Strategic Learning With Finite Automata Via The EWA-Lite Model

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

University of Southampton Purdue University

This draft: July 4, 2012

Abstract

A drawback of most adaptive models is that they incorporate learning only of stage-game actions rather than of 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 probable performance of a set of one-state and 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.

JEL Classification: C51, C92, C72, D03

Keywords: Adaptive Models, Reinforcement Learning, Belief-Based Learning, Experience Weighted Attraction Model, Finite Automata

[*]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 in which strategies are stage-game actions. More specifically, Camerer, Ho, and Chong (2007) indicated that EWA-lite does as well as EWA in predicting behavior in seven different games and better than the Quantal Response Equilibrium (QRE) model benchmark of McKelvey and Palfrey (1995). In fact, recently, Chmura, Goerg, and Selten (2011) noted 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 repeated-game strategies. As Camerer and Ho (1999) acknowledged 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, also, McKelvey and Palfrey (2001), Erev and Roth (2001), and Haruvy and Stahl (2002)) rather than about the performance of stage-game actions. This is the goal of our paper. We modify the EWA-lite model so as to accommodate a richer specification of strategies and 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 imposes complexity constraints on the candidate set of strategies so as to limit the number of potential strategies considered. In addition, the players' strategies 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 based on its attractions. Initially, each of the strategies in a player's candidate set has an equal attraction and, hence, an equal probability of being chosen. The strategies' attractions are updated periodically as the payoffs resulting from strategy choices are observed. If

[1] The fictitious-play model operates on the premise that players keep track of the history of previous play by other players and thus, can 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.

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 they form by observing the history of play. Some special cases of belief-based models are fictitious play, weighted fictitious play and Cournot best-response (Cournot 1960).[2] To fit data from coordination games, Crawford (1995) developed a more general belief model, allowing idiosyncratic shocks in beliefs and time-varying weights. Crawford and Broseta (1998) extended the model to allow ARCH error-terms and applied it to coordination with preplay auctions. Other studies have concentrated 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 reinforcement spillovers 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) demonstrated that a simple reinforcement model of learning applied to a population of agents using a set of one-state and 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. For example, reinforcement does better in constant-sum games and belief learning in coordination games. However, 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 their opponents' future use of alternative strategies. In a subsequent paper, Camerer, Ho, and Chong (2002) addressed 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) addressed 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 has been the need to incorporate a richer specification of strategies and to solve the inference problem of going from histories to beliefs about opponents' repeated-game strategies. The present study addresses this issue by modifying EWA-

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

lite to allow for a richer specification of strategies and a belief-based learning rooted in players updating their beliefs on the probability distribution of the other players' strategies. We then study the probable performance of a set of one-state and two-state automata in four symmetric 2×2 games: the Prisoner's Dilemma game, the Battle of the Sexes game, the Stag-Hunt game, and the Chicken game. Finally, we validate the predictions of the modified EWA-lite model with experimental (human) data and compare them to the predictions of the (action)-EWA-lite benchmark. Relative to the 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 by 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. Section 3 explains the methodology. In Section 4, we present the results of the computational simulations. In Section 5, we discuss and validate the results with data from experiments with human subjects. In Section 6, we offer conclusions and 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 $A \equiv \overset{I}{\underset{i=1}{\times}} A_i$. 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 $a^r = (a_1^r, ..., a_n^r)$ 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 $(Q_i,\ q_i^0,\ f_i,\ \tau_i)$ 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]

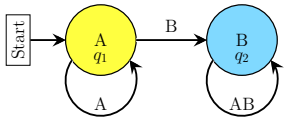
In the first period, the state is q_i^0, and the automaton chooses the action $f_i(q_i^0)$. If a_{-i} is the action chosen by the other player in the first period, then the state of i's automaton changes to $\tau_i(q_i^0, a_{-i})$, and in the second period, i chooses the action dictated by f_i in that state. Then, the state changes again according to the transition function given the other agent's action. Thus, 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.

For example, the automaton $(Q_i,\ q_i^0,\ f_i,\ \tau_i)$ in Figure 1 carries out the Grim-Trigger strategy. In the transition diagram, a vertex denotes the internal state of the automaton with the prescribed agent's action indicated in the center, and the arcs labeled with the action of the other player indicate the transition to the states. Thus, the strategy chooses A, as long as both players have

[3]Bounded rationality suggests that a player may not consider all feasible strategies but must limit himself, instead, to less-complex strategies. The complexity of finite automata may be defined in a number of ways. Our definition 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 that 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 so that $\tau_i : Q_i \times A \to Q_i$ instead of $\tau_i : Q_i \times A_{-i} \to Q_i$.



$$Q_i = \{q_1, q_2\}$$
$$q_i^0 = q_1$$
$$f_i(q) = \begin{cases} A \text{ if } q = q_1 \\ B \text{ if } q = q_2 \end{cases}$$
$$\tau_i(q, a_{-i}) = \begin{cases} q_1 \text{ if } (q, a_{-i}) = (q_1, A) \\ q_2 \text{ otherwise} \end{cases}$$

Figure 1: Grim-Trigger Automaton

chosen A in every period in the past, and chooses B otherwise.

3 Methodology

Players have attractions, or propensities, 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 that are updated at the end of a *cluster of periods*. We abstain from the common practice of updating the attractions of strategies in every period for two reasons. First, unlike actions, repeated-game strategies require several periods to be evaluated. Second, as Erev and Haruvy (2012) documented in their chapter, subjects exhibit high probability of inertia during the game-play stage. More specifically, in some games, subjects repeat their last action-choice in more than 80 percent of the plays, even after a realized bad outcome. For these two reasons, we find it appropriate that players play the stage-game for a cluster of periods before assessing their current strategy. The length of a cluster, denoted by T_χ, can be fixed or determined stochastically. In this exposition, we assume that the cluster length is fixed and is denoted by T.

The modified EWA-lite model consists of two variables that are updated once an agent switches strategies.[6] Crucially, we will assume that players update their strategies simultaneously. The first variable is $N(\chi)$, which is interpreted as the number of observation-equivalents of past experience in cluster χ. The second variable, denoted as $A_i^j(\chi)$, indicates player i's attraction to strategy j *after* the χth cluster of periods. The variables $N(\chi)$ and $A_i^j(\chi)$ begin with some prior values, $N(0)$ and $A_i^j(0)$. These prior values can be thought of as reflecting pre-game experience, due either to learning transferred from different games or to pre-play analysis. In addition, we use an indicator function $I(x, y)$ that equals 1 if $x = y$ and 0 otherwise. The evolution of learning over the χth

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

cluster with $\chi \geq 1$ is governed by the following rule:

$$A_i^j(\chi) = \frac{\phi_i(\chi) \cdot N(\chi - 1) \cdot A_i^j(\chi - 1) + I(s_i^j, s_i(\chi)) \cdot R_i(\chi) + \delta_i^j(\chi) \cdot E_i^j(\chi)}{\phi_i(\chi) \cdot N(\chi - 1) + 1}. \tag{1}$$

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, $R_i(\chi)$, is defined as the average payoff obtained by player i over the χth cluster,

$$R_i(\chi) = \frac{1}{T} \sum_{a \in h(\chi)} g_i(a),$$

where $h(\chi)$ is the sequence of action profiles played in the χth cluster. Second, the forgone payoffs in the modified EWA-lite model are not as simple as in the case of the (action)-EWA-lite model, where the opponent's action is publicly observed 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, $E_i^j(\chi)$, 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:

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

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

$$p(s_i, s_{-i}, h) = \frac{B((s_i, s_{-i}), h)}{\sum_{r \in S_{-i}} B((s_i, r), h)},$$ [7]

which can be interpreted as player i's belief that the other player chose s_{-i} when player i played s_i and the history was h. 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 χth cluster, the "Grim-Trigger" automaton as defined in Figure 1, then $h(\chi)$ consists of an all (A, A) sequence. Yet, the strategy profile in which both players play "Grim-Trigger," as well as the strategy profile in which player 1 plays

[7] 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.

"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. Finally, the expected forgone payoff for player i of repeated-game strategy j over the χth cluster is given by

$$E_i^j(\chi) = \sum_{s_{-i} \in S_{-i}} \pi_i(s_i^j, s_{-i}) \cdot p(s_i, s_{-i}, h(\chi)).$$

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 forgone 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 by the following function:

$$\delta_i^j(\chi) = \begin{cases} 1 & \text{if } E_i^j(\chi) \geq R_i(\chi) \text{ and } s_i^j \neq s_i(\chi) \\ 0 & \text{otherwise.} \end{cases}$$

Thus, on the one hand, the attention function enables subjects to reinforce only *unchosen* strategies with weakly better payoffs by a weight of one. On the other hand, the decay rate function $\phi(\cdot)$ weighs lagged attractions. When a player senses that the other player is changing behavior, 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 he chose 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, over the χ clusters, when the other player chose strategy s_{-i}. 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 χth cluster is, then,

$$\phi_i(\chi) = 1 - \frac{1}{2}\mathcal{S}_i(\chi).$$

Therefore, when player i's beliefs are not changing, $\phi_i(\chi) = 1$; that is, the player weighs previous attractions fully, 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 determine probabilities of choosing strategies. We use the logit specification to calculate the choice probability of strategy j. Thus, the probability of a player i choosing strategy j, when he updates his strategy at the beginning of cluster $\chi + 1$, is[8]

$$\mathbb{P}_i^j(\chi + 1) = \frac{e^{\lambda \cdot A_i^j(\chi)}}{\sum_k e^{\lambda \cdot A_i^k(\chi)}}.$$

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.

4 Results

We have limited our attention to the four symmetric 2×2 games: 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 from 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 $A_i^j(0) = 1.5$ (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

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

$$\mathbb{P}_i^j(\chi + 1) = \frac{A_i^j(\chi)}{\sum_k A_i^k(\chi)}.$$

	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

end of each cluster, which consists of 20 periods,[9] 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.[10] 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

[9]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 five 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 the Always-Defect automaton favorably 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.

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

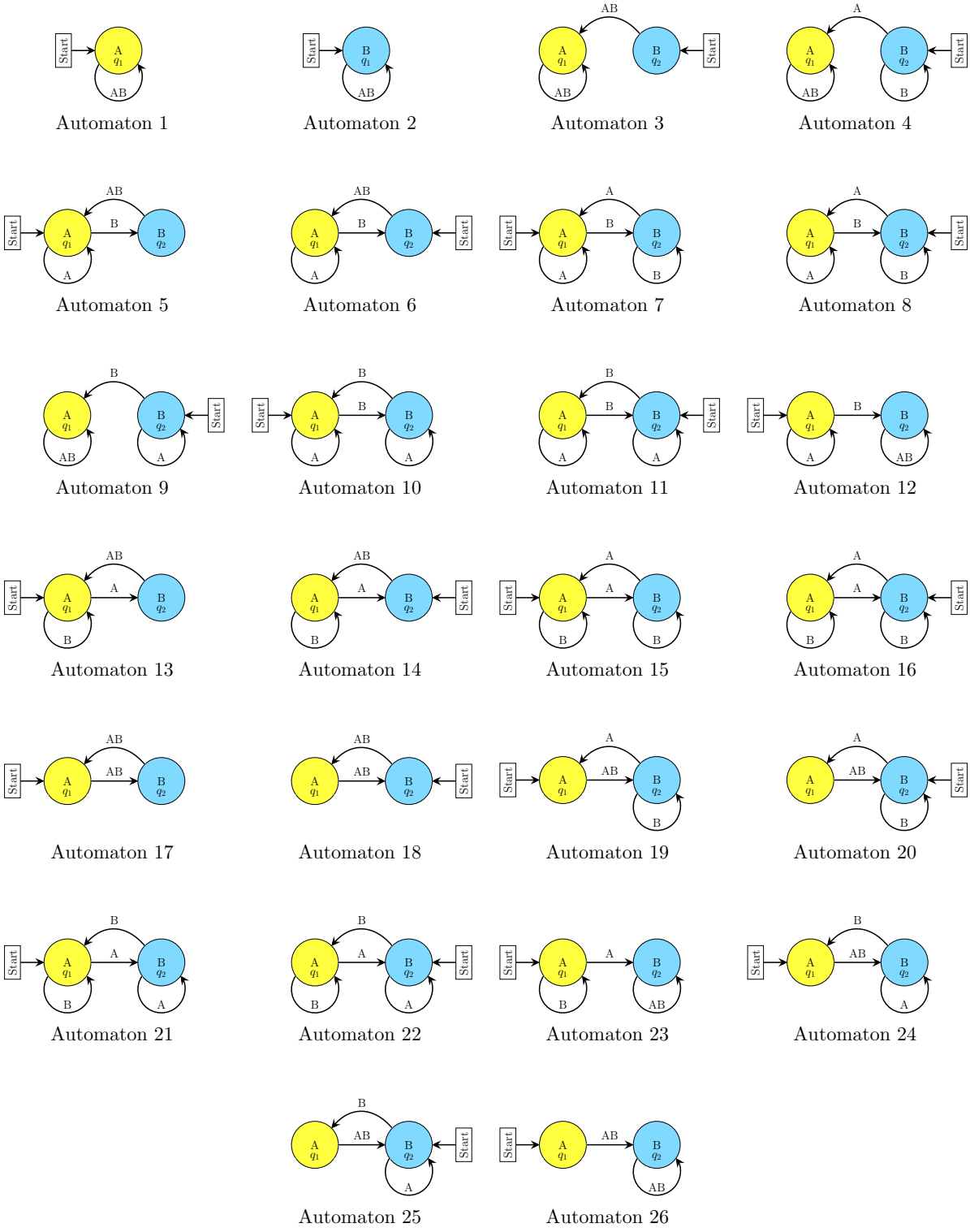
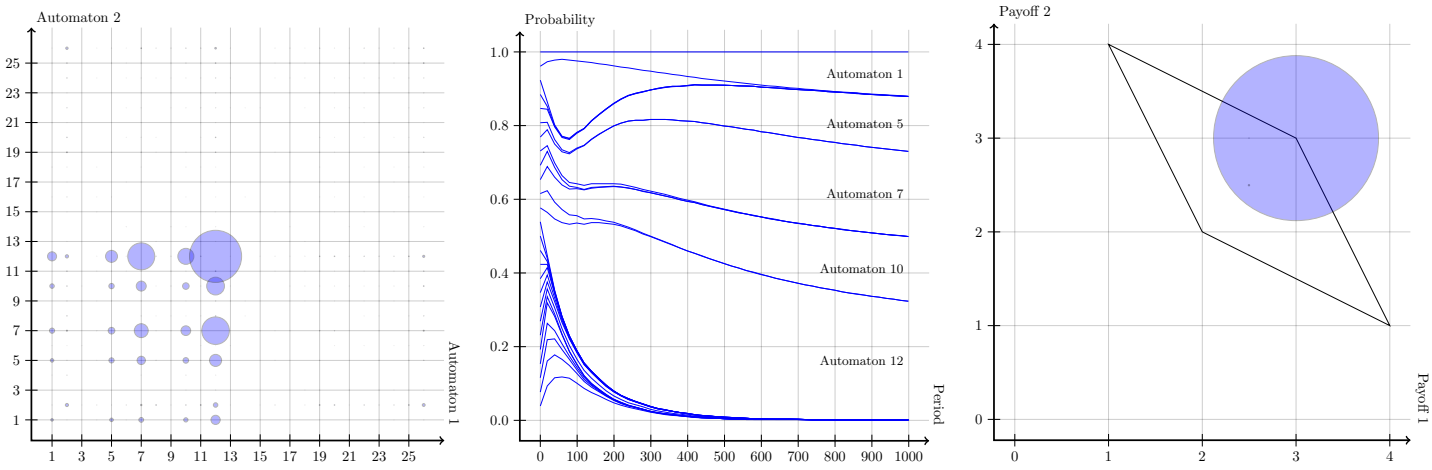
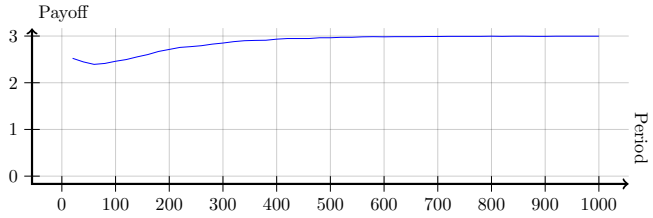


Figure 3: Two-State Automata



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

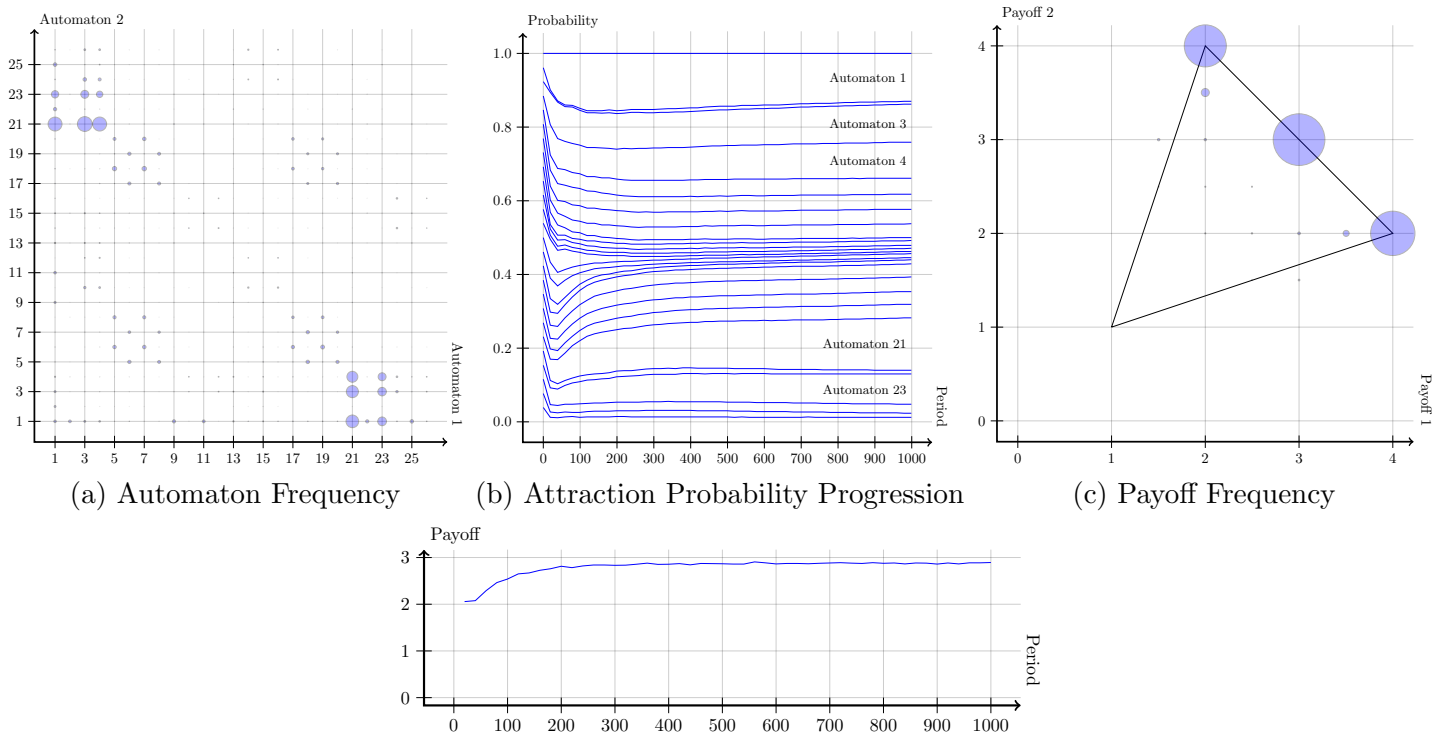


(d) Average Payoff Progression

Figure 4: 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 based on attractions as the game evolves. 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.

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 have strictly positive probabilities: Automaton $1, 5, 7, 10$, 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 over the final 200 periods of the 1,000 period-interaction. The area of the bubbe indicates 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.

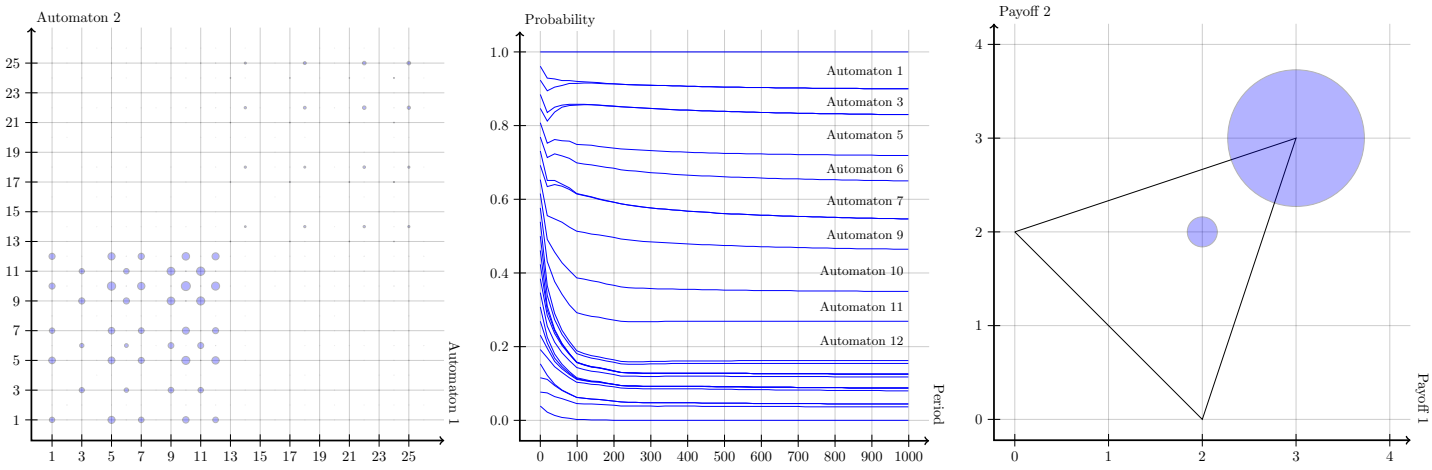


(a) Automaton Frequency (b) Attraction Probability Progression (c) Payoff Frequency

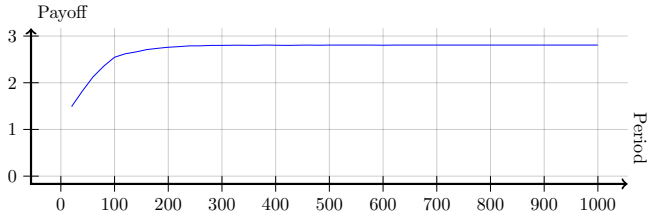
(d) Average Payoff Progression

Figure 5: Battle of the Sexes Game

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). 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, it 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 point corresponds to the situation in which players alternate 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 allows only action-learning. Notice that in spite of observing the $(3, 3)$ in



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

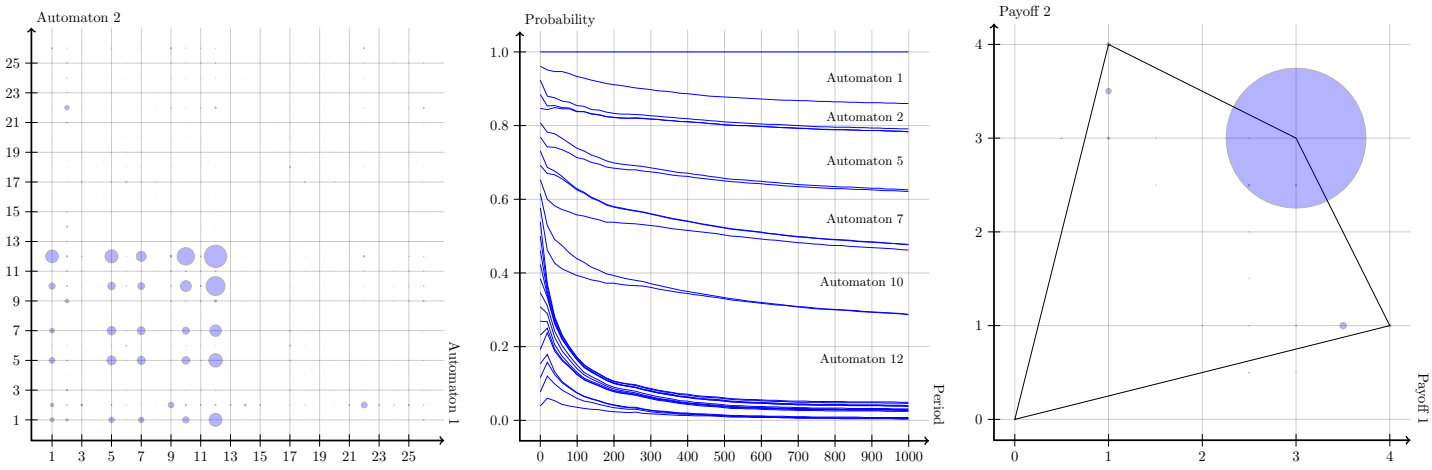


(d) Average Payoff Progression

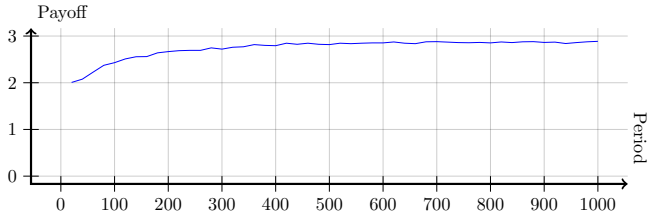
Figure 6: Stag-Hunt Game

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 number 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 the Pareto dominant equilibrium. Plots (a) and (b) in Figure 6 suggest that there is weak convergence to a small set of automata. In contrast, Figure 6(c) suggests that there is convergence to a specific payoff combination: the Pareto dominant payoff of $(3, 3)$. The payoff combination (2,2) also appears on a much smaller scale and arises from an alternation of the two pure-strategy Nash equilibria. Unlike the Prisoner's Dilemma game, in which agents play strategies that punish defectors, the Stag-Hunt 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



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



(d) Average Payoff Progression

Figure 7: Chicken Game

small set of automata. 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-than-average payoffs for each of the players when alternating between the pure-strategy Nash equilibria. The results in plots (a) and (b) of Figure 7 look similar to those in the Prisoner's Dillemma, where game-play is converging to a small set of automata attaining the cooperative outcome of (A, A). This observation is confirmed in Figure 7(c), which shows 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 below 3.

5 Discussion

In Figure 8, the predictions of the modified EWA-lite model are validated with experimental (human) data and compared to the predictions of the (action)-EWA-lite benchmark. Mathevet and Romero (2012) provide experimental data on the four games reported in Section 4. The data

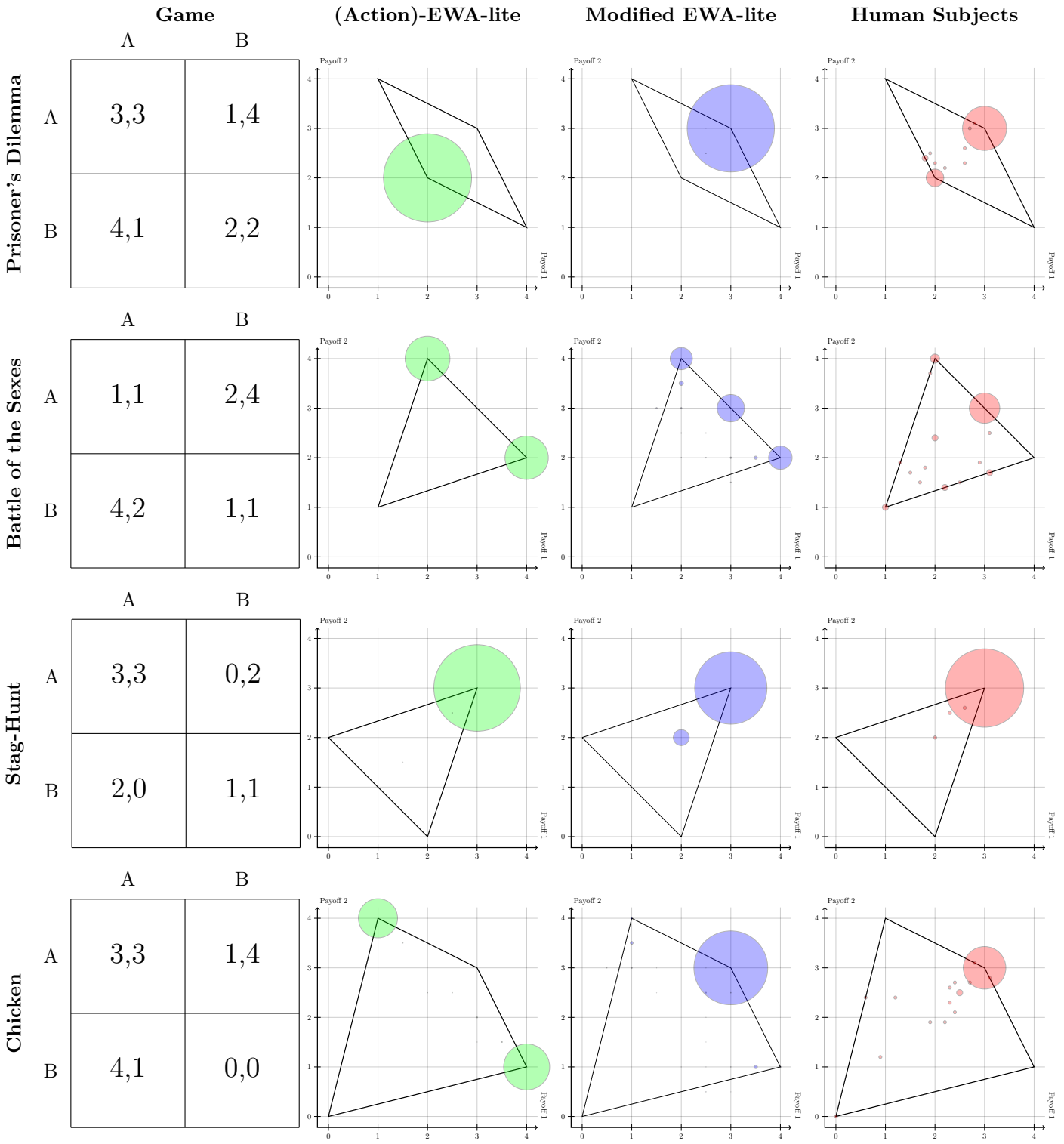


Figure 8: We compare simulations of the (action)-EWA-lite benchmark model and the modified EWA-lite model with human data from the experiments of Mathevet and Romero (2012). The simulations assume that $\lambda = 5$, and the payoffs are averaged over the last 200 periods (periods $800 - 1,000$), while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

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 equilibria in the Battle of the Sexes game,[11] 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 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 from 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. Such strategic behaviors cannot be incorporated easily in an action-learning model.

Hanaki, Sethi, Erev, and Peterhansl (2005) ran simulations with a reinforcement-learning model over a population of agents using two-state automata. The population of 1,000 agents started by playing a pre-experimental phase consisting of, at least, 20,000 periods. The players were randomly matched until the population-averaged attractions converged. Once the population attractions converged, players were randomly selected from the population to play in a fixed pair. However, the incorporation into our model of a belief-based component to complement reinforcement learning, expedites convergence while dispensing of the pre-experimental phase. Notice that 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, implementing only the reinforcement component, keeps a strong strategy unaffected in terms of attraction-weights until it gets selected. The timelines of attraction probabilities and average payoffs in Figures 4-7 suggest that the convergence in these simulations is, indeed, relatively fast. The timelines show that the simulations do not change drastically after the first 200 or 300 periods. Given that the cluster length is 20 periods, this suggests that the simulations are converging within the first ten to fifteen attraction updates.

[11] The fact that observations reside only in the vertex (2,4), and not (4,2), is a result of how players' roles were assigned. For more details, see Mathevet and Romero (2012).

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. Automaton 2 and Automaton 22 also occur with strictly 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 for. As a first step in that direction, we propose a modification of the EWA-lite model to accommodate 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.

There are many directions for future research. A stringent assumption imposed on the players is that their strategies need to be updated simultaneously. Although such a venue could be implemented with explicit communication of the players or via a reliable communication device and enforceable commitment contracts, we would like to endow the players with the ability to update their strategies with an idiosyncratic probability.

In addition, one might test the susceptibility of the results to small amounts of errors. In this study, we 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

[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.

surrounding environment. Thus, agents often 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, 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 not only where each agent's outcome depends 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.

Erev, Ido, and Ernan Haruvy. "Learning and the Economics of Small Decisions." In *The Handbook of Experimental Economics*, edited by John H. Kagel, and Alvin E. Roth, Princeton University Press, 2012, volume 2.

Erev, Ido, and Alvin E. Roth. "On simple reinforcement learning models and reciprocation in the prisoner dilemma game." In *The Adaptive Toolbox*, edited by Gerd Gigerenzer, and Reinhard Selten, Cambridge MIT Press, 2001.

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.

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 that:

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, 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,

$$R_i(t) = \frac{1}{T}\sum_{a \in h} g_i(a) = g_i(t), \tag{3}$$

where $g_i(t)$ is the payoff to player i in period t. 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

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

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

$$p(s_i, s_{-i}, a(t)) = \frac{B((s_i, s_{-i}), a(t))}{\sum_{r \in S_{-i}} B((s_i, r), a(t))} = I\left(a_{-i}, a_{-i}(t)\right).$$

Hence, we get that

$$E_i^j(t) = \sum_{s_{-i} \in S_{-i}} \pi_i(s_i^j, s_{-i}) \cdot p(s_i, s_{-i}, a(t)) = g_i\left(a_i^j, a_{-i}(t)\right). \tag{4}$$

Therefore, $R_i(t) = E_i^j(t) = g_i(a(t))$ for $a_i^j = a_i(t)$.

From equations (3) and (4), we get that the attention function is,

$$\delta_i^j(t) = \begin{cases} 1 & \text{if } g_i(a_i^j, a_{-i}(t)) \geq g_i(t) \\ 0 & \text{otherwise,} \end{cases} \tag{5}$$

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,

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

which is simply the frequency with which action profile a_{-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 $p\left(s_i, s_{-i}, a(t)\right) = I\left(a_{-i}, a_{-i}(t)\right)$ is the same as the term $r_i^k(t)$ from Camerer, Ho, and Chong (2007). Therefore, the surprise index is also the same:

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

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

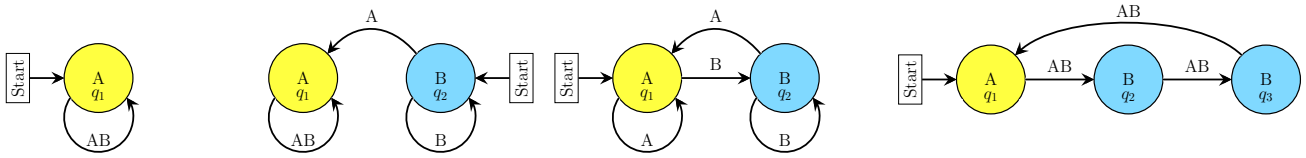
$$\begin{aligned} A_i^j(t) &= \frac{\phi_i(t) \cdot N(t-1) \cdot A_i^j(t-1) + \delta_i^j(t) \cdot E_i^j(t) + I(s_i^j, s_i(t)) \cdot R_i(t)}{\phi_i(t) \cdot N(t-1) + 1} \\ &= \frac{\phi_i(t) \cdot N(t-1) \cdot A_i^j(t-1) + \delta_i^j(t) \cdot g_i(a_i^j, a_{-i}(t))}{\phi_i(t) \cdot N(t-1) + 1}. \end{aligned}$$

This is the equation of the (action)-EWA-lite model in Camerer, Ho, and Chong (2007).[14] 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 and 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. □

[14]Note that the term $(1 - \delta_i^j(t)) \cdot I(a_i^j, a_{-i}(t))$ is always 0 when using the self-tuning functions.

Complexity

In the exercise studied, we considered one-state and two-state finite automata in order to limit the computational burden of the simulations and reduce the automaton-complexity in accord with bounded rationality. Abreu and Rubinstein (1988) defined 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. Thus, we would rather 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 $R(M_i)$. 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.[15] 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.



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

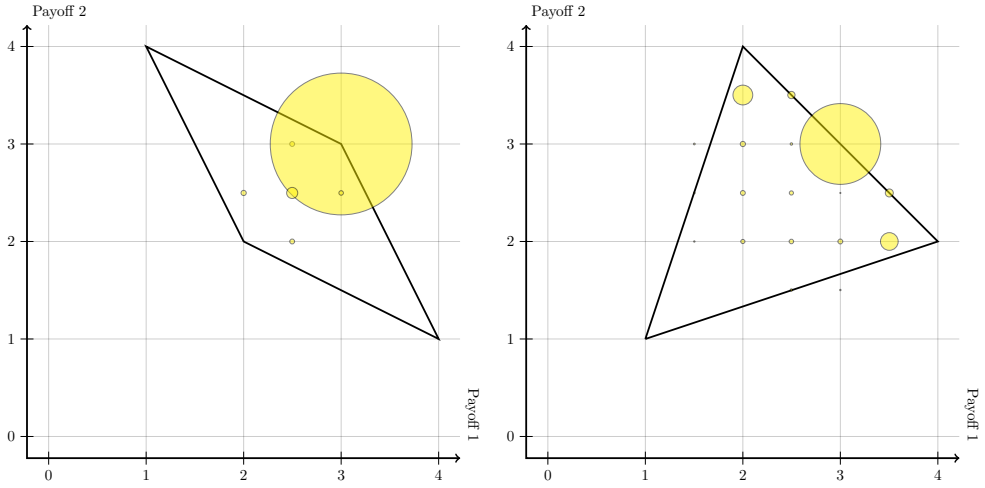
Figure 9: Automaton-Complexity

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 above, of less than or equal to 4, $(R(M_i) \leq 4)$. This set consists of 228 automata ranging from one state-automata to four-state automata. Note that the set also includes all two-state automata. The simulations using the larger set of automata are displayed in

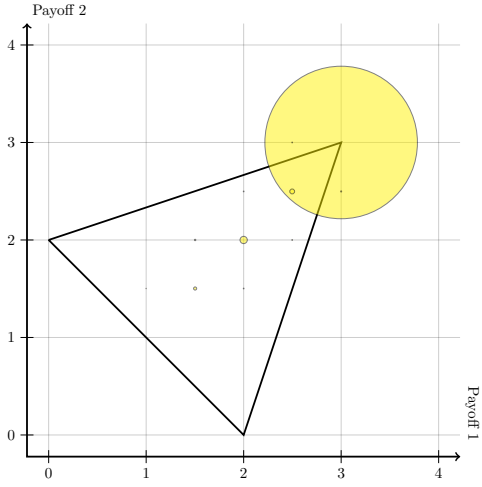
[15] Alternatively, one could use the measure of complexity suggested by Banks and Sundaram (1990), where an automaton M_i is more complex than M_j if M_i is at least as complex as M_j along one of the twin dimensions of transitional complexity and size, and strictly more complex along the other. Such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

Figure 10. The plots preserve the trends exhibited in the earlier results with two-state automata, but are more noisy. The additional noise can be attributed to the bigger cardinality of the strategy space, which leads to slower convergence.

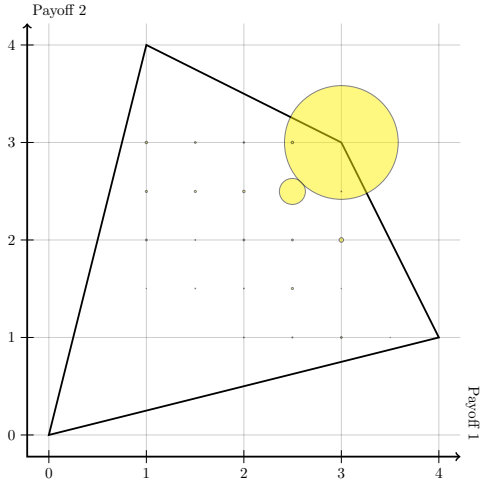


(a) Prisoner's Dilemma Game

(b) Battle of the Sexes Game



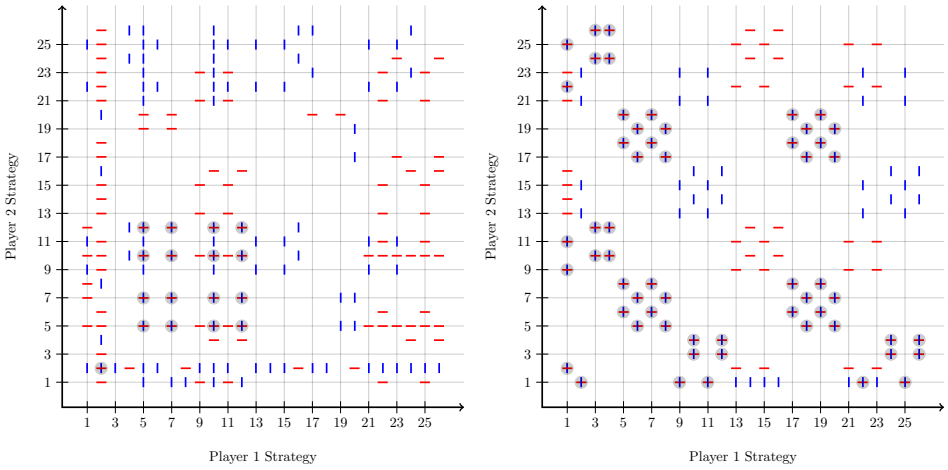
(c) Stag-Hunt Game



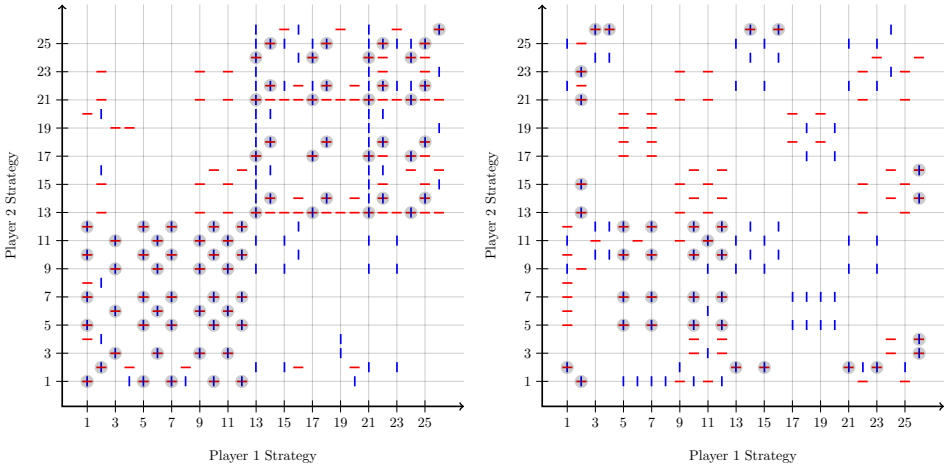
(d) Chicken Game

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

Best Responses and Nash equilibria



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



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

Figure 11: 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.