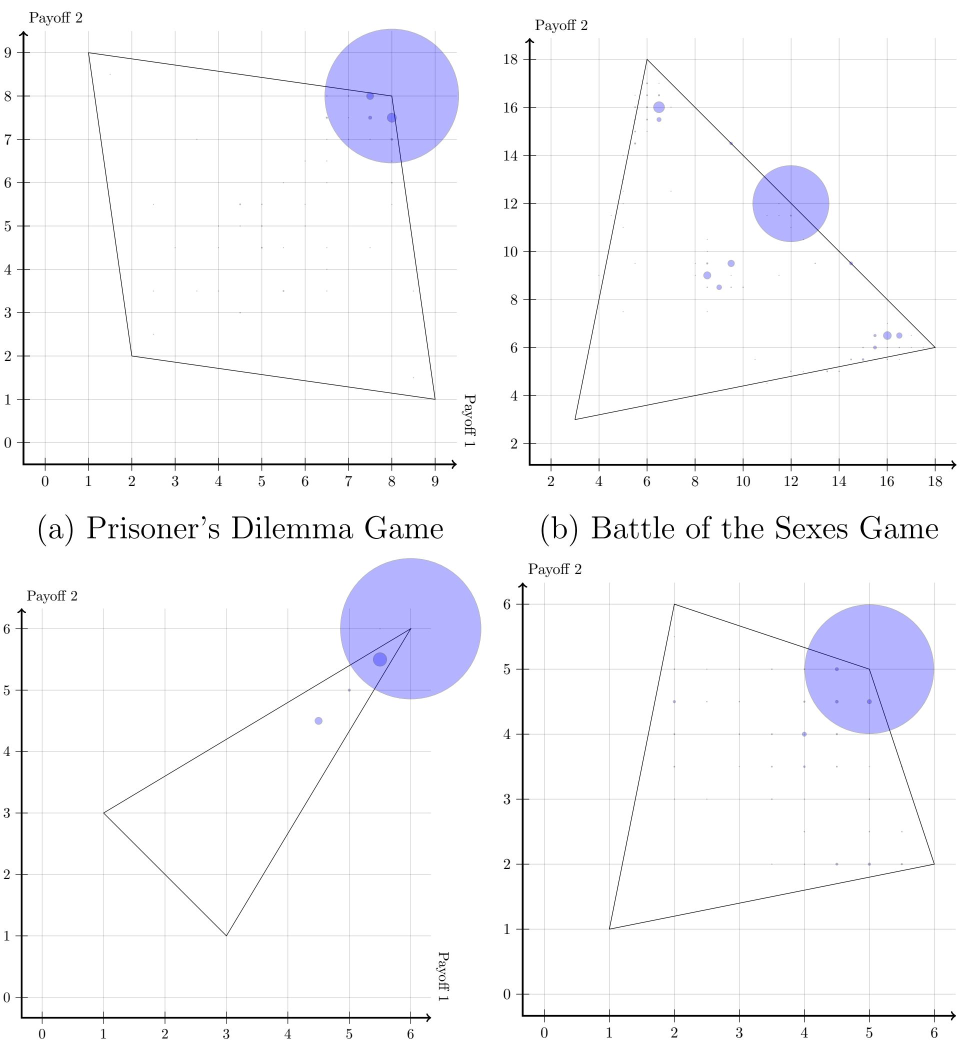
Complexity Robustness

To check for robustness of the results, we ran simulations using a larger set of automata. The natural one level up from two-state automata is to use three-state automata. However, the minimal set of three-state automata consists of 1054 automata, which proves to be computationally intensive. As an alternative, we consider all automata with complexity, as defined earlier, of less than four, $(R\left(M^i\right) \leq 4)$. This set consists of 228 automata ranging from one state-automaton to four-state automata. Note that the set also includes every two-state automaton. The simulations using the larger set of automata are displayed in Figure 10. The plots preserve the trends in the earlier results with two-state automata, yet are more

[14] 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).

noisy. The additional noise can be attributed to the bigger cardinality of the strategy space which leads to slower convergence.

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.



(a) Prisoner's Dilemma Game (b) Battle of the Sexes Game

(c) Stag-Hunt Game (d) Chicken Game

Figure 10: Modified EWA-lite simulations with automata of complexity 4 or less ($R\left(M^i\right) \leq 4$). Simulations are taken with $\lambda = 5$ and payoffs are averaged over the last 200 periods (periods $800 - 1,000$).

Payoff Robustness

McKelvey and Palfrey (2001) also provide experimental data on the four games reported in the paper, but their games consist of different payoffs. Thus, we test how sensitive are the predictions of the modified EWA-lite model to payoffs by comparing the simulations of the action-EWA-lite benchmark and the modified EWA-lite model with their experimental dataset. The data consists of 24 pairs of subjects playing a game with a fixed opponent for 50 periods. [To be completed]

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

(a) Prisoners' Dilemma Game

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

(b) Battle of the Sexes Game

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

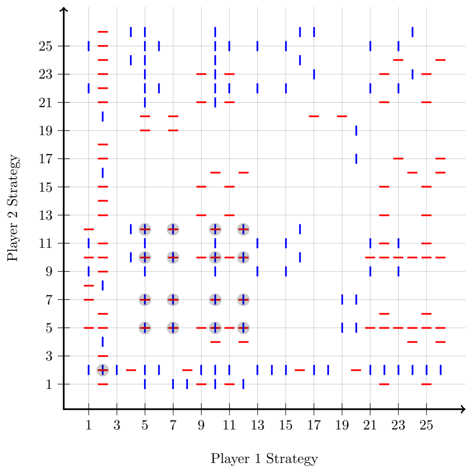
(c) Stag-Hunt Game

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

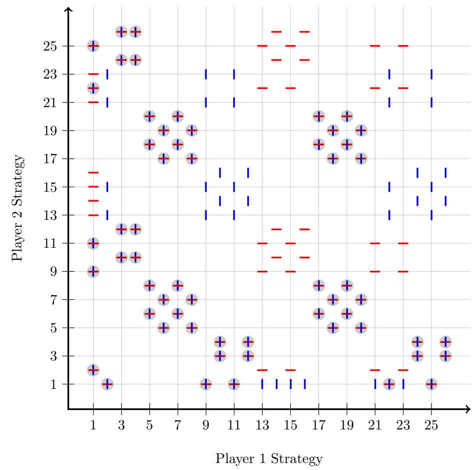
(d) Chicken Game

Figure 11: Payoff Matrices

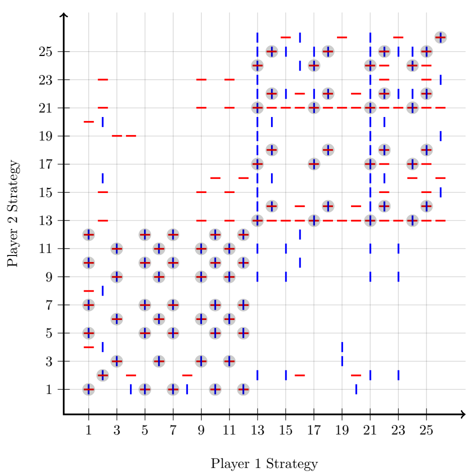
Best Responses and Nash equilibria



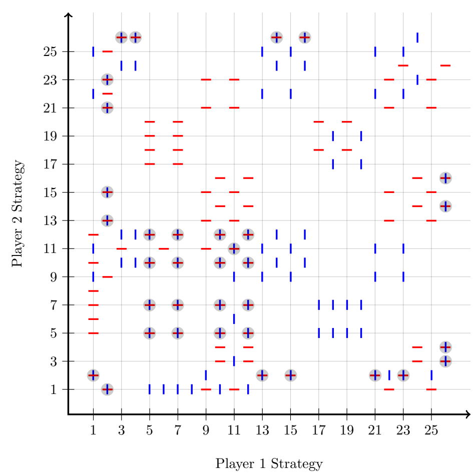
(a) Prisoner's Dilemma Game



(b) Battle of the Sexes Game



(c) Stag-Hunt Game



(d) Chicken Game

Figure 12: We indicate the Nash equilibrium automaton pairs. The red (horizontal) lines correspond to the best responses of player 2, whereas the blue (vertical) lines correspond to the best responses of player 1. The circled crosses represent Nash equilibria.