k-Level Reasoning: A Dynamic Model of Game Learning

Theodore Evans
This project performed in collaboration with David Lecutier

May 14, 2013

Abstract

After outlining the relevant key concepts of classical game theory, we define a deterministic simplification of Camerer and Ho's Experience Weighted Attraction learning model, which will serve as the basis for dynamical simulations of game learning. We propose an extension of this model, implementing depth of strategic reasoning, following reasoning employed in the game theoretic p-beauty contest.

Applying this model to various games, from the traditional game of *Rock, Paper, Scissors* and its broken symmetry variations, to large randomly generated games with well-defined statistical properties. In each case, we contrast the stability and predictability of learning dynamics observed, quantified by mathematical techniques borrowed from the field of non-linear dynamics.

We consistently observe dramatic changes in behaviour in particular games as a result of this extension, but find little net change when these results are averaged over large ensembles of random games. Under this model, we observe a small but significant net change in predictability of long-term behaviour, within parameters feasibly approximating real-world systems, at a depth of strategic thinking feasibly approximating that employed by human decision-makers.

1 Introduction

Game theory was formalised by von Neumann and Morgenstern in their work, *Theory of Games and Economic Behavior* [1], as a mathematical framework for describing and analysing strategic decision-making scenarios. This initial formulation (which we will refer to as *classical game theory*) is generally concerned with finding equilibrium states, should they exist, in scenarios involving a 'one-shot' interaction between two or more perfectly rational players. As such, this is an inherently static theory, making no statements regarding the evolution of a system over time, or to its out-of-equilibrium behaviour.

The field of game theory was extended to applications in dynamical systems with the development of *evolutionary game theory* in 1970 by Maynard Smith [2]. This approach models the process of Darwinian competition, with the time-evolution of a population of agents being governed by the outcomes of repeated interactions, and each interaction proceeding within the framework of classical game theory. It occurs that, in such models, a subset of the equilibrium states predicted by classical game theory correspond to fixed points in the population dynamics. This correspondence between evolutionary models and the classical theory implies that, despite its hyper-rational formulation, game theory may provide a means to predict the emergent behaviour of dynamical systems in which no such premise of rationality is assumed.

Behavioural economics frames the human interactions driving economic systems in the context of the cognitive processes that effect them. Analysing game-theoretic scenarios in this way raises the valuable question: under what circumstances does the decision-making behaviour of 'real' players coincide with that prescribed by the perfect rationality of classical game theory? And consequently, under what circumstances are the solution concepts found through this approach representative of real-world outcomes? Models of adaptive game learning, such as those formulated by Camerer et Al. [3], simulate the process by which a biologically plausible agent comes to behave in a particular way when confronted by a game-theoretic scenario.

Galla et Al. apply a modified version of Camerer and Ho's 1999, *Experience Weighted Attraction* (EWA) learning model, to systems of complicated games [4] with many potential equilibrium states [5]. The predictability of the resultant dynamics then gives an indication to the applicability of an equilibrium-based approach to predicting asymptotic behaviour in such systems. Since the notion of predictability is readily quantifiable when treated in the context of non-linear dynamics, we follow the example of Galla et Al. in adopting this framework under which to analyse these systems.

In this paper, we propose an extension to the EWA model, implementation *depth of strategic reasoning* based on the concept of the Keynesian beauty contest [6]. This extension is informed by studies of human reasoning in situations where they must pre-empt their opponent's process of reasoning in order to succeed [7] [8]. We contrast the stability and predictability of the dynamics described by this model with that of standard EWA learning, in games as simple as *Rock, Paper, Scissors*, through to large complicated games of the form cited above.

2 Theory

2.1 Game Theory

2.1.1 Normal Form Games

In a two-player *normal form* game, as described in [1], each player may choose one of n possible actions (called *strategies*). The outcome of the game over one round of simultaneous play is then defined by an $N \times N$ *payoff matrix*. For example, in a two-player game of Rock Paper Scissors (RPS), each player may play any one of the three available strategies: rock, paper or scissors. The result of each combination of player and opponent strategies is represented by the matrix in equation (1), with the desirability of the outcome corresponding to the positivity of the corresponding matrix element. In the nomenclature of game theory, these values are a measure of the *utility* received by a player in a given round. In this case, 1 corresponds to winning a round, -1 for losing, and 0 for a draw.

$$\Pi^{RPS} = \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix} \tag{1}$$

where the elements Π_{ij} correspond to utility received by a player for each combinations of their choice of strategy i, and their opponent's strategy j in a given round. In this case, $i, j \in \{1, 2, 3\}$ correspond to $\{\text{rock}, \text{paper}, \text{scissors}\}$. This game is *symmetric*, in that the payoff matrices are identical for both players, however this is not the general case. Normal form games such as this one, in which $\Pi = -\Pi^T$, are called *zero-sum* games, reflecting the fact that there is no net loss or gain of utility between the two players over the course of play.

2.1.2 Nash Equilibria and Mixed Strategies

Classical game theory is generally concerned with determining what combinations of player strategies, if any, would result in an equilibrium state in which no player has any incentive to change their position. The 'incentives' in this case are generated under the premise that each player desires to maximise their utility, and employs perfect rationality in order to do so. These equilibrium states are called *solution concepts*, the most general of which is the *Nash equilibrium*. A Nash equilibrium (NE) is a state in which neither player is able to unilaterally improve on their own expected utility, having full knowledge of the opponent's strategy but no power to change it by any means other than through the modification of their own intended strategy (e.g. through coercion or pre-planned cooperation).

If we make a naive attempt to find a pair of strategies that fulfil the NE condition for RPS, the as-yet incompleteness of this formulation will become apparent. If a player initially intends to play Rock, then her opponent (having perfect knowledge of this intent) may change his intended strategy to Paper in order to maximise his expected payoff when these strategies are put into effect. However, now the player may improve her own expected payoff by switching to Scissors, prompting her opponent to switch to Rock, and so on. It is clear that this process will continue indefinitely, sampling every possible combination of strategies without converging to equilibrium. The important development made by von Neumann and Morgenstern [1] was to introduce the concept of a *mixed strategy*, a discrete probability distribution over all available strategies, from which the played strategy is stochastically sampled. A player's mixed strategy is their state vector, denoted $\mathbf{x} = (x_1, x_2, \ldots x_N)$ in a game with N possible strategies, where

x_i is the probability in a given round of playing strategy $i \in \{1, 2, 3 \dots N\}$. As probability distributions, these mixed strategies obeying the probability axioms [9]:

$$\sum_{i=1}^{N} x_i = 1 \qquad\qquad x_i \in [0, 1] \tag{2}$$

In contrast with the mixed strategy, the individual strategies denoted by i are denoted *pure strategies*. This term also extends to the special case of a mixed strategy in which

$$x_i = \begin{cases} 1; & \text{if } i = j \\ 0; & \text{otherwise} \end{cases} \tag{3}$$

A player with such a strategy profile will play strategy j with absolute certainty. As per the above reasoning, there is clearly no NE for RPS in which both players play according to a pure strategy; there is no *strictly dominant strategy*. In contrast, Nash proved that every normal form game with finite n has at least one mixed strategy Nash equilibrium (MSNE) [10]. For MSNEs on the interior of a player's strategy space $x_i \in (0, 1)$ (i.e. excluding NE with the form (3)) each player's expected utility for playing a given strategy i should be equal for all i. In this state, there is no incentive for either player to modify their mixed strategy. The expected utility associated with strategy i is given by

$$R_i = \sum_j \Pi_{ij} \bar{x}_j \tag{4}$$

where $\bar{x}_j$ is the j^{th} element of the opponent's mixed strategy. To demonstrate this, we can calculate the MSNE for RPS, using the payoff matrix (1). Combining the condition for the MSNE with the normalisation condition (2) trivially gives the Nash equilibrium for RPS as

$$\bar{x}_{\text{rock}} = \bar{x}_{\text{paper}} = \bar{x}_{\text{scissors}} = \frac{1}{3}$$

i.e. strategies chosen totally at random. Since the game is symmetric, this process is identical for determining the mixed strategy of the other player, giving

$$\mathbf{x}\big|_{\text{MSNE}} = \left(\tfrac{1}{3}, \quad \tfrac{1}{3}, \quad \tfrac{1}{3}\right) \tag{5}$$

for both players.

2.1.3 p-Beauty Contest

The *Keynesian beauty contest*, named for John Maynard Keynes [6], is a game theoretic scenario demonstrating the effect of *depth of strategic reasoning* on decision making. The original formulation involves a newspaper competition asking readers to pick out the six most attractive faces from a large set of photographs, the winners being those whose choice best fits the average selection of all other players. The p-beauty contest is a simplified form of this scenario, originally posed by Moulin [11] and first performed experimentally by Nagel et Al. [7], which adequately conveys the essence of the problem.

A sample of individuals are asked to choose a number between 0 and 100, with a prize being awarded to the player whose guess is closest to $p \in (0, 1)$ times the median average of all other chosen numbers. Consider the conventional case where $p = \frac{2}{3}$. Naively, a player might

assume that all other players choose at random, and choose $\frac{2}{3} \cdot 50 = 33$. A more astute player may anticipate that his co-players will adopt this strategy and so choose $\frac{2}{3} \cdot 33 = 22$. Other players may further extrapolate this line of reasoning, recursively anticipating the strategies of their opponents who are themselves anticipating the strategies of their opponents. We signify the recursion depth of this process by the *k-level* of a player, where the (hypothetical) player choosing numbers at random has a k-level of zero, denoted $k(0)$, and the player who chooses as if all other players are $k(0)$ has a k-level of 1. A $k(n)$ player then assumes that all other players play according to $k(n-1)$ reasoning. This is a form of *bounded rationality*, in which this process of recursive anticipation of opponent strategy is truncated (either deliberately or as a result of practical limitations) at a depth of n.

If players are allowed knowledge of their opponent's intentions before they are required to make their choice, the only number that may be picked such that no player has any incentive to change their mind is zero, making this the Nash equilibrium for this system. This is also the number chosen by a $k(n)$ player in the limit $n \to \infty$. We then find a correspondence between the perfect rationality, which prescribes that a player choose the Nash Equilibrium, and the asymptotic limit of this model of bounded rationality. It is important to note, in this case, that perfect rationality only gives the best outcome (results in the maximum utility) if every other player is also employing perfect rationality. In a system where some or all players are not only employing bounded rationality, but doing so to different degrees, the best outcome results from employing a comparable form of bounded rationality oneself. Therefore, in real examples, it is reasonable to assume that the depth of strategic reasoning employed by a player reflects the depth to which he believes his co-players are reasoning, as much as it does his cognitive ability to out-reason them to a greater depth.

2.1.4 Correlated Random Games

In section 2.1.1, the elements of the payoff matrix π_{ij}^{RPS} were set such that the outcome of game play corresponds to the predefined rules of RPS. It was noted that this is a *zero-sum* game, indicating a correlation between the payoff matrices of player and opponent, in this case $\pi_{ij} = -\bar{\pi}_{ji}$. We also noted that RPS was a *symmetric* game, where the two payoff matrices are identical, $\pi_{ij} = \bar{\pi}_{ij}$. In order to follow on from the results for large, random games given in [4], it shall become necessary to generate correlated pairs of non-identical matrices of arbitrary size, A and B, with normally distributed matrix elements a_{ij} and b_{ij}, such that

$$\begin{aligned} \mathbb{E}(a_{ij}) = \mathbb{E}(b_{ij}) &= 0 \\ \mathbb{E}(a_{ij}^2) = \mathbb{E}(b_{ij}^2) &= 1 \\ \mathbb{E}(a_{ij} b_{ji}) &= \Gamma \end{aligned} \tag{6}$$

where $\mathbb{E}(X)$ is the expectation value of X and $i, j \in \{1, 2 \dots N\}$. The correlation parameter $\Gamma \in [-1, 1]$ then gives the corresponding correlation properties of the two matrices

$$\Gamma = \begin{cases} -1 & \text{; matrices totally anticorrelated,} \quad \pi_{ij} = -\bar{\pi}_{ji} \\ 0 & \text{; matrices totally uncorrelated} \\ +1 & \text{; matrices totally correlated,} \quad \pi_{ij} = \bar{\pi}_{ji} \end{cases}$$

with values in the interval $[-1, 1]$ interpolating between these special cases. To generate matrices with such properties, we define two stochastic variables, U and V, sampled from a normal

distribution $\mathcal{N}(0,1)$. The matrix elements are then given by

$$a_{ij} = U$$
$$b_{ji} = \Gamma U + \sqrt{1-\Gamma^2}V$$

giving A and B the required correlation properties. The classical game theory interpretation for the parameter Γ is as an (inexact) measure of how competitive a game is. Zero-sum games ($\Gamma = -1$) are referred to as *strictly competitive*, in the sense that one player may only increase their utility by reducing that of their opponent, i.e. all interactions have a win-lose form. Games with $\Gamma > -1$ allow the possibility of non-competitive play, wherein some combinations of strategies may correspond to win-win (or lose-lose) outcomes.

2.2 Learning Dynamics

2.2.1 Experience Weighted Attraction

The Experience Weighted Attraction (EWA) learning model is an algorithmic approach to simulating the evolution of players' mixed strategies as they learn to repeatedly play a normal form game against an opponent. It is a model of individual adaptation, as opposed to the group adaptation seen in dynamic evolutionary game theory models [12][13]. In Camerer and Ho's original formulation of EWA learning [3], a player's mixed strategy is generated from a set of predispositions, called *attractions*, towards each available strategy. These attractions are updated upon each round of play, based on a combination of factors reflecting the expected utility associated with the corresponding strategies, such that the update rule enacts the player's desire to maximise their expected utility when playing the game in following round.

For a round of play denoted τ, the strategy sampled from a player's mixed strategy $\mathbf{x}$ is denoted $s(\tau)$. Corresponding quantities and actions attributed to the opposing player are denoted with an overbar, such that all formulae are generalised to apply to any given player in a one-on-one round of game play. Under this notation, the strategy played in round τ by the opposing player is $\bar{s}(\tau)$, having been sampled from their mixed strategy $\bar{\mathbf{x}}$, with a probability $\bar{x}_i$. A round of play for a 2-player game defined by a symmetric payoff matrix Π then proceeds as follows:

1.) A player samples strategy $s(\tau)$ from $\mathbf{x}(\tau)$. Simultaneously, their opponent samples strategy $\bar{s}(\tau)$ from $\bar{\mathbf{x}}(\tau)$.

2.) Player and opponent receive utility for this round equal to $\pi_{s\bar{s}}$ and $\bar{\pi}_{\bar{s}s}$ respectively.

3.) The attraction profile for each players is updated according to the rules discussed below.

4.) The mixed strategy for each player is updated as a function of their attraction profile.

5.) This process is repeated using the new mixed strategies $\mathbf{x}(\tau+1)$ and $\bar{\mathbf{x}}(\tau+1)$

In [3], each player's attraction profile, $\mathbf{Q}(\tau) = (Q_1(\tau), Q_2(\tau) \dots Q_n(\tau))$ is updated after every round of play, with the new attractions $Q_i(\tau+1)$ being a linear superposition of the current value of that attraction, the utility received for playing the corresponding strategy (if it was played) and the hypothetical utility that would have been received had the corresponding strategy been played (if it was not). A simplified version of this model, as employed in [14][4]

may be obtained from [3] through a suitable choice of parameters, giving the attraction update rules

$$Q_i(\tau+1) = (1-\alpha)Q_i(\tau) + \pi_{i\bar{s}(\tau)} \tag{7}$$

where the parameter $\alpha \in [0,1]$ corresponds to the discounting rate of previous attraction values in favour of new values determined by the second right-hand term. The state of the attraction profile in a round $\tau' = \tau - N$ is exponentially discounted with respect to its effect on the state of the attractions in round τ, by a factor $(1-\alpha)^N$. A player's mixed strategy is then obtained by application of a logit rule to the attraction profile $Q_k(\tau)$ such that

$$x_i(\tau) = \frac{e^{\beta Q_i(\tau)}}{\sum_{j=1}^{N} e^{\beta Q_j(\tau)}} \tag{8}$$

where i, j here simply index mixed strategy components, without referring to a particular round, where $\beta \in [0,\infty]$ is a parameter controlling the intensity of selection of strategies with higher corresponding attractions. In the $\beta = 0$ limit, all strategies have equal probabilities of being played: the player acts at random. In the $\beta \to \infty$ limit, only the strategy with the largest corresponding attraction is played, as in (3). Under this rule, the mixed strategy reflects the relative magnitudes of the attractions, whilst conforming to the normalisation conditions (2).

We proceed by reformulating this update rule under the adiabatic approximation, wherein it is assumed that the time scale of interactions between players is much shorter than the time scale of adaptation of each player's mixed strategy. A time variable t is introduced, and the strategy and attraction profiles are assumed to be stationary over the course of play. The update to the attraction profile after a batch of T rounds have been played is then

$$Q_i(t+\Delta t) = (1-\alpha)Q_i(t) + \frac{1}{T}\sum_{t'=\tau}^{\tau+\Delta\tau} \pi_{i\bar{s}(t')} \tag{9}$$

where $\Delta t = T\Delta\tau$. In the limit $T \to \infty$, the final term (the *utility expectation* term) becomes

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t'=t}^{t+\delta t} \pi_{i\bar{s}(\tau')} = \lim_{T\to\infty} \sum_{i=1}^{T} \pi_{ij}\frac{\bar{T}_j}{T} \;=\; \sum_{j=1}^{T} \pi_{ij}\bar{x}_j(t)$$

where $\bar{T}_i$ is the number of times the opponent sampled the strategy i over T rounds of play. We have then used the general definition of the probability $\bar{x}_i(t)$ to retrieve the classical game theory expectation value for playing strategy i given in (4). Under this approximation, attractions are no longer updated based on stochastic sampling of strategy profiles; the game dynamics are completely deterministic and each player is given perfect knowledge of their opponent's strategy profile. The deterministic attraction update rule is then

$$Q_i(t+\Delta t) = (1-\alpha)Q_i(t) + \sum_{j=1}^{N} \pi_{ij}\bar{x}_j(t) \tag{10}$$

Substituting this into the logit rule (8) gives the deterministic strategy update rule,

$$x_i(t+\Delta t) = \frac{e^{\beta(1-\alpha)Q_i(t)}e^{\beta R_i(t)}}{\sum_{j=1}^{N} e^{\beta(1-\alpha)Q_j(t)}e^{\beta R_j(t)}} \qquad\qquad R_i(t) \equiv \sum_{j=1}^{N} \pi_{ij}\bar{x}_j(t) \tag{11}$$

which is equivalent to the discrete map

$$x_i(t+1) = \frac{x_i(t)^{(1-\alpha)} e^{\beta R_i(t)}}{\sum_j x_j(t)^{(1-\alpha)} e^{\beta R_j(t)}} \tag{12}$$

The $(t+1)$ and $(t+\Delta t)$ are used interchangeably under the premise of the adiabatic approximation, that attraction and strategy profiles are stationary between time t and $t+\Delta t$. Unless otherwise specified, the closed map 12 is the form used to govern the dynamical simulations described in this paper, supplanting any requirement to deal with attractions or attraction profiles.

Rewriting this update rule using the standard definition of a derivative with respect to t, and taking the limit as $\delta t \to 0$, gives the continuous-time differential equations

$$\frac{\dot{x}_i}{x_i} = \beta \left(\sum_k \pi_{kl} \bar{x}_l - \sum_{kl} x_k \pi_{kl} \bar{x}_l \right) - \alpha \left(\ln \bar{x}_i - \sum_k \bar{x}_k \ln \bar{x}_k \right) \tag{13}$$

and its conjugate under the exchange $\mathbf{x} \mapsto \bar{\mathbf{x}}$ and $\Pi \mapsto \bar{\Pi}$. These are referred to as the Sato-Crutchfield equations [15]. It may be noted that the form of these equations is analogous to the replicator-mutator equations that regularly appear in evolutionary game theory dynamics, with analogous behaviour emerging accordingly (for instance, comparing the results given in [14] and [12]).

2.2.2 k-Level Reasoning

As a multi-player 'one-shot' game, the format of the p-beauty contest described in section 2.1.3 is quite different to the system of iterated two-player interaction modelled by EWA learning. However, we can attempt to introduce into this model an element of k-level reasoning, by considering it in terms of recursive anticipation of an opponent's strategy. In this two-player case, this recursion has the form: 'I think that you think that I think etc...', with a $k(n)$ player truncating this sequence at a depth n. Although we might define a $k(1)$ player as a player who assumes that their opponent plays at random, by analogy with the p-beauty contest, we would like our $k(1)$ to recover the deterministic EWA map (12), which does not make this assumption. Rather, the player has a perfect knowledge of their opponent's current mixed strategy, and uses this information to improve their expected utility in the following round. Therefore, we define a $k(1)$ player as one who assumes that their opponent's strategy does not change between rounds t and $t+1$, and therefore updates their strategy using a map that is a function of $\bar{\mathbf{x}}(t)$, giving the form (12) as before. For the sake of completeness, we may define a $k(0)$ player as one who does not take their opponent's strategy into account at all (or equivalently, assumes the opponent plays at random), updating their mixed strategy components according to the average value of elements in the corresponding row of their payoff matrix.

Contestants in the p-beauty contest anticipate the group average by applying an understanding of the reasoning process employed by their co-players. Analogously, a player in our model may apply an 'understanding' of the process by which their opponent changes their mixed strategy in response to play, i.e. the EWA map. A $k(2)$ player would then employ a map that is itself a function of the opponent's map

$$\mathbf{x}^{(2)}(t+1) = \mathbf{x}^{(2)}(t+1) \left\{ \bar{\mathbf{x}}^{(1)}(t+1) \right\}$$

where $\bar{\mathbf{x}}^{(1)}(t+1)$ is the standard EWA map for the opponent. A $k(3)$ player will assume that their opponent is a $k(2)$ player and is thus employing a map which is itself a function of their own map, and so on. Higher levels of reasoning are built up from (12) through a recursive substitution of strategy with map for player and opponent, such that the map for a $k(n)$ player is given by

$$x_i^{(n)}(t+1) = \frac{1}{\mathcal{Z}^{(n)}} \cdot x_i^{(n)}(t)^{1-\alpha} \cdot \exp\left\{ \beta \sum_{j=1}^{N} \pi_{ij} \bar{x}_j^{(n-1)}(t+1) \right\} \tag{14}$$

for $n \in \{2, 3, 4 \ldots\}$, where

$$\mathcal{Z}^{(n)} \equiv \sum_k x_k^{(n)}(t)^{1-\alpha} \cdot \exp\left\{ \beta \sum_{l=1}^{N} \pi_{kl} \bar{x}_l^{(n-1)}(t+1) \right\}$$

and $x_i^{(1)}(t+1)$ is given by equation (12)*.

2.2.3 k-Level Reasoning, an alternative approach

The update rules for higher k-level reasoning may also be formulated as a form of *gradient ascent* algorithm. Gradient ascent is an algorithmic approach to finding a local maximum of some function $f\{\mathbf{x}(t)\}$, starting from point $\mathbf{x}(0)$, by taking steps in the direction of the gradient of the function at this point, $\nabla f\{\mathbf{x}(0)\}$. At step t of the algorithm, the components of the state $\mathbf{x}(t)$ are updated according to

$$x_i(t+1) = x_i(t) + \eta \frac{\partial f\{\mathbf{x}(t)\}}{\partial x_i(t)}$$

where η is some small parameter governing the step size. The attraction update rule of the EWA model (7) can be thought of as a modification of this algorithm, where the learning parameter β is equivalent to the step size η, and the function to be maximised is the expectation value of a player's utility

$$\mathbb{E}_{t,t'} = \sum_{ij} x_i(t) \Pi_{ij} \bar{x}_j(t')$$

which is differentiated with respect to the corresponding strategy component. The memory loss parameter α is also introduced to discount the current attraction profile, giving

$$Q_i(t+1) = (1-\alpha) Q_i(t) + \beta \frac{\partial \mathbb{E}_{t,t'}}{\partial x_i(t)}$$

In the case of $k(1)$ EWA, a player assumes their opponent's strategy to be static, $t = t'$, and differentiating $\mathbb{E}_{t,t}$ with respect to the strategy components $x_i(t)$ simply reproduces the deterministic EWA update rules (10) and (12). If we allow a player, as before, to pre-empt their opponent's strategy in the next round and make use of this prediction in adapting their own strategy, then this update rule resembles a gradient ascent performed on a shifting landscape of expectation values. The gradient of this landscape at a point in phase space corresponding to

*The notation used in (14) is equivalent to that defining (12), we adopt this alternative style only for the sake of clarity in representing complicated exponents.

the player's current strategy, and their opponent's *predicted* strategy in the next round is then $\mathbb{E}_{t,t+1}$, of which the individual components are

$$\frac{\partial \mathbb{E}_{t,t+1}}{\partial x_i(t)} = \sum_j \Pi_{ij} \bar{x}_j(t+1) \cdot \left\{ 1 + \beta \left(\sum_k \bar{\Pi}_{jk} x_k(t) - \sum_{lk} \bar{x}_l(t+1) \bar{\Pi}_{lk} x_k(t) \right) + \mathcal{O}(\beta^2) \right\} \quad (15)$$

which would then be substituted into the place of $R_i(t)$ in (12). This is a fairly complicated expression, the interpretation of which we will not go into in this paper. The recursive differentiation of the update rule in this approach may be truncated to an arbitrary depth, as before, with each recursion producing a term in higher order of β. The two approaches are not mutually exclusive, as this method also elicits the recursive substitution of update rules employed to generate the update rules (14). Given that the entirety of (15) has a multiplicative factor of β in the final update rule, for $\beta \ll 1$ we can reasonably disregard terms of $\mathcal{O}(\beta)$ in (15), thus recovering the previous form of k-level update rule (14).

2.3 Linear Stability Analysis

The state of the system for a given game in the EWA learning model is described by the mixed strategy vectors of the two players, whose trajectories through phase space are governed by a discrete map. Treating such a system in the context of non-linear dynamics, it is possible to examine the dynamical behaviour of the system near a fixed point (FP) in its phase space. For a discrete map defined by a vector function

$$\begin{aligned} \mathbf{x}(t+1) &= \mathbf{f}\left(\mathbf{x}(t), \bar{\mathbf{x}}(t)\right) \\ \bar{\mathbf{x}}(t+1) &= \mathbf{g}\left(\mathbf{x}(t), \bar{\mathbf{x}}(t)\right) \end{aligned} \quad (16)$$

where

$$\mathbf{x}(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{pmatrix}$$

fixed points of the dynamics, denoted $\mathbf{x}(t) = \mathbf{x}^*$, $\bar{\mathbf{x}}(t) = \bar{\mathbf{x}}^*$, are defined such that

$$\mathbf{f}(\mathbf{x}^*, \bar{\mathbf{x}}^*) = \mathbf{x}^* \qquad \mathbf{g}(\mathbf{x}^*, \bar{\mathbf{x}}^*) = \bar{\mathbf{x}}^* \quad (17)$$

To examine the stability of the FP, we examine the behaviour of small perturbations away from it, $\delta\mathbf{x}(t)$ and $\delta\bar{\mathbf{x}}(t)$, such that

$$\begin{aligned} \mathbf{x}(t) &= \mathbf{x}^* + \delta\mathbf{x}(t) \\ \bar{\mathbf{x}}(t) &= \bar{\mathbf{x}}^* + \delta\bar{\mathbf{x}}(t) \end{aligned} \quad (18)$$

The map (9) for $\mathbf{x}(t)$ near the FP is then given by

$$\mathbf{x}(t+1) = \mathbf{f}(\mathbf{x}^* + \delta\mathbf{x}(t), \bar{\mathbf{x}}^* + \delta\bar{\mathbf{x}}(t)) \quad (19)$$

which we expand as Taylor series

$$f(\mathbf{x}^* + \delta\mathbf{x}(t), \bar{\mathbf{x}}^* + \delta\bar{\mathbf{x}}(t)) = \mathbf{f}(\mathbf{x}^*, \bar{\mathbf{x}}^*) + \left.\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right|_{FP} \delta\mathbf{x}(t) + \left.\frac{\partial \mathbf{f}}{\partial \bar{\mathbf{x}}}\right|_{FP} \delta\bar{\mathbf{x}}(t) \quad (20)$$

where (arguments omitted for simplicity)

$$\frac{\partial \mathbf{f}}{\partial \mathbf{x}} \equiv \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \frac{\partial f_1}{\partial x_3} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \frac{\partial f_2}{\partial x_3} \\ \frac{\partial f_3}{\partial x_1} & \frac{\partial f_3}{\partial x_2} & \frac{\partial f_3}{\partial x_3} \end{pmatrix} \tag{21}$$

We have assumed that the dynamics near the FP are linear in $\delta\mathbf{x}(t)$ and $\delta\bar{\mathbf{x}}(t)$ provided these perturbations are small, thus allowing us to discard higher order terms. Combining equations (17–20), we see that

$$\delta\mathbf{x}(t+1) = \left.\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right|_{FP} \delta\mathbf{x}(t) + \left.\frac{\partial \mathbf{f}}{\partial \bar{\mathbf{x}}}\right|_{FP} \delta\bar{\mathbf{x}}(t) \tag{22}$$

and similarly for $\bar{\mathbf{x}}$

$$\delta\bar{\mathbf{x}}(t+1) = \left.\frac{\partial \mathbf{g}}{\partial \mathbf{x}}\right|_{FP} \delta\mathbf{x}(t) + \left.\frac{\partial \mathbf{g}}{\partial \bar{\mathbf{x}}}\right|_{FP} \delta\bar{\mathbf{x}}(t) \tag{23}$$

Equations (22) and (23) can be simultaneously represented in block matrix form

$$\begin{pmatrix} \delta\mathbf{x} \\ \delta\bar{\mathbf{x}} \end{pmatrix}_{t+1} = \left. \begin{pmatrix} \frac{\partial \mathbf{f}}{\partial \mathbf{x}} & \frac{\partial \mathbf{f}}{\partial \bar{\mathbf{x}}} \\ \frac{\partial \mathbf{g}}{\partial \mathbf{x}} & \frac{\partial \mathbf{g}}{\partial \bar{\mathbf{x}}} \end{pmatrix} \right|_{FP} \begin{pmatrix} \delta\mathbf{x} \\ \delta\bar{\mathbf{x}} \end{pmatrix}_t \tag{24}$$

The 6×6 block matrix is the Jacobian matrix of the system, J, evaluated at the fixed point. The vectors $(\delta\mathbf{x}, \delta\bar{\mathbf{x}})_{t,t+1}$ represent the displacement in the phase space of the system from the fixed point $(\mathbf{x}^*, \bar{\mathbf{x}}^*)$ at time steps t and $t+1$. These displacement vectors can be expanded as a linear combination of the eigenvectors of $\mathbf{J}$,

$$\begin{pmatrix} \delta\mathbf{x} \\ \delta\bar{\mathbf{x}} \end{pmatrix}_t = \sum_k c_t^{(k)} \mathbf{e}^{(k)} \qquad\qquad \begin{pmatrix} \delta\mathbf{x} \\ \delta\bar{\mathbf{x}} \end{pmatrix}_{t+1} = \sum_k c_{t+1}^{(k)} \mathbf{e}^{(k)} \tag{25}$$

where $c_{t,t+1}^{(k)}$ are scalar coefficients, and $\mathbf{e}^{(k)}$ is the eigenvector of $\mathbf{J}$ with a corresponding eigenvalue of $\mu^{(k)}$, such that

$$\mathbf{J} \cdot \mathbf{e}^{(k)} = \mu^{(k)} \mathbf{e}^{(k)} \tag{26}$$

Combining equations (24–26) gives us

$$c_{t+1}^{(k)} = \mu^{(k)} c_t^{(k)} \tag{27}$$

This gives three possibilities for the evolution of the displacement vectors over iterated discrete maps:

(i) $|\mathrm{Re}\{\mu^{(k)}\}| < 1$; perturbations decay exponentially in the direction of $\mathbf{e}^{(k)}$. If this is true for all k, then the fixed point is stable.

(ii) $|\mathrm{Re}\{\mu^{(k)}\}| > 1$; perturbations grow exponentially in in the direction of $\mathbf{e}^{(k)}$. If this is true for any k, then the fixed point is unstable.

(iii) $|\mathrm{Re}\{\mu^{(k)}\}| = 1$; perturbations neither grow nor decay, the displacement in the direction of $\mathbf{e}^{(k)}$ is constant. Such eigenvalues correspond to periodic cycles in the system's behaviour.

A stable fixed point is an example of an *attractor* for the system. Attractors are points or trajectories in a system's phase space that are stable to perturbations, and toward which the dynamics will generally tend. Note that for EWA learning, the Jacobian may be simplified by imposing the normalisation condition (2), such that

$$x_1 + x_2 + x_3 = 1$$

and therefore

$$\frac{\partial}{\partial x_1} = -\frac{\partial}{\partial x_2} - \frac{\partial}{\partial x_3}$$

and similarly for $\bar{\mathbf{x}}$, allowing matrix (21) to be reduced to a 2×2 matrix, and the Jacobian to a 4×4 matrix.

2.4 Lyapunov exponents

A concept that is useful in characterising the predictability of a system's long term behaviour, is that of the *Lyapunov exponent*. Consider a discrete map as before, and take two points in phase space separated initially by an infinitesimal perturbation $\delta\mathbf{x}_0$. Iterating this map with these two sets of nearby initial conditions gives a pair of trajectories which, after t iterations, are separated by a vector $\delta\mathbf{x}_t$. The relative magnitudes of $\delta\mathbf{x}_t$ and $\delta\mathbf{x}_0$ gives a measure of the convergence or divergence of the two trajectories over time.

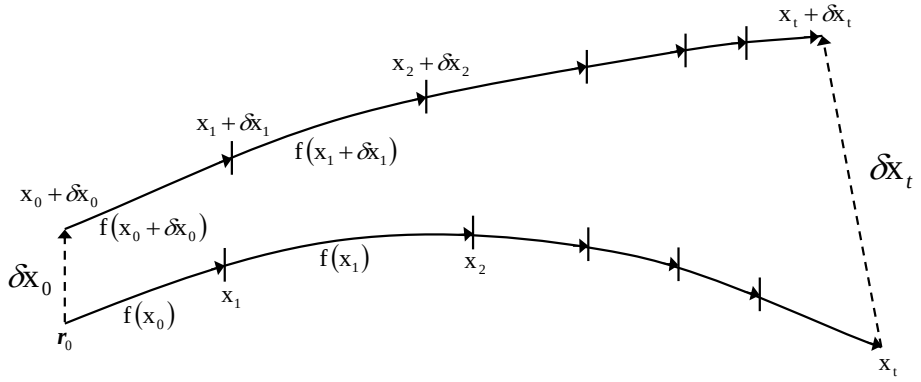


Figure 1: Trajectories diverging from an initial small separation of $\delta\mathbf{x}_0$ under the map $f(\mathbf{x}, t)$

Consider the Jacobian of the map evaluated at each point along one of these perturbed trajectories, as in section 2.3. The magnitude of a perturbation in some direction will pick up a multiplicative factor on each iteration of the map, related to the eigenvalues of the Jacobian at that point. After t iterations, the distance between the trajectories grows or decays exponentially, having the form

$$|\delta\mathbf{x}_t| = e^{\lambda t}|\delta\mathbf{x}_0| \tag{28}$$

where the parameter λ is the largest Lyapunov exponent (LLE). This quantity is given the qualifier "largest" as trajectories in an N-dimensional map will diverge at different rates in different directions, giving a Lyapunov spectrum related to the spectrum of Jacobian eigenvalues. Due to

the exponential nature of the growth (or decay), over a large number of iterations the Lyapunov exponent with the greatest magnitude will obliterate contributions from all others. Given a set of initial conditions located on the attractor of a system, converging nearby trajectories ($\lambda < 0$) indicate a predictable behaviour, in the form of fixed point or limit cycle attractors, whereas diverging trajectories ($\lambda > 0$) are an indication of chaotic, or at least unpredictable, dynamics. To rationalise this, consider a real-world scenario wherein the future state of a system is to be predicted, based on some measurement of initial conditions: If $\lambda < 0$, small deviations between measured and true initial conditions will become irrelevant, as trajectories from either will ultimately lead to identical behaviour. However, if $\lambda > 0$, even the most infinitesimal error in measurement of initial conditions (the likes of which are unavoidable) will eventually lead to a divergence of predicted and actual behaviour. This *sensitivity to initial conditions* is widely regarded as a defining characteristic of chaos behaviour, although there is by no means a one-to-one correspondence between the two concepts. As we shall see later, an attractor may have $\lambda > 0$ without being considered chaotic.

The technique used to numerically calculate largest Lyapunov exponents in this paper proceeds as follows:

1. Starting for arbitrary initial conditions (unless specified otherwise, $x_i(0) = \frac{1}{N}$), the map is iterated over t_0 rounds (where $t_0 \sim 10^4$) to allow the system to equilibrate and settle onto an attractor.

2. The trajectories of both players are perturbed by identical vectors of magnitude $|\delta \mathbf{x}_0|$ in a random direction, where $|\delta \mathbf{x}_0| \ll 1$ is a constant. States on the edge of the simplex are perturbed toward the interior, so as not to violate (2). For each random perturbation, the following steps are iterated over L rounds (where $L \sim 10^2$):

   (a) The appropriate update rules are applied to both parent trajectories and their perturbed counterparts for one round.

   (b) The total (6-dimensional) phase space distance between parent and perturbed trajectories, $|\delta \mathbf{x}_t|$ is calculated.

   (c) The ratio of this distance to the magnitude of the original perturbation, $\frac{|\delta \mathbf{x}_t|}{|\delta \mathbf{x}_0|} \equiv \phi_t$ is calculated.

   (d) The vector separating parent and perturbed trajectories is again normalised to a magnitude of $|\delta \mathbf{x}_0|$ without changing its direction.

3. Step 2 is iterated n_p times (where $n_p \sim 10^2$), with a LLE for the n-th iteration given by

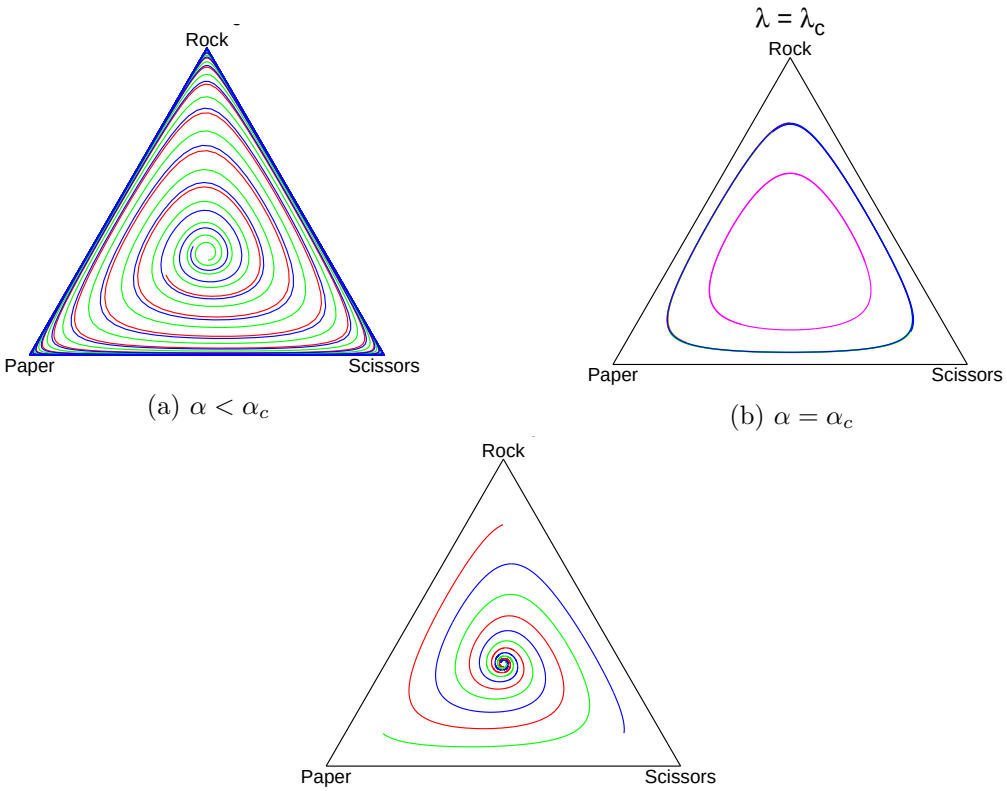
$$\lambda_n = \frac{1}{L} \ln \left(\prod_{t=nL}^{(n+1)L} \phi_t \right)$$

4. The LLE averaged over the whole attractor is then given by

$$\lambda_n = \frac{1}{n_p} \sum_{n=0}^{n_p} \lambda_n$$

3 Results and Discussion

3.1 Rock, Paper, Scissors



(a) $\alpha < \alpha_c$

(b) $\alpha = \alpha_c$

(c) $\alpha > \alpha_c$

Figure 2: An illustration of the dynamics of $k(1,1)$ EWA learning in a game of repeated RPS, over 10,000 rounds of deterministic play at different values of α relative to a critical value α_c. Trajectories begin at arbitrary initial conditions and proceed in an anticlockwise direction. The intensity of selection is $\beta = 0.01$

Figure 2 illustrates dynamics described by the deterministic EWA update rule (12), for a game of RPS, defined by the payoff matrix (1). Equivalently, these dynamics correspond to two players employing the $k(1)$ case of the update rule (14). We use the notation $k(a,b)$ to denote a system of two players employing this map with arbitrary k-levels of a and b, e.g. $k(1,1)$ in this case. The normalisation condition (2) removes one degree of freedom from each player's otherwise 3-dimensional mixed strategy, allowing their representation on a 2-dimensional ternary plot in a barycentric coordinate system.

The trajectories shown indicate a fixed point at $\mathbf{x}^* = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$, coinciding with the Nash equilibrium for RPS (5). We can confirm this by substituting these initial conditions into the map (12), thus satisfying the fixed point condition (17). The stability of this fixed point has a dependency on the value of α relative to some critical value α_c, where $\alpha_c = \alpha_c(\beta)$. Trajectories form stable spirals for $\alpha > \alpha_c$, unstable spirals for $\alpha < \alpha_c$, and neutrally stable cycles for $\alpha = \alpha_c$.

We can derive an analytic form for this stability condition by applying the techniques of linear stability analysis described in section 2.3. The Jacobian for the $k(1,1)$ map, evaluated at

the fixed point $\mathbf{x}^*$, then has the form

$$J^{(1,1)}|_{\mathbf{x}*} = \frac{1}{3}(1-\alpha)(3\mathbf{I}_6 - \mathbf{U}) + \frac{\beta}{3}\begin{pmatrix} 0 & \Pi^{RPS} \\ \Pi^{RPS} & 0 \end{pmatrix} \tag{29}$$

where $\mathbf{I}_n$ is the n-dimensional identity matrix, Π^{RPS} is the payoff matrix (1) and

$$\mathbf{U} = \begin{pmatrix} \mathbf{U}_3 & 0 \\ 0 & \mathbf{U}_3 \end{pmatrix}$$

in block matrix notation, with $\mathbf{U}_3$ being a 3×3 matrix whose elements are all equal to 1. The two non-trivial eigenvalues for this Jacobian are

$$\mu^{(1,1)} = (1-\alpha) \pm i\frac{\beta}{\sqrt{3}}$$

each with degeneracy 2.

The condition for stability of this fixed point, $|\mu^{(1,1)}| < 1$ is then given by

$$(1-\alpha)^2 + \frac{\beta^2}{3} < 1 \tag{30}$$

Applying the same technique of linear stability analysis for $k(2,2)$ dynamics gives the Jacobian at $\mathbf{x}*$

$$J^{(2,2)}|_{\mathbf{x}*} = \frac{1}{3}(1-\alpha)(3\mathbf{I}_6 - \mathbf{U}) + \frac{1}{3}\beta(1-\alpha)\begin{pmatrix} 0 & \Pi^{RPS} \\ \Pi^{RPS} & 0 \end{pmatrix} - \frac{\beta^2}{9}(3\mathbf{I}_6 - \mathbf{U})$$

which has non-trivial eigenvalues

$$\mu^{(2,2)} = \left(1 - \frac{\beta^2}{3} - \lambda\right) \pm i\frac{\beta}{\sqrt{3}}(1-\lambda) \tag{31}$$

each with degeneracy 2.

Similarly for the broken symmetry $k(2,1)$ case, the Jacobian evaluated at $\mathbf{x}*$ gives

$$\begin{aligned} J^{(2,1)}|_{\mathbf{x}*} =& \frac{1}{3}(1-\alpha)(3\mathbf{I}_6 - \mathbf{U}) + \frac{\beta}{3}\begin{pmatrix} 0 & 0 \\ \Pi^{RPS} & 0 \end{pmatrix} \\ &+ \frac{\beta}{3}(1-\alpha)\begin{pmatrix} 0 & \Pi^{RPS} \\ 0 & 0 \end{pmatrix} - \frac{\beta^2}{9}\left\{3\begin{pmatrix} \mathbf{I}_3 & 0 \\ 0 & 0 \end{pmatrix} - \begin{pmatrix} \mathbf{U}_3 & 0 \\ 0 & 0 \end{pmatrix}\right\} \end{aligned}$$

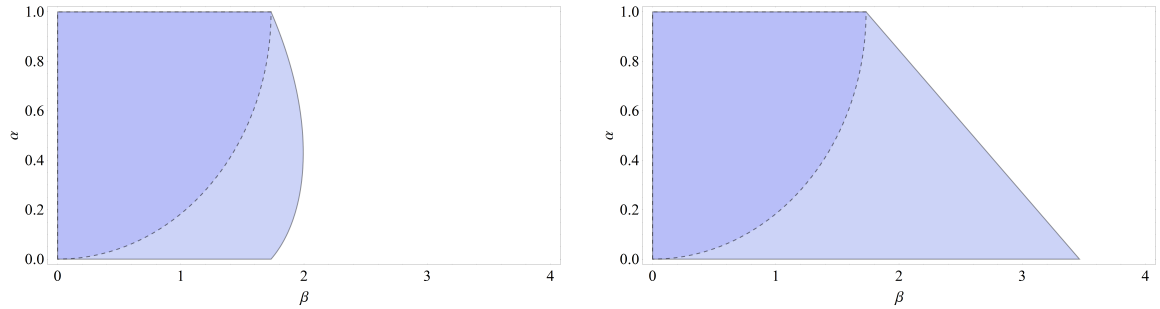
which is equivalent to

$$J^{(2,1)}|_{\mathbf{x}*} = \begin{pmatrix} \mathbf{I}_3 & 0 \\ 0 & 0 \end{pmatrix} \cdot J^{(2,2)}|_{\mathbf{x}*} + \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{I}_3 \end{pmatrix} \cdot J^{(1,1)}|_{\mathbf{x}*}$$

and has non-trivial eigenvalues

$$\mu^{(2,1)} = \left(1 - \frac{\beta^2}{6} - \lambda\right) \pm \frac{\beta}{6}\sqrt{\beta^2 - 12(1-\lambda)} \tag{32}$$

each with degeneracy 2. These eigenvalues are identical to the $k(1,2)$ case.



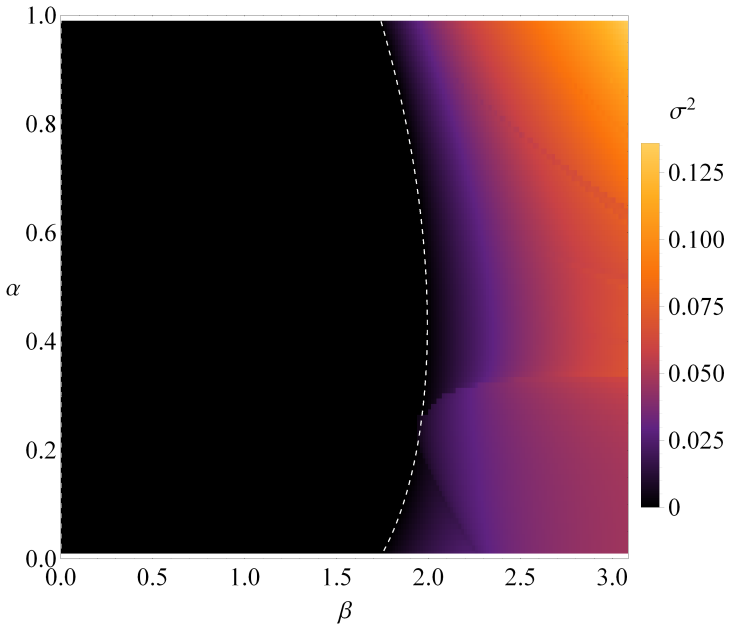
(a) $|\mu^{(2,2)}| < 1$ (b) $|\mu^{(2,1)}| < 1$

Figure 3: Analytic stability regions described by the condition $|\mu| < 1$ for the Jacobian eigenvalues (31) and (32), shaded in blue. The stability region for $k(1,1)$, defined by the inequality (30), is represented by the darker blue region in both cases

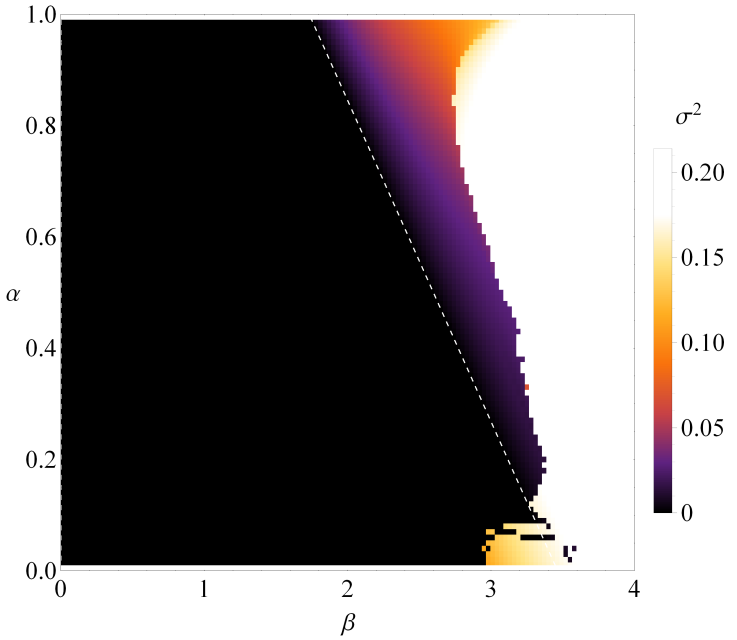
Figure 3 shows the stability regions in β–α parameter space described by these inequalities (31) and (32) compared with that of the $k(1,1)$ case (30). Figure 4 shows plots of the average mixed strategy component variance in $k(2,2)$ and $k(2,1)$ over the same parameter space. This quantity is an indicator of the type of attractor toward which the system is tending for a given set of parameters, and is equivalent to the average variance of the components of a player's mixed strategy as it evolves over an attractor, such that

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{3}{T}\sum_{t=2T/3}^{T} x_i(t)^2 - \left(\frac{3}{T}\sum_{t=2T/3}^{T} x_i(t)\right)^2\right\} \tag{33}$$

where T is some large number of iterations, of which the first two thirds are discarded such that the system be allowed to equilibrate. For the most part, the numerical in these plots results agree strongly with the predictions made from linear stability analysis; i.e. the average component variance of numerically calculated trajectories in regions in which the fixed point $\mathbf{x}^*$ is predicted to be a stable attractor is zero.



(a) $k(2,2)$



(b) $k(2,1)$

Figure 4: Average component variance, calculated according to (33) with $T = 15,000$, for a range of points in β–α parameter space. Dark regions (zero variance) correspond to fixed point attractors, whereas lighter regions correspond to higher–dimensional attractors (e.g. periodic or aperiodic cycles). The dashed white lines correspond to boundaries of the analytic stability regions shown in Figure 3.

3.2 RPS with broken symmetry

Section 3.1 shows results of EWA learning for a symmetric game, where player and opponent have identical RPS payoff matrices. Sato et Al. show in [16] that chaotic behaviour may be observed in dynamics governed by the Sato-Crutchfield equations (13), where this symmetry is broken by a small, heterogeneous 'tie-breaker' parameter for each player. The resulting payoff matrices $\Pi^{(x)}$ and $\Pi^{(y)}$ then have the elements

$$\pi_{ij}^{(x)} = \pi_{ij} + \epsilon_x \delta_{ij}$$
$$\pi_{ij}^{(y)} = \pi_{ij} + \epsilon_y \delta_{ij}$$

where π_{ij} are the elements of the base symmetric payoff matrix Π, $\epsilon_{x,y}$ are small symmetry breaking parameters and δ_{ij} is the Dirac delta function. For $\Pi = \Pi^{RPS}$, this gives the matrices

$$\Pi^{(x)} = \begin{pmatrix} \epsilon_x & -1 & 1 \\ 1 & \epsilon_x & -1 \\ -1 & 1 & \epsilon_x \end{pmatrix} \qquad \Pi^{(y)} = \begin{pmatrix} \epsilon_y & -1 & 1 \\ 1 & \epsilon_y & -1 \\ -1 & 1 & \epsilon_y \end{pmatrix} \tag{34}$$

Figure 6 contrasts the dynamics for such a broken symmetry case with $\epsilon_x = 0.1$, $\epsilon_y = -0.05$, for $k(1,1)$. The parameters values $\alpha = 0.0001$ and $\beta = 0.1$ are chosen such that behaviour in the $k(1,1)$ dynamics approximates those observed in [16]. The corresponding behaviour for the $\epsilon_x = \epsilon_y = 0$ symmetric case is demonstrated in Figure 2a, in which the fixed point $\mathbf{x}^*$ is unstable, and the system tends asymptotically towards *heteroclinic cycles* between the three vertices, with a player spending a linearly increasing number of rounds at each vertex as the system evolves.

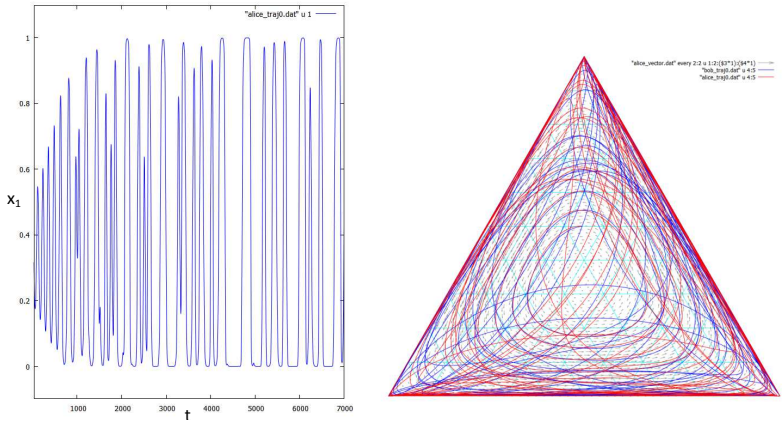
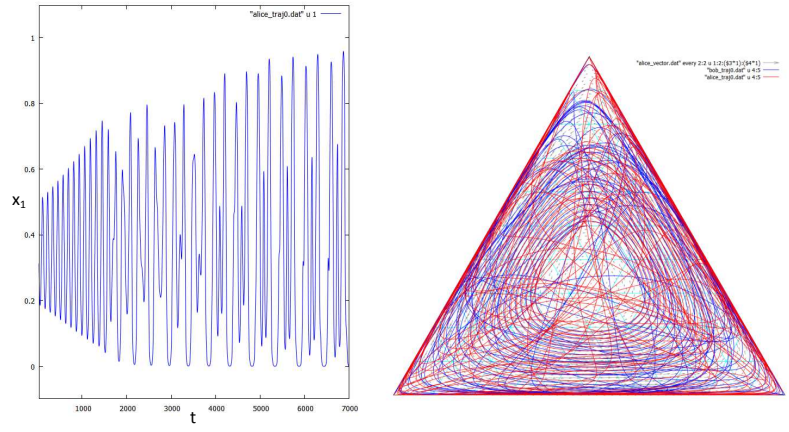


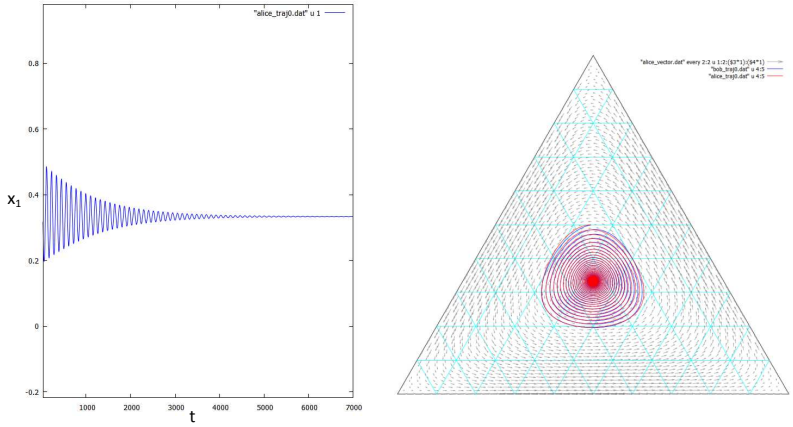
Figure 5: Dynamics for broken symmetry RPS, with $\alpha = 0.0001$, $\beta = 0.1$, $\epsilon_x = 0.1$, $\epsilon_y = -0.05$, under the $k(1,1)$ update rule. The left hand plot show the evolution of a single component of a player's mixed strategy. Right hand plots show the evolution of the entire mixed strategy. The behaviour observed here is comparable to the chaotic transients shown in [16].

In the regime $\epsilon_x + \epsilon_y > 0$, Sato et Al. found that the Sato-Crutchfield dynamics displayed chaotic behaviour similar to that shown in Figure 5. This form of chaotic behaviour, whereby the system tends towards heteroclinic cycles as before, but in an irregular and unpredictable way, is described in detail by Chawanya [17]. Dynamics in the $\epsilon_x + \epsilon_y < 0$ regime tend toward heteroclinic cycles (as in symmetric RPS) once transient behaviour has died away.

Comparison of this behaviour with that shown in Figure 6, shows that changing to an update rule with a different k-level has the potential to significantly alter the system's behaviour and subsequent stability. The $k(2,2)$ case in Figure 6b exhibits an overall change in the stability of



(a) $k(2,1)$



(b) $k(2,2)$

Figure 6: Dynamics for broken symmetry RPS, as in Figure 5 (and with the same parameters), for different k-level update rules. In the case of asymmetric update rules $k(2,1)$, the left hand plot gives one mixed strategy component for the player with k-level a in $k(a,b)$. The right hand plot shows the mixed strategy of a in red and b in blue.

the fixed point, as seen before in section 3.1, and a corresponding lack of chaotic behaviour. The $k(2,1)$ case exhibits irregular behaviour, yet still on average tends towards heteroclinic cycles, despite falling within the corresponding analytic stability region for symmetric RPS (Figure 3). Figure 7 shows plots of LLEs for the three cases given above, over the parameter space of

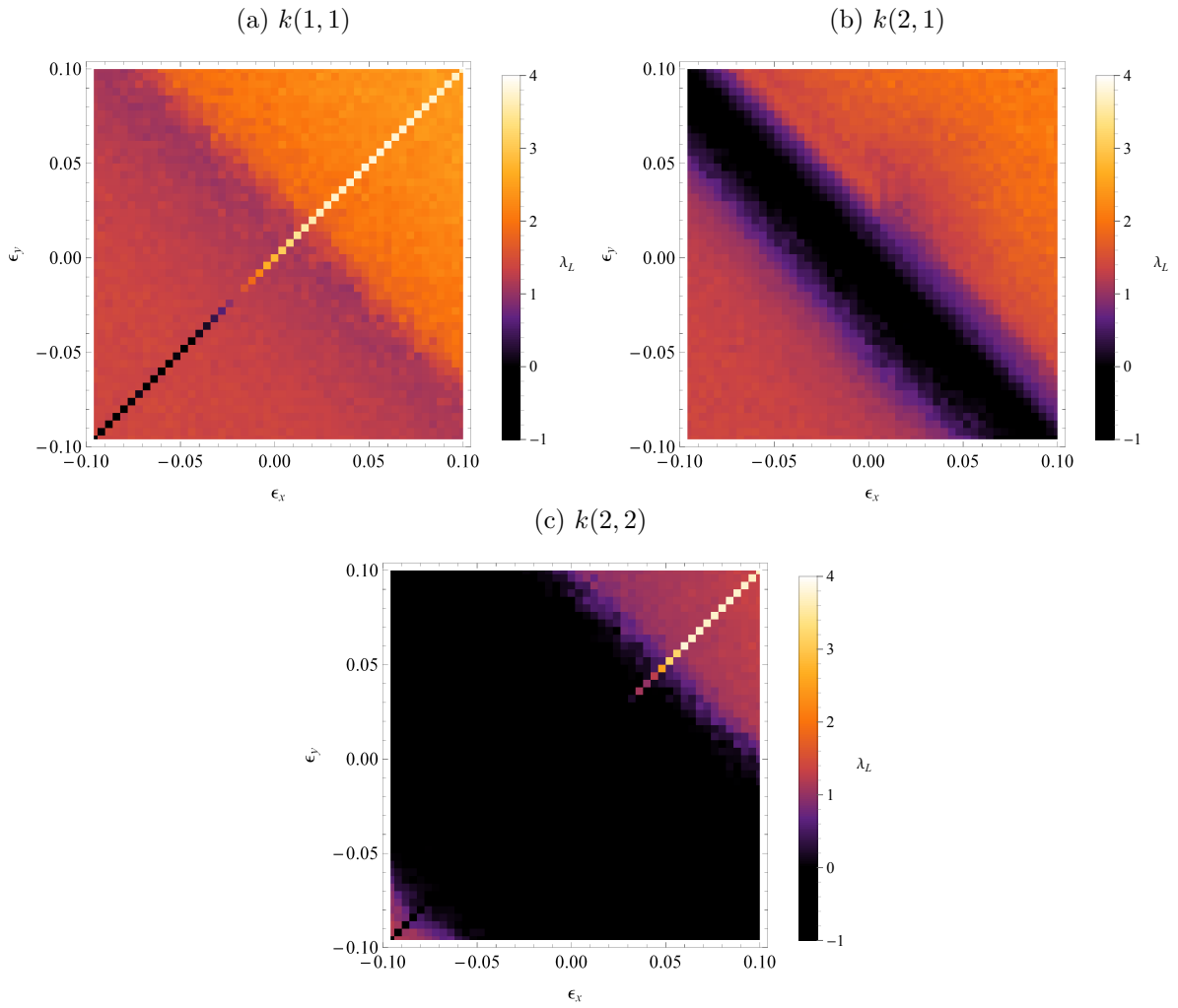


Figure 7: LLEs over the parameter space of tie-breaker parameters $\epsilon_{x,y}$ for different k-level update rules. Learning parameters are fixed at $\alpha = 0.0001$ and $\beta = 0.1$. Points represent an averaging over an ensemble of 100 initial conditions, uniformly distributed over the simplex.

tie-breaker parameters $\epsilon_{x,y}$. The appearance of chaos in a region with the form $\epsilon_x + \epsilon_y > \sigma$ for $k(1,1)$ is reflected in the sharp increase in LLEs for this region compared to that of $\epsilon_x + \epsilon_y < \sigma$ in Figure 7a, where σ is some small value. The disparity between this model and the results given in [16], where $\sigma = 0$, might feasible be attributed to a mismatching of our models. Note that the LLEs in the latter region are still positive, albeit less so than in the former, indicating the divergence of nearby trajectories as they tend toward heteroclinic cycles. This might be explained if we are allowed to speculate that, in-keeping with our observations, the time spent near each vertex is a function of distance from the edge of the simplex. Then, nearby trajectories at infinitesimally different distances from the edge will have a tendency to diverge as they fall "out of step", e.g. with one remaining at a given vertex while the other moves off to the next. In such a case, the interpretation of the positive LLE is that it is impossible to predict which *particular* vertex a player's mixed strategy will be situated near to at some point in the future, given initial conditions close the boundary of the simplex.

In all three cases, the stability behaviour is largely uniform along lines of constant $\epsilon_x + \epsilon_y$, with dynamical behaviour changing as a function of this quantity. Notable exceptions are the special cases in which the system is symmetrised; i.e. $\epsilon_x = \epsilon_y$ for $k(A, A)$. In these cases, player and opponent are identical and the dimensionality of the system is reduced from 4 to 2. The Poincaré-Bendixson theorem [18] implies that chaos may only be observed in a continuous dynamical system with 3 or more dimensions. While our model is based on a discrete map, with a small enough step size we expect it to approximate continuous dynamics. In line with this assumption, we see that for symmetrised systems in our model (corresponding to the $\epsilon_x = \epsilon_y$ lines in Figures 7a and 7c), only fixed points (dark) and asymptotic progression toward heteroclinic cycles (light) are observed. The symmetry is broken in $k(2, 1)$ by the dissimilar update rules for player and opponent, and so no such discontinuity of behaviour is seen for $\epsilon_x = \epsilon_y$ in this case.

Figures 7b and 7c show the change in stability behaviour on moving from the $k(1, 1)$ to $k(2, 1)$ and $k(2, 2)$ update rules. The $k(2, 2)$ case shows a marked increase in stability over much of the $\epsilon_{x,y} \in [-0.1, 0.1]$ parameter space, both in the sense that the fixed point $\mathbf{x}^*$ becomes stable for $-0.15 < \epsilon_x + \epsilon_y \lesssim 0.08$, and in that chaotic behaviour is no longer observed in this region. The $k(2, 1)$ case exhibits similar behaviour to that of $k(1, 1)$, with chaotic dynamics in $\epsilon_x + \epsilon_y \gtrsim 0$ and heteroclinic cycles $\epsilon_x + \epsilon_y \lesssim -0.02$, but with an additional region in which the dynamics evolve to a stable fixed point at $\mathbf{x}^*$, for $\epsilon_x + \epsilon_y \approx -0.01$.

3.3 Large random games

The results so far demonstrate an increase in the parameter space volume for which stability is observed upon implementing higher k-level update rules. In order to determine whether this is a general property of these update rules, or whether it is specific to the cases we have looked at so far, we extend our model to large, randomly generated games. These large random games are defined by pairs of payoff matrices, with joint statistical properties given by (6). In order to compare our results for those of Galla et al. [4], we fix the learning parameter β and observe changes in dynamical behaviour as the learning parameter α and the payoff matrix correlation parameter Γ are changed. In [4], this behaviour is characterised by the Kaplan-Yorke dimension of the attractors in these systems, calculated from their corresponding Lyapunov spectra. Fixed points and periodic attractors have a Kaplan-Yorke dimension close to zero, whereas chaotic attractors are higher-dimensional. For comparison, we only consider the LLE and variance of the corresponding attractors, under the premise that this is a good indicator for chaos, or at least unpredictable behaviour.

Figure 8 shows the average variance in mixed strategy components for the attractors of such games under the $k(1, 1)$ update rule, while Figure 9 shows their corresponding averaged LLEs. Each point in the parameter space of these plots corresponds to an averaging over an ensemble of instances of the system, each with a different randomly generated game. A fresh ensemble of games is generated, by the method described in section 2.1.4, for each point. These plot gives a good agreement with the results for attractor dimension in [4], with regions in which $\lambda > 0$ corresponding to those with high average Kaplan-Yorke dimension. Also shown on these plots is the line $\alpha = \xi(\Gamma)$, corresponding to the theoretical onset of instability in random games in the limit $N \to \infty$, as calculated by a path integral method in the appendix of [4]. Both variance and LLE data show good agreement with this theoretical prediction, with almost all points with $\alpha > \xi(\Gamma)$ having $\sigma^2 = 0$ and $\lambda < 0$, implying fixed point attractors. This both corroborates our results, and lends confidence to the choice of game size, $N = 50$, as adequate approximation to arbitrarily large games.

In the region of $\Gamma \approx 0$ we observe stable attractors with $\lambda < 0$ and $\sigma^2 > 0$, indicative of

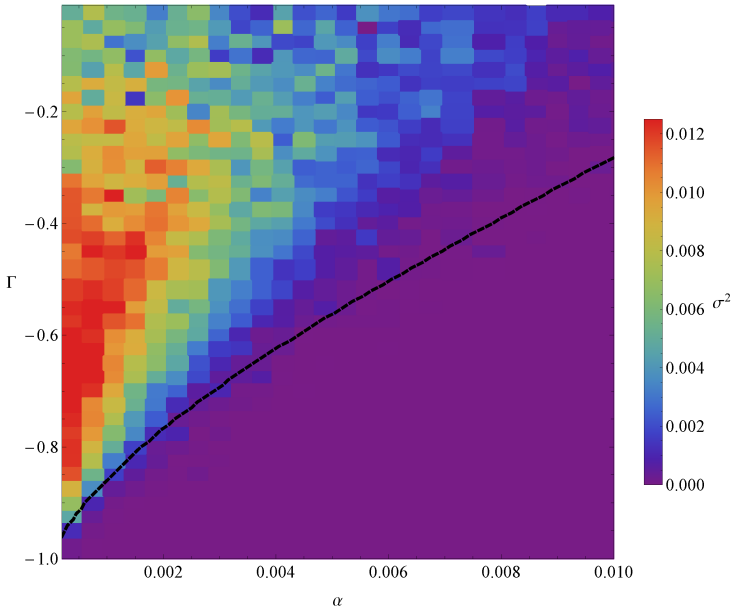


Figure 8: Average mixed strategy component variance in rounds 10,000 – 15,000, averaged over 10 large random games per point, for $N = 50$ and $\beta = 0.07$. Values of zero correspond to attractors with no extent in phase space, i.e. fixed points. Points with $\sigma^2 > 0$ may equally correspond to periodic or aperiodic attractors, we refer to Figure 9 to break this degeneracy.

limit cycles. We see a region of highly unpredictable dynamics where α is small and $\Gamma \approx 0.75$. In [4], it is suggested that learning models in this region of α-β-Γ parameter space may be a good approximation to real-world decision making. In light of this, we designate $\alpha \in [0, 0.002]$, $\Gamma \in [-1, -0.5]$ as a region of interest for the purpose of constraining our results in later plots.

The corresponding variance and LLE plots for the $k(2,1)$ and $k(2,2)$ cases are similar to the point of being superficially indistinguishable, both to one another and to the $k(1,1)$ case. This implies that there is no dramatic net change in stability as a result of moving from the $k(1)$ to $k(2)$ update rule for either or both players, contrary to what earlier results may have suggested. Figure 10 shows correlations between λ in the dynamics described by $k(2,2)$ and $k(1,1)$, for every individual game in each ensemble represented in Figure 9. Points in the two plots are coloured according to the values of Γ and α for that particular game, in order to give a better idea of where we observe different characteristic behaviour. Note that points with greatest LLE for either update rule are located within our previously defined area of interest. Points to the right of the $\lambda_{k(1,1)} = \lambda_{k(2,2)}$ diagonal correspond to games for which λ decreases as a result of moving to a higher k-level, whereas points to the left correspond to games where λ increases. Points on the diagonal represent games in which there was no change λ, a region mostly populated by stable games with either high α, falling within the stability region delineated by $\xi(\Gamma)$, or $\Gamma \approx 0$, corresponding to the stable attractors previously noted. Most important to note are points in the upper-left and lower-right hand quadrants of these plots, indicating transitions between positive and negative λ on changing from one update rule to the other. These correspond to transitions between high- and low-dimensional attractors, and accordingly, significant changes in overall system stability. The symmetry of this plot about $\lambda_{k(1,1)} = \lambda_{k(2,2)}$ indicates that, while changing update rule may often result in a dramatic shift in stability for given game, this shift is equally likely to happen in either direction, giving no net change when averaged over an ensemble.

Figure 11 again show correlations in λ, but now restricted to points within the region of

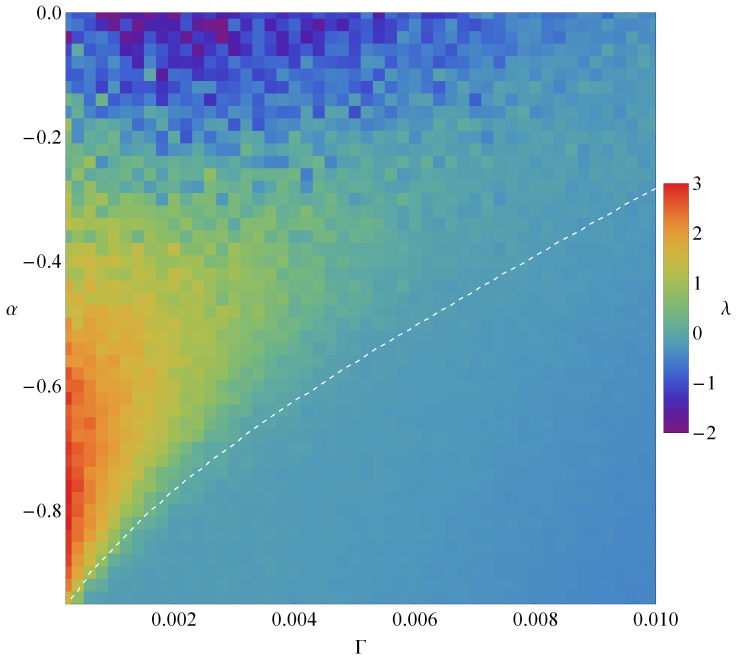


Figure 9: Largest Lyapunov exponents averaged over 25 large random games per point, for $N = 50$ and $\beta = 0.07$. In each instantiation, the system is allowed to run for 15,000 rounds in order to equilibrate. The dashed line shows the theoretical boundary for stability predicted by a path integral method, corresponding to the limit $N \to \infty$, with matrix elements scaled accordingly [4]. Regions in blue, where $\lambda < 0$, corresponds to fixed points and other low-dimensional attractors. Regions in green through to red indicate higher-dimensional attractors and unpredictable behaviour.

interest $\alpha \in [0.000, 0.002]$, $\Gamma \in [-1.0, -0.5]$, and extending up to higher k-level update rules. Also shown are the corresponding changes in distribution of λ values for games in this region. For the $k(1,1) \to k(2,2)$ case in both representations, we again observe a lack of net change in stability. However, in moving from $k(1,1) \to k(3,3)$, an asymmetry develops in the λ value correlations. Many points now fall within the critical lower right quadrant, indicating a net shift in λ toward negativity, and a corresponding increase in stability and predictability of dynamical behaviour. The plot showing change in distribution also reflects this. The $k(1,1) \to k(5,5)$ plots show a similar asymmetric shift, but not to a significantly greater degree than in the case of $k(3,3)$.

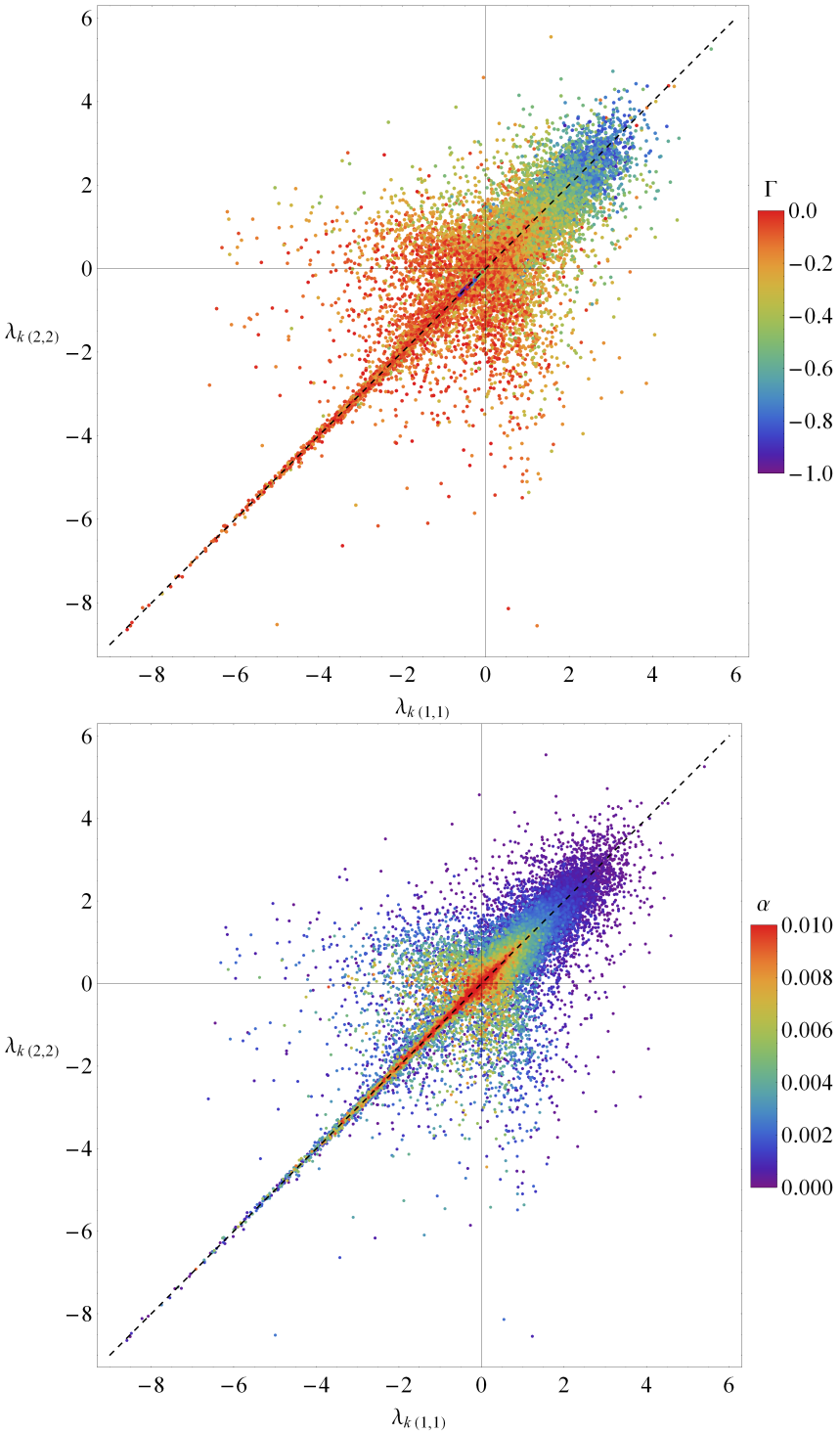


Figure 10: Correlations in λ between $k(1,1)$ and $k(2,2)$ in 625,000 unique random large games, uniformly distributed over the parameter region $\Gamma \in [-1.0, 0.0]$, $\alpha \in [0.00, 0.01]$. Top and bottom plots are colour coded so as to reflect the value of Γ and α for each point, respectively.

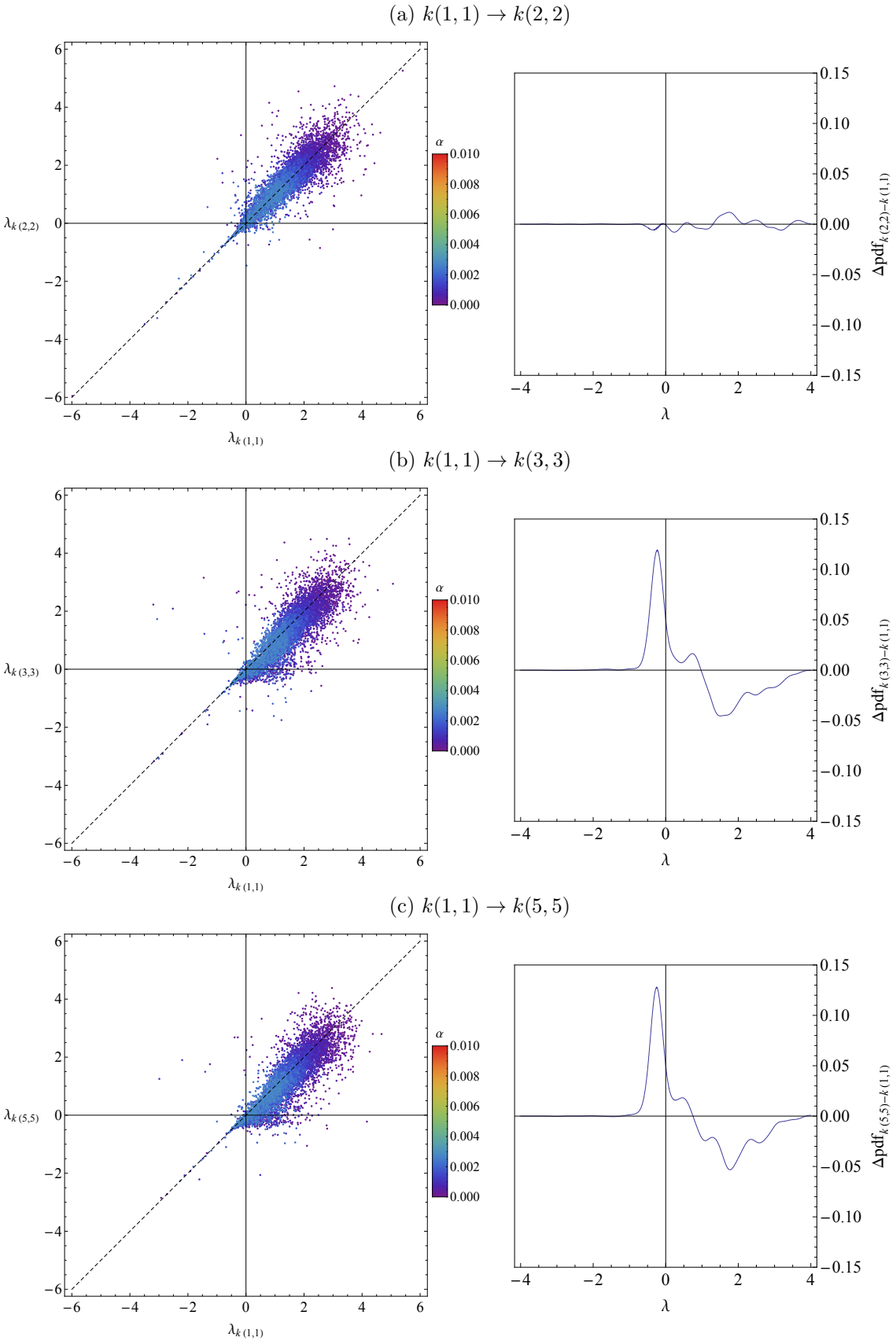


Figure 11: Correlations between LLEs of k(1,1) and higher k-level update rules, in the parameter range $\alpha \in [0, 0.003]$, $\Gamma \in [-1, -0.5]$. Right hand plots show the change in λ distribution over this region on moving between update rules.

4 Summary and Conclusion

In section 2.1 we introduce the basic terminology and mathematical framework of classical game theory, including the key concepts: normal form games, mixed strategies and solution concepts. In particular, we looked at the criteria for a Nash Equilibrium, the most general and commonly cited form of equilibrium state for a game theoretic system. In such a state, no player has any incentive to change their intended strategy, given perfect information of their opponent's intended strategy. We go on to introduce the idea of depth of strategic reasoning, motivated by the game theoretic Keynesian beauty contest and its simplified form: the p-beauty contest. In such a contest, a successful player must not only out-wit her opponents by pre-empting their intending strategies, using an understanding of their reasoning process, but also gauge the degree of *bounded rationality* to employ so as not to 'over-think' the problem. In order generalise away from specific games, and to later compare our results to those demonstrated in [4], we describe a process for generating large random games with particular statistical properties. These large random games will go on to form the basis of our final results.

Section 2.2 introduces the dynamical approach to game learning used throughout the paper, defining update rules for the EWA learning model in the deterministic limit. These rules model the process by which a pair of players learn to play against one another in a given game. In moving to the deterministic limit, we assume a state of perfect information, in which both players may utilise all the relevant information about the system in order to adapt their behaviour. This eliminates any potential effects due to uncertainty and stochastic sampling from the dynamics. Such effects introduce an element of unpredictability into a system by their inherent nature. With these removed, we may be confident that any unpredictable behaviour displayed is a result of entirely deterministic processes. We propose an extension to the EWA learning model, implementing a model of *k-level reasoning* based on the process of rationalisation employed by contestants in a p-beauty contest. The original set of experiments by Nagel [7] and more recent work by Camerer [8] regarding the biological feasibility of such an extension lend confidence to our approach.

By constructing a dynamic model of such game theoretic systems, we area able to draw on techniques from non-linear dynamics, analysing their long term behaviour in terms of *attractors*. In section 2.3 we describe these techniques, such as linear stability analysis and calculation of largest Lyapunov exponents, that allow us to quantify the stability and predictability of this behaviour.

Section 3.1 illustrates the dynamics in a simulation of two players adapting their behaviour as they learn to competitively play a simple game of RPS, using the EWA with and without k-level reasoning. Using linear stability analysis we derive analytic predictions for stability, which are closely corroborated by numerical simulation. We observe that the stability of potential equilibrium states, including those predicted by classical game theory, are highly dependent on the learning model and its parameters. In section 3.2 we extend these dynamics to games with asymmetric payoff matrices and reproduce results of [16], illustrating the possibility of chaotic behaviour emerging even in a very simple game.

Section 3.3 shows the results of extending these models to games defined by pairs of large, randomly generated matrices with specific correlation properties. This is motivated by a desire to generalise away from specific cases, so as to better examine the properties of the learning model independent of the particular game being played. It is also suggested that such games represent real-world scenarios, where many possible options are available to decision-making agent, and the number of possible equilibrium states is very large. We observe different regions within the α - Γ parameter space, with highly unpredictable behaviour being observed for games

that are competitive, but not zero-sum. A lack of correlation between player and opponent payoff matrices, $\Gamma \approx 0$, is observed to result in predictable, periodic behaviour. We find that, while the predictability of learning behaviour in a given game may be dramatically affected by the k-level employed by players, there is little net change when averaged over a large ensemble of random games, and none at all for $k(n < 3)$. For $k(3, 3)$ and $k(5, 5)$, we do find a significant shift toward predictability over a small range of parameters.

In conclusion, we find that implementing a biologically feasible model of k-level reasoning can result in an overall increase in predictability of the long term behaviour of players learning a complicated game, where this predictability is quantified by the negativity of the largest Lyapunov exponent for an attractor in that system. This increase is observed over a range of parameters in which the EWA model is suggested to approximate real-world decision making processes [19] and at depth of k-level reasoning deemed realistic by experimental studies [8]. In spite of this change, the qualitative schema of stability and predictability with respect to these parameters remains invariant. However, little change in predictability is observed as the depth of strategic reasoning is further increased, from $k(3)$ to $k(5)$, giving no indication that moving to arbitrarily high k-level would make the system's behaviour arbitrarily more predictable. Future In any case, the results of the studies in [8] imply that it is unlikely for human players to employ an greater depth of strategic reasoning than that corresponding to the $k(5)$ model. We consider this paper to be a response to the suggestion made in [4] that a wider range of biologically feasible learning models should be employed in probing the inherent unpredictability in learning to play complicated games. As such, our results corroborate those of [4].

It is worth noting that for each instance of the systems simulated, only one set of initial conditions was considered. These initial conditions corresponded to random play, based upon the premise that players have no a priori bias toward any particular strategy. It is possible that, given different initial conditions, these systems may settle into one of many alternative attractors, each with potentially different properties.

5 Acknowledgements

We would like to thank Ian Cottam, and all other faculty members involved in the running and maintaining the EPS Condor High Throughput Computing network. Without access to this incredible resource, this paper would have displayed a relative paucity of notable results.

We thank Tobias Galla for his support and supervision throughout the course of our research, and Stefan Soldner-Rembold for his constructive feedback on our early results.

References

[1] J. v. Neumann and O. Morgenstern. *Theory of games and economic behavior*. Princeton, NJ: Princeton Univiversity Press, 1944.

[2] J.M. Smith. *Evolution and the Theory of Games*. Cambridge university press, 1982.

[3] C. Camerer and T. Hua Ho. Experience-weighted attraction learning in normal form games. *Econometrica*, 67(4):827–874, 2003.

[4] Tobias Galla and J Doyne Farmer. Complex dynamics in learning complicated games. *Proceedings of the National Academy of Sciences*, 110(4):1232–1236, 2013.

[5] Andrew McLennan and Johannes Berg. Asymptotic expected number of nash equilibria of two-player normal form games. *Games and Economic Behavior*, 51(2):264–295, 2005.

[6] John Maynard Keynes. The general theory of employment. *The Quarterly Journal of Economics*, 51(2):209–223, 1937.

[7] John Duffy and Rosemarie Chariklia Nagel. On the robustness of behaviour in experimental" beauty contest" games. *Economic Journal*, 107(445):1684–1700, 1997.

[8] Colin F Camerer, Teck-Hua Ho, and Juin-Kuan Chong. A cognitive hierarchy model of games. *The Quarterly Journal of Economics*, 119(3):861–898, 2004.

[9] A.N. Kolmogorov. *Grundbegriffe der wahrscheinlichkeitsrechnung*. Springer, 1933.

[10] J. Nash. Non-cooperative games. *Annals of mathematics*, 54(2):286–295, 1951.

[11] Hervé Moulin. *Game theory for the social sciences*. NYU press, 1986.

[12] L.A. Imhof, D. Fudenberg, and M.A. Nowak. Evolutionary cycles of cooperation and defection. *Proceedings of the National Academy of Sciences of the United States of America*, 102(31):10797–10800, 2005.

[13] A.J. Bladon, T. Galla, and A.J. McKane. Evolutionary dynamics, intrinsic noise, and cycles of cooperation. *Physical Review E*, 81(6):066122, 2010.

[14] T. Galla. Cycles of cooperation and defection in imperfect learning. *Journal of Statistical Mechanics: Theory and Experiment*, 2011(08):P08007, 2011.

[15] Y. Sato, E. Akiyama, and J.P. Crutchfield. Stability and diversity in collective adaptation. *Physica D: Nonlinear Phenomena*, 210(1):21–57, 2005.

[16] Yuzuru Sato, Eizo Akiyama, and J Doyne Farmer. Chaos in learning a simple two-person game. *Proceedings of the National Academy of Sciences*, 99(7):4748–4751, 2002.

[17] Tsuyoshi Chawanya. A new type of irregular motion in a class of game dynamics systems. *Progress of Theoretical Physics*, 94(2):163–179, 1995.

[18] Ivar Bendixson. Sur les courbes définies par des équations différentielles. *Acta Mathematica*, 24(1):1–88, 1901.

[19] Teck H Ho, Colin F Camerer, and Juin-Kuan Chong. Self-tuning experience weighted attraction learning in games. *Journal of Economic Theory*, 133(1):177–198, 2007.