Institute for
New Economic Thinking
AT THE OXFORD MARTIN SCHOOL

A TAXONOMY OF LEARNING DYNAMICS IN 2 × 2 GAMES

Marco Pangallo, James Sanders,
Tobias Galla, and J. Doyne Farmer

2nd May 2017

INET Oxford Working Paper no. 2017-06

Complexity Economics

A taxonomy of learning dynamics in 2×2 games

Marco Pangallo[⋆,1,2], James Sanders[3], Tobias Galla[3], and J. Doyne Farmer[1,2,4]

[1]Institute for New Economic Thinking at the Oxford Martin School, University of Oxford, Oxford OX2 6ED, UK

[2]Mathematical Institute, University of Oxford, Oxford OX1 3LP, UK

[3]Theoretical Physics, School of Physics and Astronomy, University of Manchester, Manchester M13 9PL, UK

[4]Santa Fe Institute, Santa Fe, NM 87501, US

Abstract

Learning would be a convincing method to achieve coordination on an equilibrium. But does learning converge, and to what? We answer this question in generic 2-player, 2-strategy games, using Experience-Weighted Attraction (EWA), which encompasses many extensively studied learning algorithms. We exhaustively characterize the parameter space of EWA learning, for any payoff matrix, and we understand the generic properties that imply convergent or non-convergent behaviour in 2×2 games.

Irrational choice and lack of incentives imply convergence to a mixed strategy in the centre of the strategy simplex, possibly far from the Nash Equilibrium (NE). In the opposite limit, in which the players quickly modify their strategies, the behaviour depends on the payoff matrix: (i) a strong discrepancy between the pure strategies describes dominance-solvable games, which show convergence to a unique fixed point close to the NE; (ii) a preference towards profiles of strategies along the main diagonal describes coordination games, with multiple stable fixed points corresponding to the NE; (iii) a cycle of best responses defines discoordination games, which commonly yield limit cycles or low-dimensional chaos.

While it is well known that mixed strategy equilibria may be unstable, our approach is novel from several perspectives: we fully analyse EWA and provide explicit thresholds that define the onset of instability; we find an emerging taxonomy of the learning dynamics, without focusing on specific classes of games ex-ante; we show that chaos can occur even in the simplest games; we make a precise theoretical prediction that can be tested against data on experimental learning of discoordination games.

Key Words: Behavioural Game Theory, EWA Learning, Convergence, Equilibrium, Chaos.

JEL Class.: C62, C73, D83.

*Corresponding author: marco.pangallo@maths.ox.ac.uk. We thank Vince Crawford, Mike Harré, Cars Hommes, Peiran Jiao, Robin Nicole, Mousavi Shabnam and Peyton Young, as well as seminar participants at Nuffield College, INET YSI Plenary, Herbert Simon Society International Workshop, Conference on Complex Systems 2016 and King's College, for helpful comments and suggestions.

Contents

1 Introduction

How do players coordinate on specific profiles of strategies in non-cooperative games, and why should they coordinate on an equilibrium profile? If the game is simple or one-shot, a reasonable explanation is provided by strategic thinking and introspection. Another justification, which is more generally valid in complicated and repeated games, is learning and interaction. However, as it is fairly well known since the contribution of Shapley (1964), the learning dynamics may fail to converge to an equilibrium. This questions the validity of equilibrium thinking in game theory: at least in some contexts, strategic interactions might be governed by learning in an ever-changing environment, rather than by rational and fully-informed decision making. The literature has faced the dilemma about the convergence of the learning dynamics to Nash Equilibria (NE) in several ways. Most theoretical work has identified classes of games and learning algorithms in which the dynamics succeeds to converge; some authors provided counter-examples in which learning would not converge.[1] Little has been said about the *generic properties* of games and learning algorithms which yield a convergent or non-convergent dynamics. Recent work (Galla and Farmer, 2013) addressed this issue by considering *ensembles* of 2-person, N-strategy games and finding the regions of the parameter space where learning was less likely to converge: negatively correlated payoffs and "rational" long-memory learning implied limit cycles and high-dimensional chaos in the learning dynamics. However, little understanding of the reasons for non-convergent behaviour was provided.

In order to shed light on the mechanisms behind (non-)convergence, this paper investigates the drivers of instability in the simplest possible non-trivial setting, that is generic 2-person, 2-strategy normal form games, trying to capture the typical features of the payoff matrix and of the learning behaviour that yield a cycling or an irregular dynamics. We study a slightly simplified version of Experience-Weighted Attraction (EWA), which is general enough to encompass both reinforcement and belief learning and has been shown to be in accord with experimental data (Camerer and Ho, 1999). In short, we find that the existence of a *cycle of best responses* in the payoff matrix,[2] coupled with a quick enough learning dynamics (in a sense that will be specified later), is a sufficient condition for the non-convergence of learning. In particular, in games with a unique mixed strategy equilibrium (to which we refer as *discoordination games*, lacking an established terminology in the literature) the players follow the cycle of best responses and never converge to the NE: we rather observe limit cycles or low-dimensional chaos. Lack of convergence is driven by the players adapting too quickly to the moves of their opponent. In the same learning scenario, if the payoff matrix is *acyclic* (there is at least one fixed point in terms of best responses, that is a profile of strategies which is the best response by both players to some beliefs on their opponent), as in dominance-solvable and coordination games, convergence to a pure strategy NE occurs immediately. On the contrary, if the players are "irrational" and/or do not have enough incentives to switch their moves, they do not recognize that a pure strategy may be better and simply randomize between their possible moves, reaching a mixed strategy fixed point.

We find such a taxonomy of the learning dynamics by looking at relevant combinations of parameters, which naturally emerge from the mathematical analysis. Figure 1 illustrates our approach and provides a qualitative characterization of the parameter space. We denote by "irrationality" the ratio of two parameters of EWA, namely the memory loss of past performance α divided by the closeness to optimal decision making β (payoff sensitivity or *intensity of choice*). "Coordination" (AC) depends on the payoff matrix and quantifies the preference of the players for "diagonal" outcomes: if we denote their pure strategies by 1

[1] Robinson (1951); Miyazawa (1961); Shapley (1964); Crawford (1974); Stahl (1988); Nachbar (1990); Milgrom and Roberts (1991); Krishna (1992); Conlisk (1993a); Monderer and Shapley (1996); Hahn (1999); Arieli and Young (2016).

[2] For instance, in Matching Pennies, if player Row (who wins if the pennies are matched) thinks that player Column would play Heads, the best response for Row would be to play Heads. The best response for Column to this move of player Row is to play Tails. Row would then switch to Tails as well, and so on.

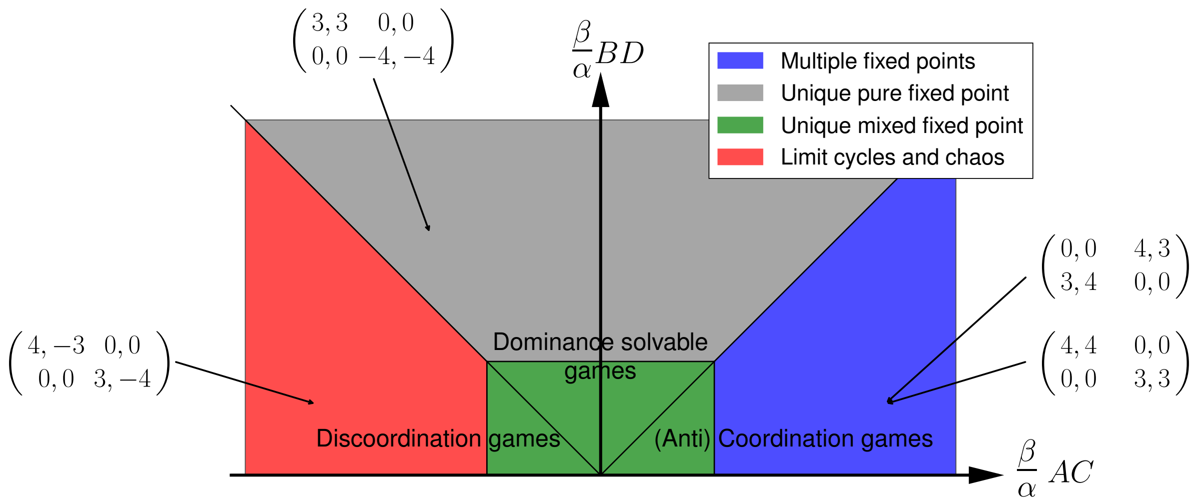


Figure 1: Qualitative characterization of the parameter space. The *irrationality* α/β refers to the intrinsic noise in the learning algorithm. *Coordination* (AC) and *dominance* ($|BD|$) quantify properties of the payoff matrix. The combinations of these parameters characterize the learning dynamics and relate to specific classes of 2×2 games.

and 2, coordination is large when the payoffs associated with the profiles of strategies $(1,1)$ and $(2,2)$ are much larger than the payoffs for $(1,2)$ and $(2,1)$.[3] "Dominance" ($|BD|$) on the other hand quantifies the relative strength of a pure strategy with respect to the other one. Coordination and dominance naturally relate to well-known classes of 2×2 games (see Table 1 and Section 2). Dominance is large for dominance-solvable games; coordination is positive and large for coordination and anticoordination games, whereas it is negative and large for discoordination games.

The learning behaviours in the taxonomy are quite intuitive and generally known in the literature. The main contribution of this work lies in its methodology, namely in the exhaustive characterization of the space of games which makes it possible to identify specific structures in the payoff matrix and find various classes of games *ex-post*, based on the convergence properties of the learning algorithm. Further details on the relations of this work with the existing literature are provided below.

Related literature The first example of a normal form game where convergence of fictitious play (Brown, 1951; Robinson, 1951) did not occur was provided by Shapley (1964). He considered a 3×3 payoff matrix similar to the one in Rock-Paper-Scissors, and showed that fictitious play would spiral out of the mixed strategy NE, by following cycles of exponentially increasing period. Crawford (1974) considered another adaptive learning algorithm (related to gradient learning), which was not able to reach mixed strategy NE, but Conlisk (1993a,b) showed how some amendments to Crawford's adaptive dynamics lead instead to convergence. From the point of view of the properties of the payoff matrix, Nachbar (1990) showed convergence to Nash Equilibria in dominance-solvable games, Monderer and Shapley (1996) proved convergence for potential games, Milgrom and Roberts (1991) considered games with strategic complementarities (or supermodular games) and demonstrated that learning algorithms consistent with adaptive learning would converge to the serially undominated set of pure strategies. Arieli and Young (2016) studied stochastic better-reply dynamics in weakly acyclic games, which encompass all classes of games considered above, and showed that it is possible to put bounds on the speed of convergence.

[3] Or viceversa: coordination is also large if the payoffs for $(1,2)$ and $(2,1)$ are much larger than the payoffs for $(1,1)$ and $(2,2)$.

Type of game	Payoff matrix: $\begin{pmatrix} a,e & b,g \\ c,f & d,h \end{pmatrix}$	Example	Parameters
Coordination	$a > c$, $b < d$, $e > g$, $f < h$. Two pure strategy $(1, 1)$; $(2, 2)$ and one mixed strategy NE.	$\begin{pmatrix} 5,2 & -1,1 \\ 0,-3 & 3,4 \end{pmatrix}$	$AC = 72$ $\lvert BD\rvert = 6$
Anticoordination	$a < c$, $b > d$, $e < g$, $f > h$. Two pure strategy $(1, 2)$; $(2, 1)$ and one mixed strategy NE.	$\begin{pmatrix} 1,0 & 5,4 \\ 2,3 & 4,1 \end{pmatrix}$	$AC = 12$ $\lvert BD\rvert = 0$
Discoordination	$a > c$, $e < g$, $b < d$, $f > h$; $a < c$, $e > g$, $b > d$, $f < h$. Unique mixed strategy NE.	$\begin{pmatrix} 4,-3 & -1,-2 \\ -3,2 & 3,-5 \end{pmatrix}$	$AC = -88$ $\lvert BD\rvert = 18$
Dominance-solvable	All other possible orderings. E.g. $a > c$, $b > d$, $e > g$, $f > h$. Unique pure strategy NE.	$\begin{pmatrix} 5,3 & -1,2 \\ 0,-1 & -2,-3 \end{pmatrix}$	$AC = 4$ $\lvert BD\rvert = 18$

Table 1: Games in the taxonomy. The games are defined in terms of the orderings in the payoff matrix. Coordination is AC, where $A = a + d - b - c$ and $C = e + h - f - g$, while dominance is $|BD|$, $B = a + b - c - d$, $D = e + f - g - h$. In dominance-solvable games, $|BD| > |AC|$; in coordination and anticoordination games, $|BD| < |AC|$, $AC > 0$; in discoordination games, $|BD| < |AC|$, $AC < 0$. Note that there are some exceptions: see Proposition 1.

Another literature focused on the generic properties of the payoff matrices and learning algorithms that were associated with multiplicity of NE or non-convergent behaviour. Berg and Weigt (1999) showed how the number of NE increases exponentially with the correlation of the payoffs, while Opper and Diederich (1992) considered the replicator dynamics with a large number of species and used techniques from the statistical physics of disordered systems to show how, below a certain level of cooperation pressure (a parameter characterizing the learning algorithm), the dynamics becomes unstable. More recently, Galla and Farmer (2013) analysed random games and EWA learning, showing that high-dimensional chaos and limit cycles could be observed in a significant portion of the parameter space, for negatively correlated payoffs.

This paper bridges the two described literatures in that we exhaustively characterize the parameter space of EWA in generic 2×2 games and we connect *ex-post* the learning dynamics to specific classes of games based on the convergence properties of the learning algorithm, rather than focusing *ex-ante* on any specific class of games. We show that convergence occurs in acyclic 2×2 games, such as dominance-solvable, potential, coordination and supermodular games (Nachbar, 1990; Monderer and Shapley, 1996; Milgrom and Roberts, 1991; Arieli and Young, 2016), and that such games are more common if the payoffs are positively correlated; the higher the correlation, the more likely the payoff matrix describes a coordination game, with multiple fixed points (Galla and Farmer, 2013; Berg and Weigt, 1999). Moreover, we generalize a number of results on the well-known instability of mixed strategy equilibria (Stahl, 1988; Benaïm et al., 2009) by tuning the free parameters of EWA. For instance, fictitious play converges to mixed strategy NE in 2×2 discoordination games (Miyazawa, 1961), whereas best-response dynamics (Cournot, 1838) does not. Both fictitious play and best-response dynamics are limiting cases of EWA (Camerer and Ho, 1999) and we recover their convergence properties through our analysis, but we also provide explicit thresholds that define the onset of instability in between the two limits.

An important finding of this paper is that the unstable dynamics might be chaotic even in 2×2 games (note that deterministic learning is not necessary for the dynamics to display this property: chaos is well defined even in the presence of noise. See Crutchfield et al. 1982). While Vilone et al. (2011) obtained this result by analysing a specific discretization of the replicator dynamics, we are the first to find chaotic behaviour in a 2-dimensional strategy space with reinforcement and belief learning algorithms. Due to the reduced number of pure strategies available to each player, we find low-dimensional chaos,[4] in contrast with Galla and Farmer (2013), who find high-dimensional chaotic attractors (which are consistent with an essentially random and unpredictable learning dynamics) in games with many pure strategies. Since we find a quasi-cyclical learning dynamics, it can sensibly be argued that the pattern can be guessed by one of the players, who could then take advantage of her forecast of the moves of her opponent in order to systematically outguess his choices, and thereby perform better than him. In evolutionary terms, the player who can guess the cyclical behaviour of her opponent has higher fitness and is eventually expected to take over the entire population. This is the rational expectations argument of Muth (1961) and would suggest that the cyclic behaviour is expected to die out. However, in line with the view of the *rational route to randomness* (Brock and Hommes, 1997), this is not an obvious outcome. The information cost for guessing the moves of the other player and the interaction between two or more forecasting strategies easily yield complex dynamics, preventing rational and perfectly informed players to outperform less sophisticated players. Hommes et al. (2016) apply this formalism to the theory of learning in games by considering the interplay between rational play and a short memory adjustment process such as best-response dynamics or fictitious play in Cournot games. Rational players are able to outguess the choices of their opponents, but complex dynamics may still occur. In a different context, Huberman and Hogg (1988) show that more sophisticated learning algorithms may lead to chaotic dynamics.

Another understandable critique is whether our learning algorithm can be considered as representative of how players learn in reality, and whether limit cycles or chaos in the learning dynamics play a role in the real world and could be detected in experiments. Camerer and Ho (1999) and Ho et al. (2007) fit the EWA model to experimental data in several classes of games and show that it outperforms other learning models in most cases. However, it is likely that the players would change their learning strategy as the game evolves, implying that they *learn how to learn*. Stahl (1996) considered a model of *rule learning* where the players are of different k-levels (Nagel, 1995) and change their k-level using reinforcement learning. Crawford (1995) proposed a generalization of the standard belief learning algorithms to take into account time-varying memory and idiosyncratic shocks. Would we find the same qualitative learning dynamics if we used more sophisticated learning algorithms? Our analysis suggests that limit cycles and chaos may theoretically be observed as long as the players are willing to quickly switch their moves, independently of the reason why they behave so. A property of the cycling behaviour, as opposed to the convergence to a mixed strategy equilibrium, is the slower decay in the autocorrelation function of the moves chosen by each player. In the language of time series, the sequence of moves by each player exhibits *persistence*. This is a precise theoretical prediction that can be tested against data on experimental learning of discoordination games.

Organization of the paper The rest of this paper is organized as follows: in Section 2 we define the classes of 2×2 games that we use in this paper; in Section 3 we describe the learning model and a number of simplifications that help the subsequent analysis; in Section 4 we fully characterize the learning dynamics as a function of the parameters and we connect the classes of 2×2 games to randomly generated payoff matrices; in Section 5 we relax some strong assumptions used in the previous section and we prove that our results are robust to stochasticity. Section 6 concludes.

2 Classes of 2-person, 2-strategy games

Based on the properties one wants to look at, it is possible to construct several classifications of 2-person, 2-strategy (2×2) games. Rapoport et al. (1976) find 78 classes of games, which can be reduced to 24 when less properties are considered. Here we are only concerned with the number of Nash Equilibria (NE) and with their type, i.e. whether they are pure or

[4] The dimensionality of the chaotic attractors quantifies the departure from regular oscillations.

mixed strategy NE. We only find 3 classes of 2×2 games: all such games belong to one of these classes.

Consider the general 2×2 payoff bi-matrix (henceforth called payoff matrix):

$$\begin{pmatrix} a, e & b, g \\ c, f & d, h \end{pmatrix}. \tag{1}$$

The number and type of the NE depend on the pairwise ordering of the payoffs each player compares, namely (a, c) and (b, d) for player Row, (e, g) and (f, h) for player Column. There are $2^4 = 16$ such orderings. We find the following classes of 2×2 games:

- **Coordination and anticoordination games** are respectively defined by the orderings $a > c$, $d > b$, $e > g$, $h > f$ and $a < c$, $d < b$, $e < g$, $h < f$. Coordination 2×2 games have 2 pure strategy NE where the players choose the strategies with the same labels, i.e. $(1, 1)$ and $(2, 2)$, and one mixed strategy NE. Two well-known examples of coordination games are Stag-Hunt and Battle of the Sexes. What we name anticoordination 2×2 games (there exists no standard terminology for such games) are largely similar to coordination games, in that they also have two pure strategy and one mixed strategy NE, but in the pure strategy NE the players choose strategies with different labels, i.e. $(1, 2)$ and $(2, 1)$. A well known example of an anticoordination game is Chicken. From a mathematical point of view, coordination and anticoordination games are largely similar, so we group them together in most cases.
- **Discoordination games** are defined by the orderings $a > c$, $d > b$, $e < g$, $h < f$ and $a < c$, $d < b$, $e > g$, $h > f$ (again, there exists no standard terminology for this class of games). They have a unique mixed strategy NE and no pure strategy NE because the players have incentives to coordinate on different profiles of strategies. The prototypical discoordination game is Matching Pennies.
- **Dominance-solvable games** are defined by all 12 remaining possible orderings. They have a unique pure strategy NE, obtainable from the elimination of strongly dominated strategies. For instance, if $a > c$, $d < b$, $e > g$, $h < f$, the NE is $(1, 1)$. The Prisoner Dilemma is a 2×2 dominance-solvable game.

Hofbauer and Sigmund (1998, Chapter 10, p. 120) find a very similar classification with respect to the replicator dynamics. The only difference is that in their setting coordination and anticoordination games are completely equivalent. This is correct for the relatively simple functional form of the replicator dynamics, but inexact for more complex learning algorithms such as the one analysed in this paper.

It will be very useful in the following to relate the classes of 2×2 games described above to the following key parameters:

$$\begin{aligned} A &= \frac{1}{4}(a + d - b - c), \\ B &= \frac{1}{4}(a + b - c - d), \\ C &= \frac{1}{4}(e + h - f - g), \\ D &= \frac{1}{4}(e + f - g - h). \end{aligned} \tag{2}$$

The parameter A indicates the preference of player Row for outcomes of the type $(1, 1)$ or $(2, 2)$ over the cases $(1, 2)$ and $(2, 1)$. Similarly C is a measure for the preference of player Column for the same "diagonal" outcomes. It is then sensible to use the product AC as a measure of overall coordination. We then name AC as the "coordination" parameter. If both A and C are positive and large, coordination is positive and large and both players prefer outcomes $(1, 1)$ and $(2, 2)$. We have a coordination game and learning may intuitively display multiple fixed points. If both A and C are negative and large, coordination is still positive and large. The payoff matrix describes an anticoordination game and both players prefer outcomes $(1, 2)$ and $(2, 1)$. If A is positive and large and C is negative and large, coordination is negative and large, one player prefers outcomes $(1, 1)$ and $(2, 2)$, and the other prefers $(1, 2)$ and $(2, 1)$. This is a discoordination game; intuitively, learning may not converge to a fixed point.

The parameter B is a measure for the dominance of player Row's first strategy over her second, and similarly D measures the dominance of player Column's first strategy over her second. We refer to $|BD|$ as "dominance" parameter (we take the absolute value of the product BD because its sign only determines which profile of strategies is selected as the NE, but does not change the type of game). If dominance is large the payoff matrix describes a dominance-solvable game and it is sensible that the learning dynamics is characterized by a unique fixed point, close to the pure strategy NE.

These statements are made more precise in the following proposition:

Proposition 1. *(i) In symmetric games ($A = C$, $B = D$), where coordination (A^2) and dominance (B^2) are positive, it is equivalent to consider $|A|$ as the coordination parameter and $|B|$ as the dominance parameter. If coordination is larger than dominance ($|A| > |B|$), the payoff matrix describes a coordination (if $A > 0$) or anticoordination (if $A < 0$) game. Viceversa, if $|A| < |B|$, it describes a dominance-solvable game.*

(ii) In asymmetric games ($A \neq C$, $B \neq D$), if coordination in absolute value is smaller than dominance ($|AC| < |BD|$), the game is dominance-solvable; in the opposite case ($|AC| > |BD|$), we cannot disambiguate between the classes of games using only these parameters. In particular, if both $|B| < |A|$ and $|D| < |C|$, the payoff matrix describes a coordination (if $AC > 0$, $A > 0$, $C > 0$), anticoordination (if $AC > 0$, $A < 0$, $C < 0$) or discoordination (if $AC < 0$) game. On the other hand, if $|B| > |A|$ or $|D| > |C|$, even if $|AC| > |BD|$, the game is dominance-solvable. However, the larger the value of coordination (compared to dominance), the less likely the payoff matrix describes a dominance-solvable game.

The proof of Proposition 1 is in Appendix A, where we also show that there are only 4 effective degrees of freedom in the payoff matrix, for what concerns the NE and the dynamical properties of EWA learning.

3 The learning model

In this section we describe Experience-Weighted Attraction (EWA) learning and we list all mathematical simplifications that ease the subsequent analysis. In Section 3.1 we provide a formal definition of EWA and discuss the meaning of its parameters. In Section 3.2 we start to simplify the dynamics by assuming that the experience (one of the EWA components) has already reached a steady state and by taking a deterministic limit. In Section 3.3 we specify a diffeomorphism that allows to substantially simplify the equations governing the learning dynamics, with no loss in generality.

3.1 Experience-Weighted Attraction learning

Camerer and Ho (1999) proposed EWA as a hybrid of reinforcement (the players learn on the basis of the performance of their actions) and belief learning (the players construct beliefs on the possible actions of their opponents and respond to these beliefs). They noticed that the two largely studied classes of learning algorithms are in fact equivalent if the players also consider forgone payoffs.[5] Thanks to the generality of EWA, the fit with experimental data is better than with pure reinforcement or pure belief learning. The reason is that real players learn using both information about performance and beliefs.

We now introduce some notation. Consider a 2-person, 2-strategy normal form game. We index the players by $\mu \in \{\text{Row} = R, \text{Column} = C\}$ and the pure strategies by $i = 1, 2$. We denote by $x(t)$ the probability for player R to play pure strategy 1 at time t, and by

[5] The other requirement is that they average between the current payoff for a certain strategy and the past tendency to play the same strategy, see Section 3.2.

$y(t)$ the probability for player C to play pure strategy 1 at time t.[6] We further denote by $s_\mu(t)$ the pure strategy which is actually chosen by player μ at time t, so that $\Pi^\mu(i, s_{-\mu}(t))$ represents the payoff that player μ receives at t if she plays pure strategy i and the other player chooses the pure strategy $s_{-\mu}(t)$.

In EWA, the mixed strategies are determined from the so-called *attractions* or *propensities* $Q_i^\mu(t)$ following a logit rule. For example, the probability for player R to play pure strategy 1 is given by

$$x(t+1) = \frac{e^{\beta Q_1^R(t+1)}}{e^{\beta Q_1^R(t+1)} + e^{\beta Q_2^R(t+1)}}, \tag{3}$$

where β is the payoff sensitivity or *intensity of choice*[7] and a similar expression holds for $y(t+1)$. The propensities update as follows:

$$Q_i^\mu(t+1) = \frac{(1-\alpha)\mathcal{N}(t)Q_i^\mu(t) + (\delta + (1-\delta)I(i, s_\mu(t+1))\Pi^\mu(i, s_{-\mu}(t+1))}{\mathcal{N}(t+1)}, \tag{4}$$

where $\mathcal{N}(t+1) = (1-\alpha)(1-\kappa)\mathcal{N}(t)+1$. Here $\mathcal{N}(t)$ represents *experience* and increases with the number of rounds played; the more it grows, the smaller becomes the influence of the received payoffs on the attractions. The propensities change according to the received payoff when playing action i against the strategies $s_{-\mu}$ by the other players, i.e. $\Pi^\mu(i, s_{-\mu}(t+1))$. The indicator function $I(i, s_\mu(t+1))$ is equal to 1 if i is the actual pure strategy that was played by μ at time $t+1$, that is $i = s_\mu(t+1)$. All attractions (those corresponding to strategies that were and were not played) are updated with weight δ, while an additional weight $1-\delta$ is given to the specific attraction corresponding to the strategy that was actually played. Finally, the memory loss parameter α determines how quickly previous attraction and experience are discounted and the parameter κ interpolates between cumulative and average reinforcement learning (see below).

3.2 Steady state experience and deterministic limit

Here we make two substantial, albeit rather innocuous, simplifications. First, EWA has two state variables: attraction and experience. The dynamics of the latter is trivial, as it reaches a fixed point extremely fast (for many combinations of parameters, the time scale of convergence is of the order of 2-3 time steps). Therefore we assume, with a small loss in generality, that experience has already reached a fixed point $\mathcal{N}^\star$ when the dynamics starts. To ensure the existence of such a fixed point we need to assume that $(1-\alpha)(1-\kappa) < 1$. This restriction only rules out standard fictitious play, in which all past actions are taken into account with the same weight and therefore the relative weight of the most recent actions becomes smaller and smaller. There is no further loss in generality, as all other reinforcement and belief learning algorithms can still be viewed as a particular case of the EWA dynamics once $\mathcal{N}(t)$ has reached a fixed point.

The update rule (4) now reads:

$$Q_i^\mu(t+1) = (1-\alpha)Q_i^\mu(t) + (1-(1-\alpha)(1-\kappa))(\delta + (1-\delta)I(i, s_\mu(t+1))\Pi^\mu(i, s_{-\mu}(t+1)). \tag{5}$$

The interpretation for κ is now more transparent: if $\kappa = 1$ the past payoffs are cumulated, hence cumulative reinforcement learning; if $\kappa = 0$ the past attraction and the current payoff are averaged with weight given by the memory loss parameter α, hence average reinforcement learning. Note that the two learning algorithms can be made equivalent by rescaling the propensities (or equivalently the intensity of choice) by α (see Galla and Farmer 2013).

[6] Due to the normalization condition, the learning dynamics is fully characterized by $\{x(t), y(t)\}_{t=0}^\infty$.

[7] The larger β, the more the players consider the attractions in determining their strategy. In the limit $\beta \to \infty$ the players choose with certainty the pure strategy with the larger attraction. In the limit $\beta \to 0$ they choose randomly, disregarding the attractions.

Following Camerer and Ho (1999), note that belief learning is recovered if $\delta = 1$ and at least one of the following conditions is satisfied:

- There is no memory ($\alpha = 1$). If, in addition, $\beta \to \infty$, one recovers best-response dynamics (Cournot, 1838), in that the players best respond to the last period beliefs only;
- Average reinforcement learning ($\kappa = 0$).

Therefore, by studying the dynamic properties of (3) and (5) we are considering a wide class of learning algorithms, including reinforcement learning, best-response dynamics and weighted fictitious play. As a benchmark case we consider cumulative reinforcement learning (Section 4), which excludes belief learning, but we allow for average reinforcement learning in Section 5.2, where we generalize the results to belief learning.

We make another bold assumption in this section, which will then be relaxed in Section 5.1: we assume that the players play against each other many times before updating their propensities, so that the empirical frequency of their moves corresponds to their mixed strategy. This sort of argument was already made by Crawford (1974) and justified by Conlisk (1993a) in terms of "two-rooms experiments": the players only interact through a computer console and need to specify several moves before they know the moves of their opponent. This assumption is useful from a theoretical point of view and does not affect the results in most cases (Section 5.1): the only difference when noise is allowed is a blurring of the dynamical properties.

We denote by $\overline{\Pi_i^\mu}$ the expected payoff for player μ playing pure strategy i at time t, given that player $-\mu$ plays a distribution of strategies given by her mixed strategy. An important remark is that, under the deterministic assumption, it is intended that $\delta = 1$, as it would be ambiguous to distinguish between the strategies which were and were not played (as long as the players choose a non-degenerate mixed strategy, both pure strategies would be chosen by each player with non-zero frequency), so in order to recover belief learning it is really just enough to consider average reinforcement learning ($\kappa = 0$).

Finally, it is useful to combine (3) and (5) and to write the probabilities $x(t+1)$ and $y(t+1)$ directly in terms of the same probabilities at time t, that is $x(t)$ and $y(t)$. In the deterministic limit (and so with $\delta = 1$) we get

$$x(t+1) = \frac{x(t)^{1-\alpha} e^{\beta(1-(1-\alpha)(1-\kappa))\overline{\Pi_1^R}(y(t))}}{\mathcal{Z}_x}, \tag{6}$$

where $\mathcal{Z}_x = x(t)^{1-\alpha} e^{\beta(1-(1-\alpha)(1-\kappa))\overline{\Pi_1^R}(y(t))} + (1 - x(t))^{1-\alpha} e^{\beta(1-(1-\alpha)(1-\kappa))\overline{\Pi_2^R}(y(t))}$ and an analogous expression holds for $y(t+1)$.

3.3 Transformed coordinates

The remaining simplification implies no loss of generality and matters for technical reasons: we propose a diffeomorphism that transforms the coordinates of the learning dynamics and leads to a simpler set of equations. As we consider the combinations of parameters in the transformed coordinates, the taxonomy of the learning dynamics starts naturally to emerge. A diffeomorphism between a coordinate space (x, y), henceforth denoted by *original coordinates*, to a coordinate space $(\tilde{x}, \tilde{y})$, henceforth denoted by *transformed coordinates*, leaves the dynamical properties (e.g. Jacobian, Lyapunov Exponents) in (x, y) unchanged in $(\tilde{x}, \tilde{y})$, thanks to a well-known property in dynamical systems theory (Ott, 2002).

We consider the generic 2×2 payoff bimatrix (1) and the diffeomorphism

$$\begin{aligned} \tilde{x} &= -\frac{1}{2} \ln \left(\frac{1}{x} - 1 \right), \\ \tilde{y} &= -\frac{1}{2} \ln \left(\frac{1}{y} - 1 \right). \end{aligned} \tag{7}$$

In terms of the transformed coordinates, the map (6) writes:

$$\begin{aligned}\tilde{x}(t+1) &= (1-\alpha)\tilde{x}(t) + \beta(1-(1-\alpha)(1-\kappa))(A\tanh\tilde{y}(t)+B),\\ \tilde{y}(t+1) &= (1-\alpha)\tilde{y}(t) + \beta(1-(1-\alpha)(1-\kappa))(C\tanh\tilde{x}(t)+D),\end{aligned} \tag{8}$$

where A, B, C and D have been defined in (2).

The original coordinates are restricted to $x(t) \in [0,1]$ and $y(t) \in [0,1]$, the transformed coordinates on the other hand take values on the entire real axis. Pure strategies $(x, y) \in \{(0,0),(0,1),(1,0),(1,1)\}$ in the original coordinates map to $(\tilde{x}, \tilde{y}) \in \{(\pm\infty, \pm\infty)\}$ in the transformed coordinates. Note also that mixed strategies where the players choose among their actions with the same probability, i.e. $x, y = 1/2$, are mapped to $\tilde{x}, \tilde{y} = 0$.

The inverse transformation is given by

$$\begin{aligned} x &= \frac{1}{1+e^{-2\tilde{x}}},\\ y &= \frac{1}{1+e^{-2\tilde{y}}}.\end{aligned} \tag{9}$$

4 Taxonomy of learning dynamics

We analyse the dynamical properties of EWA learning in generic 2×2 games. In Section 4.1 we analyse the existence and the stability of fixed points. Depending on the class of game and on the learning parameters, the fixed points may be stable or unstable. In Section 4.2 we simulate the learning dynamics in settings where the fixed points are unstable. We find limit cycles or low-dimensional chaos in discoordination games. In order to quantify the prevalence of each class of games, in Section 4.3 we draw payoff matrices at random, showing that dominance-solvable games are the most common, coordination games can be observed if the payoffs are positively correlated, while discoordination games are associated with negatively correlated payoffs. Finally, in Section 4.4 we state a formal result: under EWA learning, the profiles of pure strategies, despite being fixed points of the learning dynamics, are always unstable (but a nearby stable fixed point may exist).

For convenience, in this section we focus on deterministic learning (which implies $\delta = 1$) and consider cumulative reinforcement learning ($\kappa = 1$). The extensions are considered in Section 5.

4.1 Fixed point analysis

We first analyse the existence and the position of the fixed points in the strategy simplex, and then we consider their stability. In Section 4.1.1 we focus on symmetric games. We find that there exists always at least one stable fixed point, which may or may not correspond to the NE. In Section 4.1.2 we consider "antisymmetric" games, where, for any combination of strategies, the payoffs received by one player are the opposite of the payoffs received by the other player (this does not necessarily correspond to zero-sum games, see below). For discoordination games, the learning dynamics may not settle to a fixed point. Finally, in Section 4.1.3, we analyse the most general class of asymmetric games.

4.1.1 Symmetric games

We start from the simplest case from the point of view of the analysis, namely symmetric 2×2 games. Due to symmetry, both players have identical payoff parameters, $\Pi^R_{ij} = \Pi^C_{ij}$, so $A = C$ and $B = D$. Therefore, coordination is A^2 and dominance is B^2. Recall from Proposition 1 that if $|A| > |B|$ and $A > 0$, the payoff matrix describes a coordination game; if $|A| > |B|$ and $A < 0$, the payoff matrix describes an anticoordination game; if $|B| > |A|$, the game is dominance-solvable.

The fixed points in the transformed coordinates can be obtained from (8), by setting $\tilde{x}(t+1) = \tilde{x}(t) = \tilde{x}^\star$ and $\tilde{y}(t+1) = \tilde{y}(t) = \tilde{y}^\star$. The fixed point equation is $\tilde{x}^\star = \Psi(\tilde{x}^\star)$, where

$$\Psi(\tilde{x}^\star) = \frac{\beta}{\alpha}\left[A \tanh\left(\frac{\beta}{\alpha}(A \tanh \tilde{x}^\star + B)\right) + B\right]. \tag{10}$$

An identical expression holds for $\tilde{y}^\star$. Note that the EWA parameters α and β combine as the ratio α/β (or β/α). It makes sense to define α/β as the "irrationality" parameter because it is large if there is substantial memory loss and/or small intensity of choice. Eq. (10) can have either 1 or 3 solutions. If there are 3 intersections between $\Psi(\tilde{x}^\star)$ and the $\tilde{x}^\star$ line, we denote as *central solution* the intersection with an intermediate value for $\tilde{x}^\star$ and by *lateral solutions* the intersections with the maximum and minimum values. Note that the fixed points are a vector $(\tilde{x}^\star, \tilde{y}^\star)$, so it is not enough to compute the solutions of Eq. (10), one also needs to find the right couplings by replacing the possible combinations in (8).[8]

Thanks to the fact that the maps (6) and (8) are topologically conjugate, their Jacobian is the same. We compute it from (8):

$$J|_{\tilde{x}^\star, \tilde{y}^\star} = \begin{pmatrix} 1-\alpha & \frac{A\beta}{\cosh^2(\tilde{y}^\star)} \\ \frac{C\beta}{\cosh^2(\tilde{x}^\star)} & 1-\alpha \end{pmatrix}. \tag{11}$$

The eigenvalues are

$$\lambda_\pm = 1 - \alpha \pm |A|\,\beta \frac{1}{\cosh(\tilde{x}^\star)\cosh(\tilde{y}^\star)}. \tag{12}$$

Since $1 - \alpha > 0$, the leading eigenvalue is λ_+ and it is enough to study that for the stability properties. After a little algebra we get the stability condition

$$\frac{\alpha}{\beta}\cosh(\tilde{x}^\star)\cosh(\tilde{y}^\star) - |A| \geq 0. \tag{13}$$

The shape of $\Psi(\tilde{x}^\star)$ varies according to the irrationality (α/β), coordination $(|A|)$ and dominance $(|B|)$ parameters. Due to the strong non-linearity of $\Psi(\tilde{x}^\star)$, it is not possible to study it analytically in full. Therefore, we first solve Eq. (10) numerically, and then provide a mathematical analysis of a number of specific cases. Figure 2 shows the properties of the fixed points obtained from the numerical solution of (10), keeping irrationality constant, i.e. $\alpha/\beta = 1$ (since the parameters combine as $\frac{\beta}{\alpha}A$ and $\frac{\beta}{\alpha}B$, it is equivalent to change the values of A and B). We also check the stability of the fixed points by using Eq. (13). We find that there is always at least one stable fixed point. If there are multiple fixed points, only the lateral solutions are stable. For small values of the payoffs, such that the players do not have strong incentives to choose a specific pure strategy, learning converges to a mixed strategy fixed point, where the players randomly choose between the pure strategies. If dominance is larger than coordination, the payoff matrix describes a dominance-solvable game and learning converges to a pure strategy fixed point corresponding to the NE. If coordination is larger than dominance, the payoff matrix may represent a coordination or an anticoordination game. Note that multiple fixed points are much more likely in anticoordination games. To see why this is the case, consider the following payoff matrices, with $A = C = \pm 1.75$ and $B = D = 1.25$:

$$\Pi_{coord} = \begin{pmatrix} 6,6 & 0,0 \\ 0,0 & 1,1 \end{pmatrix}; \Pi_{anticoord} = \begin{pmatrix} 0,0 & 6,1 \\ 1,6 & 0,0 \end{pmatrix}. \tag{14}$$

While in Π_{coord} the NE which yields payoffs (6,6) is to be clearly preferred over the NE yielding (1,1), so it is reasonable that learning only converges to the preferred outcome (unique pure strategy fixed point), in $\Pi_{anticoord}$ there is no preferred NE, so it is sensible that learning displays multiple fixed points. This is indeed what happens in the top right and top left corners of Figure 2, which correspond to the payoff matrices in (14).

[8] It never occurs that the components of a fixed point are the central and lateral solutions: either both components are central solutions, or both components are lateral solutions.

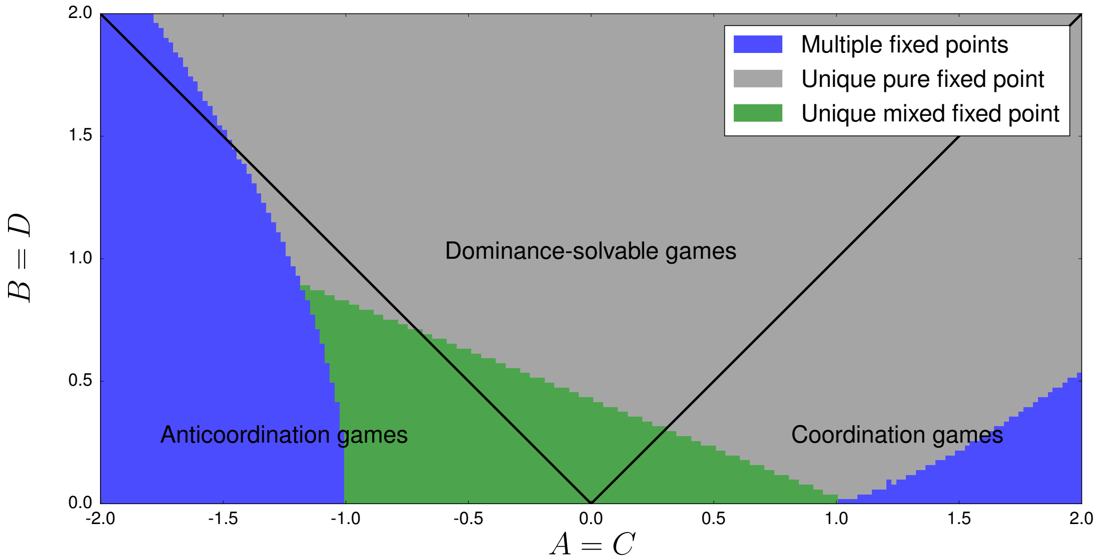


Figure 2: Numerical solution of Eq. (10) for $\alpha/\beta = 1$ and several values of $A = C$ and $B = D$. If $0.3 < x^\star < 0.7$ and $0.3 < y^\star < 0.7$, the solution is classified as a "mixed strategy fixed point". All unique fixed points are stable; if there are multiple fixed points, only the "lateral solutions" are stable. The overall picture is that for small values of the payoffs learning converges to a mixed strategy fixed point; if dominance is strong, to a pure strategy fixed point; if coordination is strong, to multiple fixed points. There are noticeable differences between coordination and anticoordination games. This figure corresponds to the $\frac{\beta}{\alpha}AC > 0$ semiplane in Figure 1, up to folding along the y-axis (because both $A = C < 0$ and $A = C > 0$ yield $AC > 0$).

We now proceed with the mathematical solution for a number of specific cases. We first set $B = 0$, and study the interplay of the coordination parameter, $|A|$, and the irrationality parameter, α/β. The lateral solutions do not exist if

$$\frac{\beta}{\alpha}|A| \leq 1. \tag{15}$$

The interpretation is straightforward: if irrationality is large (so its inverse is small) and/or coordination is small (i.e. the absolute value of the payoffs is small), there is a unique fixed point in the centre of the strategy simplex. This transition can be seen for $B = 0$ and $A = \pm 1$ in Figure 2.

We now consider $B > 0$. It is possible to check in Eq. (10) that a large value of $\frac{\beta}{\alpha}|B|$ flattens $\Psi(\tilde{x}^\star)$ (because it makes the argument less sensitive to the values of $\tilde{x}^\star$) and moves the offset $\Psi(0)$ away from zero. Therefore, for a sufficiently large value of $\frac{\beta}{\alpha}|B|$ there is a unique fixed point far from the centre of the simplex. This is indeed what happens in Figure 2.

Stability is addressed in the following proposition.

Proposition 2. *We consider a symmetric* 2×2 *game. The following results hold:*

(i) if $B = 0$ *and* $\frac{\beta}{\alpha}|A| \leq 1$*, the unique fixed point is stable.*

(ii) if $B = 0$ *and* $\frac{\beta}{\alpha}|A| \to 1^+$ *or* $\frac{\beta}{\alpha}|A| \to +\infty$*, the fixed point whose components are the central solutions is unstable and the fixed points whose components are the lateral solutions are stable. In particular, at* $\frac{\beta}{\alpha}|A| = 1$ *a supercritical pitchfork bifurcation occurs.*

(iii) if $\frac{\beta}{\alpha}|B| \to +\infty$ *and* $B \gg A$*, the unique fixed point is stable.*

The proof is in Appendix B.

4.1.2 "Antisymmetric" games

So far, we analysed the learning dynamics in dominance-solvable, coordination and anti-coordination games. We now want to focus on the remaining class of 2×2 games, namely discoordination games. As these are not symmetric by definition, we consider "antisymmetric" games, where $\Pi^R_{ij} = -\Pi^C_{ij}$, and so $A = -C$, $B = -D$. Note that this condition does not generally define zero-sum games, as the latter are rather defined by the equality $\Pi^R_{ij} = -\Pi^C_{ji}$ (so the two classes of games correspond only if $\Pi^R_{ij} = -\Pi^C_{ij} = 0$ for $i \neq j$). Again, if $B > A$ the game is dominance-solvable, but if $A > B$ we have a discoordination game.

The fixed points $(\tilde{x}^\star, \tilde{y}^\star)$ are again obtained from (8):

$$\begin{aligned} \tilde{x}^\star &= \frac{\beta}{\alpha}\left[-A\tanh\left(\frac{\beta}{\alpha}(A\tanh\tilde{x}^\star + B)\right) + B\right], \\ \tilde{y}^\star &= \frac{\beta}{\alpha}\left[-A\tanh\left(\frac{\beta}{\alpha}(A\tanh\tilde{y}^\star + B)\right) - B\right], \end{aligned} \tag{16}$$

where we have used the identity $\tanh(-x) = -\tanh(x)$. It is immediate to note from (16) that there exists a unique fixed point. Indeed, the functions on the RHS are monotonically decreasing, so only one intersection with the $\tilde{x}^\star$ and $\tilde{y}^\star$ lines is possible. Moreover, given $AC = -A^2 < 0$, the eigenvalues of the Jacobian (11) are complex:

$$\lambda_\pm = 1 - \alpha \pm i\frac{\beta\,|A|}{\cosh(\tilde{x}^\star)\cosh(\tilde{y}^\star)}. \tag{17}$$

The stability condition reads:

$$\frac{\beta}{\sqrt{2\alpha - \alpha^2}}\frac{|A|}{\cosh(\tilde{x}^\star)\cosh(\tilde{y}^\star)} \leq 1. \tag{18}$$

In Figure 3 we show the properties of the unique fixed point obtained from the numerical solution of (16), for several values of A and B. We also check the stability of the fixed points by using Eq. (18). Focusing on small B, a larger value of $|A|$ or a smaller value of $\sqrt{2\alpha - \alpha^2}/\beta$ (which is close to the irrationality parameter α/β) imply a more likely instability. The intuition is straightforward: if the players are rational and/or have strong incentives to switch a strategy which is not performing well, they follow the cycle of best-responses in the payoff matrix and keep switching their moves, rather than smoothly converging to a fixed point in the centre of the simplex, where they would randomize between the pure strategies. On the contrary, if B is large (with respect to A), the learning dynamics simply converges towards a fixed point close to the pure strategy NE.

We conclude this section by focusing on one specific example of discoordination games, where dominance is null: $B = D = 0$, $C = -A$. The unique fixed point is $(0, 0)$ and is stable if (we assume without loss of generality $A > 0$, $C < 0$)

$$\frac{\beta}{\sqrt{2\alpha - \alpha^2}}A \leq 1. \tag{19}$$

Replacing the values of α and β used in Figure 3, the fixed point becomes unstable for $A^\star = 1.224$, which corresponds to the transition observed in Figure 3.

4.1.3 Asymmetric games

We now consider asymmetric 2×2 games. In general, the payoff parameters are different for the two players, so $A \neq C$ and $B \neq D$. There is a larger variety of behaviours, but in general asymmetric games are widely similar to their symmetric counterparts (e.g. asymmetric dominance-solvable games are widely similar to symmetric dominance-solvable games), except that the player with the strongest incentive to play a certain move plays mixed strategies farther from the centre of the strategy simplex.

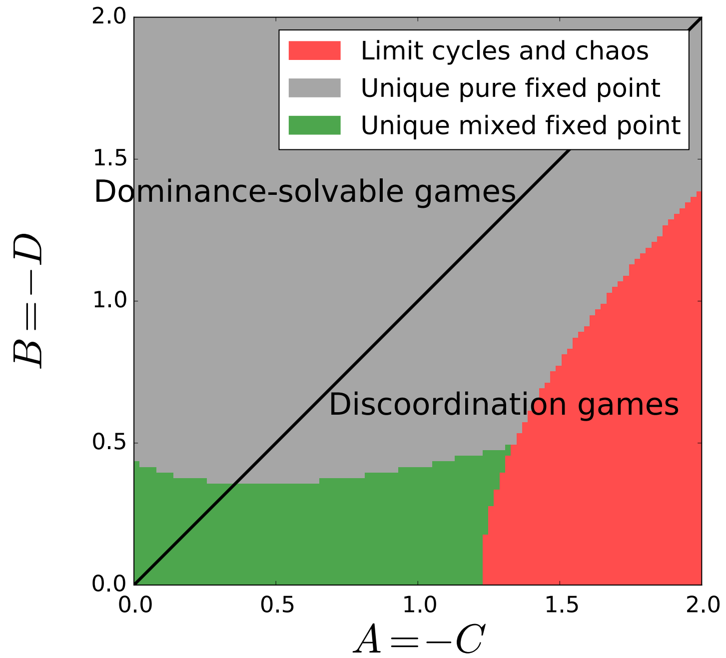


Figure 3: Numerical solution of Eq. (20) for $\alpha = \beta = 0.8$ and several values of $A = -C$ and $B = -D$. If $0.3 < x^\star < 0.7$ and $0.3 < y^\star < 0.7$, the solution is classified as a "mixed strategy fixed point". There is always a unique fixed point, which may become unstable in discoordination games for low values of irrationality and/or high (absolute) values of coordination. The intuition is that the players have strong incentives to try and improve their payoffs, so they fail to coordinate to the mixed strategy NE and the learning dynamics keeps cycling. This figure corresponds to the $\frac{\beta}{\alpha}AC < 0$ semiplane in Figure 1.

The fixed points $(\tilde{x}^\star, \tilde{y}^\star)$ are given by

$$\begin{aligned} \tilde{x}^\star &= \frac{\beta}{\alpha}\left[A \tanh\left(\frac{\beta}{\alpha}(C \tanh \tilde{x}^\star + D)\right) + B\right], \\ \tilde{y}^\star &= \frac{\beta}{\alpha}\left[C \tanh\left(\frac{\beta}{\alpha}(A \tanh \tilde{y}^\star + B)\right) + D\right]. \end{aligned} \tag{20}$$

Without loss of generality, we can write the (combinations of) payoffs of player Column as a rescaled version of the (combinations of) payoffs of player Row, that is $C = WA$ and $D = ZB$, with W and Z scale factors. The magnitudes of W and Z quantify the imbalance in coordination and dominance for the two players. For instance, if W is large, player Column has stronger incentives to converge on one of the pure strategy NE (just consider the payoff matrix (1), with $a = d = 1$, $e = h = 5$, $b = c = f = g = 0$). Consistently, the height of the hyperbolic tangent (20) for player Column is larger, leading to an intersection with the $\tilde{y}^\star$ line which is farther away from zero ($\tilde{y}^\star \gg \tilde{x}^\star$). Therefore, player Column will choose a mixed strategy farther from the centre of the simplex. Likewise, if Z is large (consider $a = 1$, $d = -1$, $e = 5$, $h = -5$, $b = c = f = g = 0$), player Column ends up to a fixed point closer to the pure strategy NE. Concerning the signs, the sign of Z does not matter in determining the stability properties, while the sign of W has a substantial effect. If $W > 0$, we find little difference with the symmetric case; if $W < 0$ the payoff matrix may define discoordination games (the additional condition is that $|B| < |A|$ and $|D| < |C|$), which may have no stable fixed points, as shown in Section 4.1.2.

4.2 Simulations of the unstable dynamics

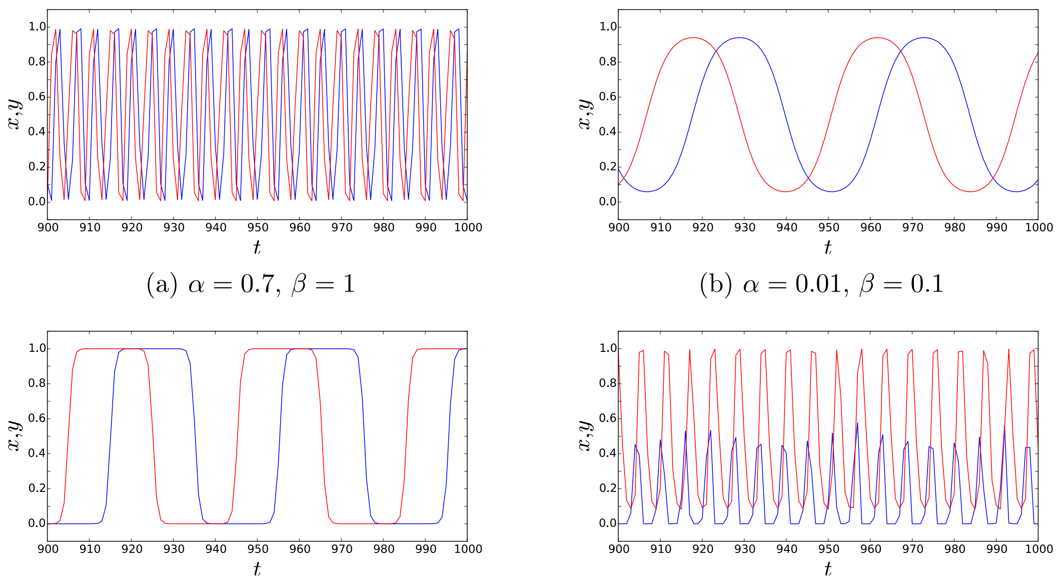
We choose a parameter setting where the fixed point of the discoordination game is unstable. Figure 4 shows some examples of the dynamics for some values of α and β, for a given payoff matrix. The dynamics superficially looks like following a limit cycle, whose shape is governed by α and β: Fig. 4a shows that, for high α and β, the players frequently change their strategies, whereas in Fig. 4b, for low values of α and β, the dynamics is smoother; in Fig. 4c, where α is very small but β is reasonably high, the players spend a lot of time playing mostly one pure strategy and then quickly switch to the other one (because they have long memory). Finally, in Fig. 4d we choose $B \neq 0$: the discrepancy between the pure strategies seems to yield the most irregular dynamics.

In order to get further insights into the learning dynamics, Figure 5 represents the bifurcation diagrams and the largest Lyapunov Exponent (LLE),[9] varying α and β. We focus on the values of the payoff matrix in Fig. 4d, as the behaviour of the learning dynamics in Fig. 4a is only marginally chaotic. Figures 5a and 5c refer to a parameter setting where the fixed point is unstable, and we observe alternating limit cycles and chaotic bands. On the other hand, in Figures 5b and 5d, for small intensity of choice, that is $\beta \in (0, 0.5)$, we observe convergence towards the fixed point in the centre of the simplex. Interestingly, for values $\beta \in (0.5, 0.8)$ the dynamics is not periodic, but the LLE is almost null. This case corresponds to a marginally chaotic behaviour, like the one in Fig. 4a. For larger values of β we observe again chaotic bands and limit cycles. At the points where the limit cycles become chaotic we can observe a higher density of trajectories, probably related to the intermittency scenario of transition to chaos (Pomeau and Manneville, 1980).

Figure 6 shows[10] that chaos is more frequently observed if one of the pure strategies is dominant over the other, $B > 0$. Moreover, the LLE is always negative if $B > A$ (consistently

[9] Since the system is 2-dimensional, in order to compute the Lyapunov exponents it is necessary to periodically orthogonalize the unit vectors using a Gram-Schmidt procedure, see Benettin et al. (1980). Note that, while this is strictly necessary only in order to obtain the whole Lyapunov spectrum, and so compute the Kaplan-Yorke dimension, in practice the estimate of the LLE is much more accurate if one uses the orthogonalization method even just to compute the LLE.

[10] Since we choose $C = -A$ and $D = -B$, there is a 4-fold symmetry in the AB plane, so we only plot the 1st quadrant.



(a) $\alpha = 0.7$, $\beta = 1$

(b) $\alpha = 0.01$, $\beta = 0.1$

(c) $\alpha = 0.01$, $\beta = 0.5$

(d) $\alpha = 0.7$, $\beta = 1$

Figure 4: Time series of the probabilities x (in blue) and y (in red). The payoff parameters are: $b = c = f = g = 0$, $a = d = 4$ and $e = h = -4$ in panels (a)-(c), and $a = -11.8$, $d = -1.8$, $e = 11.8$ and $h = 1.8$ in panel (d). Cyclical and chaotic dynamics occur.

with the diagram depicted in Figure 3) and is larger for high absolute values of the payoffs (so that A and B are larger).

4.3 Random 2×2 payoff matrices

In the above sections, the taxonomy of learning dynamics is determined by three classes of 2×2 games: dominance-solvable, (anti)coordination and discoordination games. It is interesting to understand the relative frequency of each class of games, under some appropriate probability distribution over the payoff matrices. While it is true that game theory is usually concerned with specific scenarios of strategic interaction, which are properly described by a unique payoff matrix, finding which games are prevalent under a random generation of the payoff matrix would shed light on biological and social phenomena where the nature of the interactions is not known *a priori*. We choose an *ensemble* of payoff matrices obtained by constraining the mean, variance and correlation of the payoff elements. In particular, we assume that the mean is 0, the variance is 1 and the two payoffs for the same profile of pure strategies in the payoff matrix are correlated by a parameter Γ. A value $\Gamma = -1$ would imply anticorrelation and describe a zero-sum game, while negative values for the correlation, $\Gamma < 0$, are more generally associated with competitive games; on the contrary $\Gamma = 1$ reveals perfect correlation and positive values of Γ are related to cooperative games. Finally, $\Gamma = 0$ implies lack of correlation. Under these constraints, the maximum entropy distribution is a bivariate Gaussian with specified mean and covariance matrix (Galla and Farmer, 2013).

Figure 7 represents the fraction of games which belong to each of the three classes, as a function of the correlation parameter Γ. We see that for all values of Γ, dominance-solvable games are the most likely. Positive values of Γ are associated with (anti)coordination games, which display multiple fixed points under EWA learning, whereas for negative values of Γ it is more likely to obtain a discoordination game, and consequently limit cycles or chaos in the learning dynamics. Indeed, this was observed by Galla and Farmer (2013), who find convergence to multiple fixed points in a semiplane $\Gamma > 0$ and unstable dynamics in a semiplane $\Gamma < 0$ (in both cases, α/β should be low). A difference with Galla and

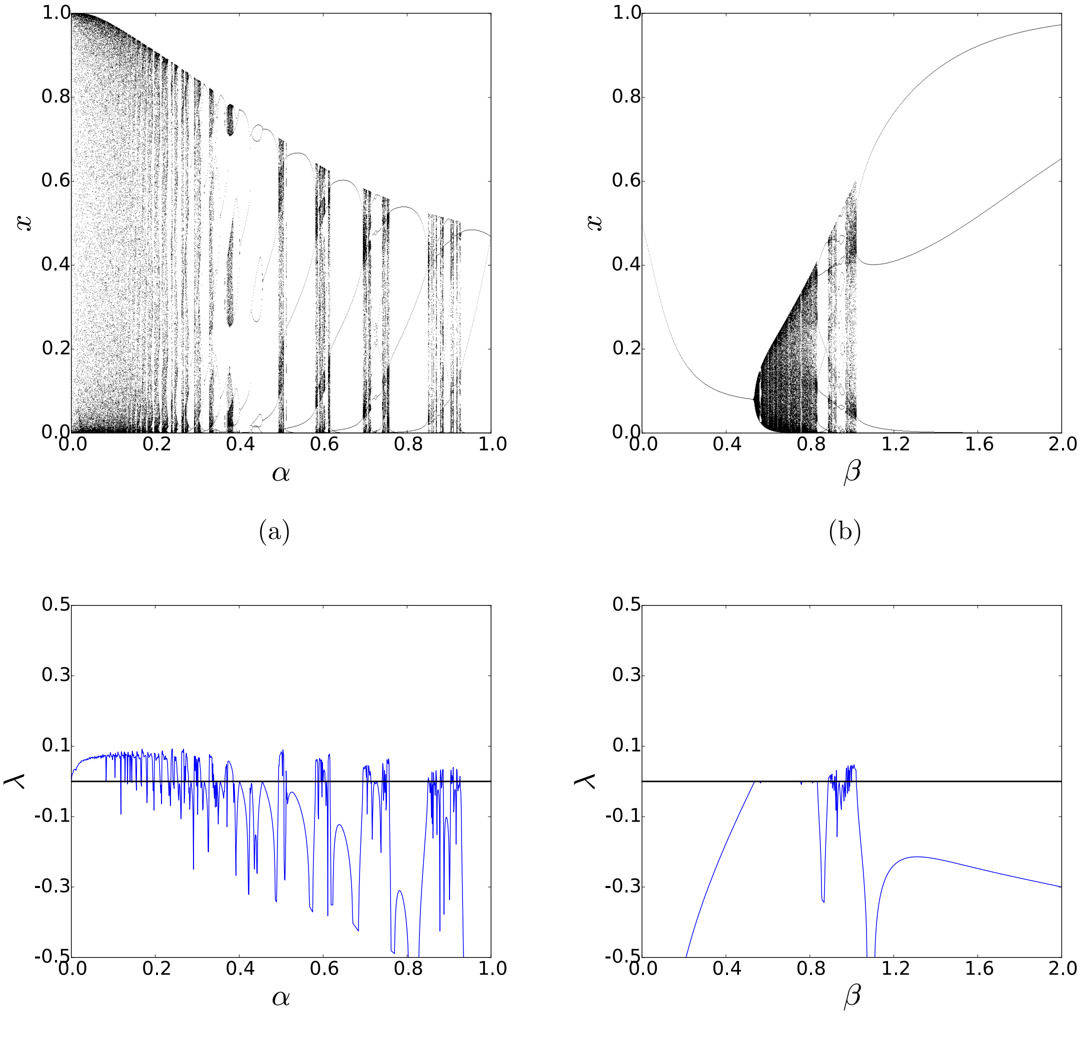
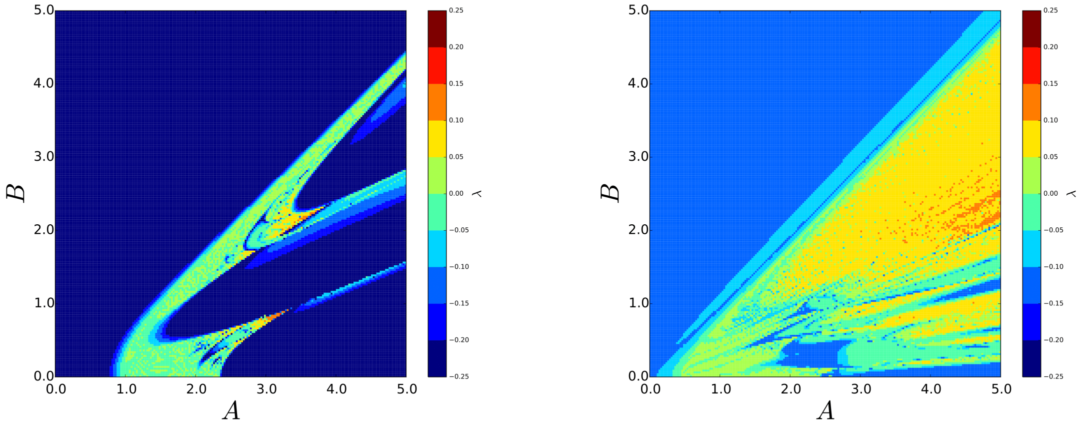


Figure 5: Bifurcation diagrams and largest Lyapunov exponent for $b = c = f = g = 0$, $a = -11.8$, $d = -1.8$, $e = 11.8$ and $h = 1.8$. Panels (a)-(c): $\beta = 1$, α varying. Panels (b)-(d): $\alpha = 0.7$, β varying. Low-dimensional chaos may be observed.



(a) (b)

Figure 6: Largest Lyapunov exponent as a function of the parameters A and B. The colour scale is set such that chaos is observed from green to red. The parameters are: $C = -A$, $D = -B$, $\beta = 1$, $\alpha = 0.7$ in panel (a), $\alpha = 0.1$ in panel (b). Chaos is more likely if there is a discrepancy between the pure strategies, $B > 0$, and the payoffs are quite large in absolute value.

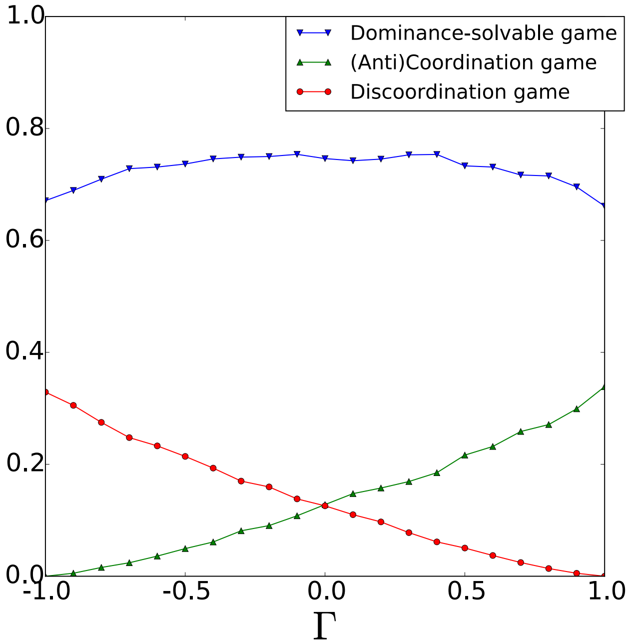


Figure 7: Fraction of dominance-solvable, (anti)coordination and discoordination games, as a function of the correlation Γ. These results are averaged over 10000 random draws of the (Gaussian) payoff matrix.

Farmer (2013) is that, whereas they find consistently unstable behaviour in certain regions of the parameter space, we cannot rule out convergence to a fixed point for $\Gamma < 0$. In fact, for all values of Γ, most payoff matrices describe dominance-solvable games, which always display a stable fixed point. This difference might be explained by the fact that Galla and Farmer (2013) consider high-dimensional strategy spaces, whereas we are restricted to two pure strategies. A reasonable conjecture would be that by increasing the number of pure strategies the fixed points of the learning dynamics may become unstable. We leave the exploration of this conjecture to future work.

4.4 The pure strategy NE are unstable

Here we show that the NE in pure strategies are "infinitely" unstable.

Proposition 3. *Consider a generic 2×2 game and the learning dynamics in the original coordinates* (6). *At the profiles of pure strategies, $(x, y) \in \{(0,0), (0,1), (1,0), (1,1)\}$, for positive memory loss, $\alpha > 0$, the Jacobian has infinite elements along the main diagonal and null elements along the antidiagonal.*

The proof of Proposition 3 is in Appendix C.

A clarification is worth here: while the NE in pure strategies are formally unstable (unless $\alpha = 0$), for most values of the parameters there is a fixed point nearby. In particular, if irrationality is not too high and the absolute values of the payoffs are not too small, it is likely that one of the fixed points will be quite close to the NE in pure strategies. This result could be anticipated, since for, e.g., dominance-solvable games, a reasonable learning dynamics is expected to converge sufficiently close to the NE.

5 Extensions

We generalize the results in Section 4 by relaxing two seemingly restricting assumptions. In Section 5.1 we drop the simplification of deterministic learning and we analyse the stochastic learning dynamics. All previous results still hold and the only effect of this extension is a blurring of the dynamical properties. In Section 5.2 we analyse the more general case where the parameter κ, which interpolates between average and cumulative reinforcement learning, is not restricted to be $\kappa = 1$. This allows us to recover belief learning and to reproduce the well-known result about the convergence of fictitious play in 2×2 games. Again, the analysis is not substantially different, in that it is enough to rescale the intensity of choice by a value proportional to κ.

5.1 Stochastic learning

When playing a game, except for very specific experimental arrangements (Conlisk, 1993a), the players would update their propensities after observing each move by their opponent. This questions whether the deterministic dynamics (6), which assumes that the participants of the game play against each other many times before updating their propensities, provides robust conclusions. We interpolate from the deterministic limit by considering batches of size T, where the players sample their mixed strategies. The limit $T \to \infty$ recovers deterministic learning, whereas actual learning would occur with $T = 1$. As noted in Section 3.2, unless $T = 1$, the meaning of the parameter δ in unclear. Indeed, a value of δ different from 1 implies that the players give an additional update to the attraction corresponding to the move which they chose. This rule is not well defined if they play against each other many times before updating their attractions, as they might choose both pure strategies at least once. However, for $T = 1$ we consider several values of δ and we show that, the lower the value of δ, the more noisy becomes the learning dynamics, as there is an additional source of stochasticity given by which strategy the player randomly chooses, further to which strategy is randomly chosen by his opponent.

It is beyond the scope of this paper to systematically study the effect of noise on the learning dynamics, and we refer the reader to Galla (2009) for a study on the effect of noise on learning, and to Crutchfield et al. (1982) for a more general discussion on the effect of noise on the properties of dynamical systems. In the following we show a few numerical examples where we investigate what happens as we progressively increase the level of noise. We simply describe our findings and we leave most of the numerical support to Appendix D. We stress that the dynamical properties in the deterministic limit, in order to be considered as robust, need to hold down to $T = 1$, as that is the natural choice for a realistic learning dynamics. We focus on the three classes of games which we identified in the paper.

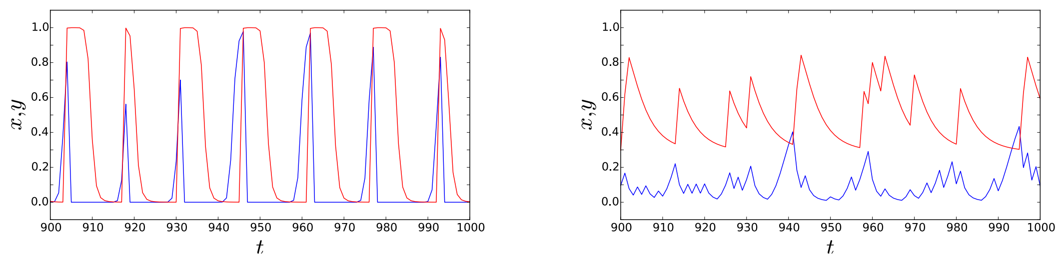
Dominance-solvable games Provided that the irrationality parameter is not too high, the players converge close to the pure strategy NE (Figures D.1 (a)-(d)). After an irregular transient, as the learning dynamics moves close to the faces of the simplex, it becomes remarkably stable. On the contrary, if α/β is high, the players converge to the centre of the simplex, as it occurs with deterministic learning (Figures D.1 (e)-(f)). However, the learning dynamics is much more irregular. The asymptotic learning behaviour is explained by two factors: deviations from the previous moves, and their effect. If the players always played the same moves, the learning dynamics would converge to a fixed point. But as one of them switched her move, we would observe a perturbation from such a fixed point. This explains in part why, close to the centre of the simplex, the learning dynamics is more irregular: the players converge to a mixed strategy where they choose each move roughly with the same probability. The other factor is that the attractions are large at the faces of the simplex, so the relative magnitude of their update (due to the deviation) is smaller. We also observe another pattern in Figures D.1 (a)-(d): the higher the level of noise (i.e., the smaller T and/or the smaller δ), the more irregular is the transient.

(Anti)Coordination games As for dominance-solvable games, we observe convergence to a fixed point close to one of the pure strategy NE (for low levels of irrationality). We investigate whether noise can help reaching the Pareto-Optimal NE, as it does in the theory of stochastic stability (Young, 1993). Given the previous remark on the effect of the noise near the faces of the simplex, we expect that stochastic learning can help reaching the Pareto-Optimal NE only in the first steps of the dynamics. This conjecture is confirmed by the numerical simulations in Figure D.2. We find that EWA is path dependent, differently from the learning algorithms introduced by (Young, 1993), which are based on ergodic Markov Chains. With EWA, the learning dynamics reaches the Pareto-Optimal NE only if there is a favourable fluctuation in the first stage of the dynamics.

Discoordination games In Section 4 we identify two learning behaviours: if irrationality is high, the dynamics converges to the centre of the strategy simplex and the players simply randomize between their moves; if irrationality is low, the players do not converge to an equilibrium and the mixed strategies keep oscillating. This distinction survives when we allow for noise. In Figure 8 we plot the stochastic time series for both behaviours. In Figure 8a the mixed strategy fixed point of the corresponding deterministic dynamics is unstable and the stochastic learning dynamics is chaotic (the parameters are the same as in Fig. 5), whereas in Figure 8b the mixed strategy fixed point is an attractor of the (deterministic) dynamics. It is immediately clear that in the latter case there is a total lack of autocorrelation in the moves by each player (because the dynamics does not spend much time near the faces of the simplex), whereas in the former the autocorrelation function decays more slowly as a function of the time lag. These results are confirmed in Figure 9 and constitute a precise theoretical prediction that can be tested against data on experimental learning of discoordination games. Finally, Figure 10 represents the same bifurcation diagram and largest Lyapunov exponents as in Figs. 5a and 5c respectively, with the only difference that we consider stochastic learning with $T = 1$. For small values of α the LLE is still positive, so the dynamics is chaotic. We consider several values of T in Figure D.3. We observe the equivalence between parametric and additive[11] noise (Crutchfield et al., 1982): the effect of

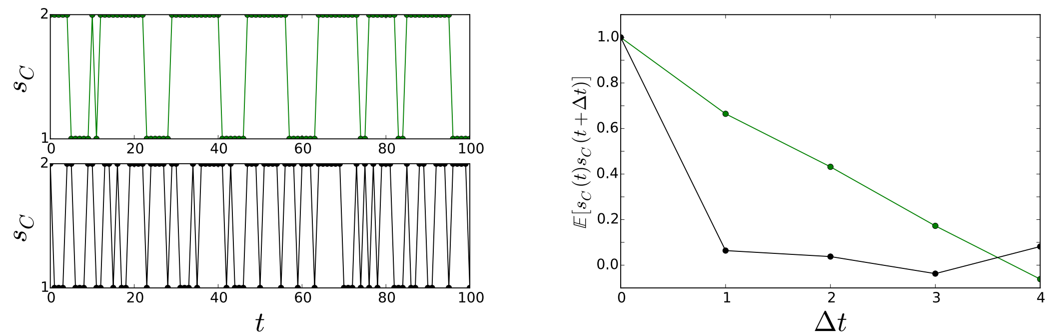
[11] In fact, the noise source induced by finite T is not additive, but it is always possible to express the noise through a properly defined additive stochastic term in the dynamical equations.

noise on the properties of dynamical systems equivalently occurs as a perturbation of their trajectories or as a perturbation of their parameter values. By progressively increasing the level of noise, we observe the smoothing of both the bifurcation diagram and the plot representing the LLE, losing the finely alternating structure with bands of chaos and windows of regularity.



(a) (b)

Figure 8: Time series of the probabilities x (in blue) and y (in red), for stochastic learning with $T = 1$. The payoff parameters are $b = c = f = g = 0$, $a = -11.8$, $d = -1.8$, $e = 11.8$ and $h = 1.8$. The memory loss is $\alpha = 0.2$, the intensity of choice is $\beta = 1$ in panel (a), implying deterministic chaotic behaviour, and $\beta = 0.1$ in panel (b), implying that in the deterministic limit the players reach a fixed point. The two sequences are qualitatively different.



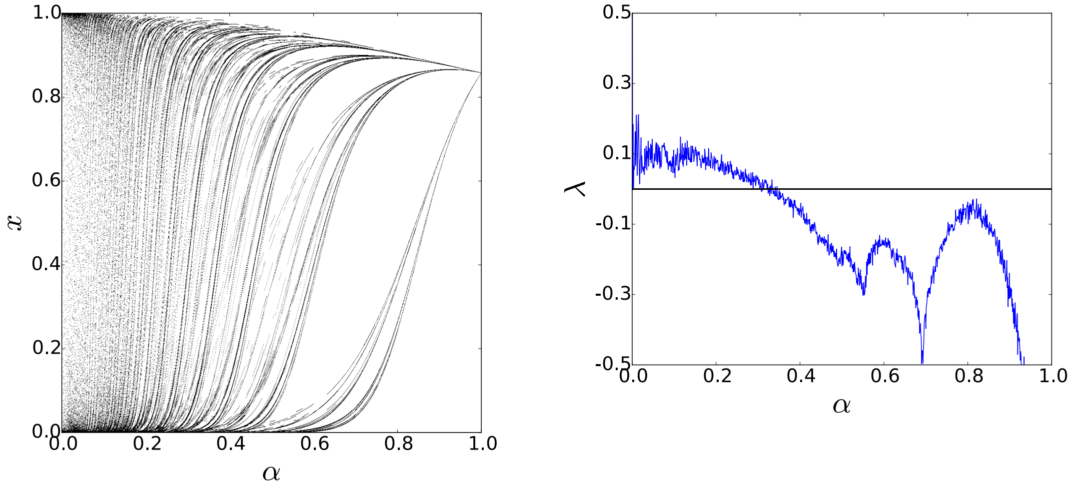
(a) (b)

Figure 9: (a) Time series of the moves of player Column, for stochastic learning with $T = 1$. The upper (lower) panel corresponds to the stochastic learning dynamics in Fig. 8a (8b). (b) Autocorrelation function of the moves of player Column, for both learning dynamics represented in the left panel. If irrationality is high, the players randomize between their moves and the autocorrelation decays instantaneously.

5.2 Belief learning

We drop the assumption of cumulative reinforcement learning ($\kappa = 1$) and we analyse other learning algorithms in the EWA family. Looking at Eqs. (6) and (8), in order to consider a general value for κ, it is sufficient to rescale the intensity of choice β and to replace it by $\tilde{\beta} = \beta\,[1 - (1-\alpha)(1-\kappa)]$. As the quantity multiplying β is lower than one, the intensity of choice is smaller and so the irrationality parameter is larger. Therefore, the learning dynamics is generally more stable, and it is easier to converge to a fixed point in the centre of the simplex.

If $\kappa = 0$ and $\delta = 1$ we recover most forms of belief learning ($\alpha = 1$: best-response dynamics; $\alpha = 0$: fictitious play; $0 < \alpha < 1$: weighted fictitious play). The rescaled



(a) (b)

Figure 10: Bifurcation diagram and largest Lyapunov exponent as a function of α for stochastic learning with $T = 1$. The payoff parameters are $H = -11.8$, $K = -1.8$, $L = 11.8$ and $M = 1.8$, the intensity of choice is $\beta = 1$. Chaos survives for small values of α and we observe the equivalence between additive and parametric noise.

intensity of choice is $\tilde{\beta} = \beta\alpha$. First of all, this means that the coordinates of the fixed points do not depend on α any more (Eqs. (10) and (20)). However, the memory loss sets the timescale of convergence to the fixed points, so in the limit $\alpha \to 0$ convergence may take exponentially long time. Moreover, the stability threshold (19) in discoordination games depends non-linearly on α, which does not cancel out:

$$\frac{\beta\alpha}{\sqrt{2\alpha - \alpha^2}} A \leq 1. \tag{21}$$

The derivative of the LHS with respect to α is positive, and so smaller and smaller values of α make stability more and more likely. In other words, the parameter space where it is possible to observe unstable behaviour shrinks as α is reduced. In the limit $\alpha \to 0$, the LHS goes to zero, so stability is ensured for all parameter values. Note that the case $\alpha = 0$, $\kappa = 0$ is the standard fictitious play learning algorithm (see Ho et al. (2007), Fig. 1) that was ruled out by obtaining the steady state dynamical equations (5), from the more general EWA rule (4). However, by taking the limit $\alpha \to 0$ ex-post, we recover the well known result of Miyazawa (1961) and Monderer and Sela (1996), namely that in non-degenerate 2×2 games fictitious play would converge to the NE.

6 Conclusion

In this paper we have exhaustively characterized the dynamics of EWA learning in generic 2×2 games. We have shown that a variety, or a *taxonomy*, of different behaviours can be observed, according to the properties of the payoff matrix and to the value of the parameters of the learning algorithm. The taxonomy naturally relates to classes of games that have been extensively studied in the literature: in dominance-solvable games we observe convergence towards the unique pure strategy NE; in coordination games we find multiple fixed points corresponding to the NE; in discoordination games the unique mixed strategy NE may be unstable and the learning dynamics may settle in a limit cycle or a low dimensional chaotic attractor. However, for all classes of games, if the players cannot choose with certainty the best performing strategy (because of finite intensity of choice), quickly forget the past performance of their moves and/or have little incentives in terms of payoffs, the learning dynamics converges to a fixed point well in the centre of the simplex, where the players simply randomize between the pure strategies.

The novelty of this work is first of all in its approach: we have identified a number of relevant parameters and classified the learning dynamics accordingly, by *ex-post* relating the values of the parameters to the classes of games described above. In particular, we have found that *irrationality*, defined as the ratio of memory loss α to intensity of choice β, if large implies the convergence to a mixed strategy in the centre of the simplex. We have then defined a *coordination* parameter by computing the difference of the diagonal and the off-diagonal elements in the payoff matrix for both players (A for row and C for column), and multiplying the two numbers (AC). A large positive value of coordination is related to a coordination or an anticoordination game, where the players try and coordinate on the same profiles of pure strategies.[12] If coordination is negative (A is positive and C is negative, or viceversa), the players try and coordinate on different profiles of strategies and there is no pure strategy NE. The payoff matrix defines a discoordination game and, for a good level of rationality, is related to an unstable learning dynamics. The third parameter is called *dominance*. It is obtained as the absolute value of the product of the difference between the payoffs associated with one pure strategy and the payoffs associated with the other one, for both players (B for row and D for column, so dominance is $|BD|$). If large, it is likely that the payoff matrix describes a dominance-solvable game.

Thanks to the exhaustive characterization of 2×2 games, we have found general properties about the features of the payoff matrix that relate to a convergent or non convergent learning dynamics, without the need to assume symmetry or zero or constant sum, as most previous works did; we have provided explicit values for the parameters which define the onset of instability. For the first time, we have fully analysed EWA, showing that its learning properties are more general than those of the learning algorithms it generalizes. In particular, EWA has several free parameters that can be tuned to govern the "speed" of the learning dynamics, thereby facilitating convergence or rather unstable behaviour. Moreover, we have found chaos in a two strategy set, showing that non periodic dynamics can occur even in simple games.

The perspectives of this work span multiple dimensions. From the point of view of the learning algorithms, we aim at generalizing our approach to a form of rule learning, so that the players select the best performing learning algorithms, e.g. through evolutionary dynamics. More sophisticated algorithms may involve a larger information cost, and evolutionary dynamics may select a learning algorithm which implies convergence to the NE, or rather another learning algorithm whereby the players quickly switch their moves as soon as there is a better performing strategy, so that instability is more likely. It is also possible that a fixed point in the space of learning algorithms is not reached, so that the players keep switching their learning rule. From the point of view of the payoff matrices, it is interesting to understand which are the features of larger strategy sets that are associated with stable or unstable dynamics and to see their relative prevalence under some properly defined ensemble. The common feature of the classes of games with most convergence results in the literature is that they are acyclical (Arieli and Young, 2016), so that a chain of best responses converges to a profile of strategies (equivalently, there are no cycles in best responses). However, Foster and Young (1998) show a counterexample of a coordination game where fictitious play does not converge, implying that some less trivial and higher-order cyclical properties probably matter. Game theory has largely focused on classes of games where learning converges, but whether such classes are *typical* has not been thoroughly explored. It is sensible that, by increasing the size of the strategy sets and/or the number of players, unstable dynamics may become prevalent. Some work has in part confirmed this conjecture (Sanders et al., 2016), but a more systematic investigation is required.

[12] In coordination games, where A and C are positive, they try and coordinate on profiles where they play the same strategy. On the contrary, in anticoordination games, where A and C are negative, they try and coordinate on profiles of strategies where their moves are different.

The ultimate goal of this line of research is to test whether learning converges in experiments. Most experiments show approximate aggregate convergence, but the underlying games have usually distinct equilibria and paths of convergent best replies. For general payoff matrices with cycles in best-responses and several players, the players may just endlessly cycle between the profiles of strategies, even in reality, and equilibrium concepts would be meaningless. For what concerns 2×2 games, we may test our theoretical prediction on the unstable behaviour in discoordination games, by looking at the persistence of the experimental individual time series. A footprint of the stability of a mixed strategy fixed point is a lack of autocorrelation between successive moves by the players, whereas cycling behaviour is associated with a slower decay in the autocorrelation function. It is also possible to estimate the EWA parameters given the time series (Camerer and Ho, 1999), and we could check if the learning dynamics would be in the unstable region of the parameter space. Only experiments will be able to provide a definitive answer on the dilemma about the convergence of learning in games.

7 Bibliography

Arieli, I. and Young, H. P. (2016) "Stochastic Learning Dynamics and Speed of Convergence in Population Games," *Econometrica*, Vol. 84, pp. 627–676.

Benaïm, M., Hofbauer, J., and Hopkins, E. (2009) "Learning in games with unstable equilibria," *Journal of Economic Theory*, Vol. 144, pp. 1694–1709.

Benettin, G., Galgani, L., Giorgilli, A., and Strelcyn, J.-M. (1980) "Lyapunov characteristic exponents for smooth dynamical systems and for Hamiltonian systems; a method for computing all of them. Part 1: Theory," *Meccanica*, Vol. 15, pp. 9–20.

Berg, J. and Weigt, M. (1999) "Entropy and typical properties of Nash equilibria in two-player games," *EPL (Europhysics Letters)*, Vol. 48, pp. 129–135.

Brock, W. A. and Hommes, C. H. (1997) "A rational route to randomness," *Econometrica: Journal of the Econometric Society*, pp. 1059–1095.

Brown, G. W. (1951) "Iterative solution of games by fictitious play," in T. Koopmans ed. *Activity analysis of production and allocation*, New York: Wiley, pp. 374–376.

Camerer, C. and Ho, T. (1999) "Experience-weighted attraction learning in normal form games," *Econometrica*, Vol. 67, pp. 827–874.

Conlisk, J. (1993a) "Adaptation in games: Two solutions to the Crawford puzzle," *Journal of Economic Behavior & Organization*, Vol. 22, pp. 25–50.

——— (1993b) "Adaptive tactics in games: Further solutions to the Crawford puzzle," *Journal of Economic Behavior & Organization*, Vol. 22, pp. 51–68.

Cournot, A.-A. (1838) *Recherches sur les principes mathématiques de la théorie des richesses*: L. Hachette.

Crawford, V. P. (1974) "Learning the optimal strategy in a zero-sum game," *Econometrica: Journal of the Econometric Society*, pp. 885–891.

——— (1995) "Adaptive dynamics in coordination games," *Econometrica: Journal of the Econometric Society*, pp. 103–143.

Crutchfield, J. P., Farmer, J. D., and Huberman, B. A. (1982) "Fluctuations and simple chaotic dynamics," *Physics Reports*, Vol. 92, pp. 45–82.

Foster, D. P. and Young, H. P. (1998) "On the nonconvergence of fictitious play in coordination games," *Games and Economic Behavior*, Vol. 25, pp. 79–96.

Galla, T. (2009) "Intrinsic noise in game dynamical learning," *Physical review letters*, Vol. 103, p. 198702.

Galla, T. and Farmer, J. D. (2013) "Complex dynamics in learning complicated games," *Proceedings of the National Academy of Sciences*, Vol. 110, pp. 1232–1236.

Hahn, S. (1999) "The convergence of fictitious play in 3×3 games with strategic complementarities," *Economics Letters*, Vol. 64, pp. 57–60.

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

Hofbauer, J. and Sigmund, K. (1998) *Evolutionary games and population dynamics*: Cambridge university press.

Hommes, C. H., Ochea, M. I., and Tuinstra, J. (2016) "Evolutionary Competition between Adjustment Processes in Cournot Oligopoly: Instability and Complex Dynamics,"Technical report, THEMA (THéorie Economique, Modélisation et Applications), Université de Cergy-Pontoise.

Huberman, B. A. and Hogg, T. (1988) "The behavior of computational ecologies," in *The Ecology of Computation*, pp. 77–115: North-Holland.

Krishna, V. (1992) "Learning in games with strategic complementarities,"Technical report, Harvard Business School.

Milgrom, P. and Roberts, J. (1991) "Adaptive and sophisticated learning in normal form games," *Games and economic Behavior*, Vol. 3, pp. 82–100.

Miyazawa, K. (1961) "On the Convergence of the Learning Process in a 2×2 Non-Zero-sum Two-person Game,"Technical Report Research Memorandum No. 33, Econometric Research Program, Princeton University.

Monderer, D. and Sela, A. (1996) "A $2{\times}2$ Game without the Fictitious Play Property," *Games and Economic Behavior*, Vol. 14, pp. 144–148.

Monderer, D. and Shapley, L. S. (1996) "Fictitious play property for games with identical interests," *Journal of economic theory*, Vol. 68, pp. 258–265.

Muth, J. F. (1961) "Rational expectations and the theory of price movements," *Econometrica: Journal of the Econometric Society*, Vol. 29, pp. 315–335.

Nachbar, J. H. (1990) ""Evolutionary" selection dynamics in games: Convergence and limit properties," *International journal of game theory*, Vol. 19, pp. 59–89.

Nagel, R. (1995) "Unraveling in guessing games: An experimental study," *The American Economic Review*, Vol. 85, pp. 1313–1326.

Opper, M. and Diederich, S. (1992) "Phase transition and 1/f noise in a game dynamical model," *Physical review letters*, Vol. 69, pp. 1616–1619.

Ott, E. (2002) *Chaos in dynamical systems*: Cambridge university press.

Pomeau, Y. and Manneville, P. (1980) "Intermittent transition to turbulence in dissipative dynamical systems," *Communications in Mathematical Physics*, Vol. 74, pp. 189–197.

Rapoport, A., Guyer, M. J., and Gordon, D. G. (1976) "The $2{\times}2$ Game," *Ann Arbor: University of Michigan Press.*

Robinson, J. (1951) "An iterative method of solving a game," *Annals of mathematics*, pp. 296–301.

Sanders, J. B., Galla, T., and Farmer, J. D. (2016) "The prevalence of complex dynamics in games with many players," *in preparation.*

Shapley, L. S. (1964) "Some topics in two-person games," *Advances in game theory, Annals of Mathematical Studies*, Vol. 52, pp. 1–29.

Stahl, D. O. (1988) "On the instability of mixed-strategy Nash equilibria," *Journal of Economic Behavior & Organization*, Vol. 9, pp. 59–69.

——— (1996) "Boundedly rational rule learning in a guessing game," *Games and Economic Behavior*, Vol. 16, pp. 303–330.

Vilone, D., Robledo, A., and Sánchez, A. (2011) "Chaos and unpredictability in evolutionary dynamics in discrete time," *Physical review letters*, Vol. 107, p. 038101.

Young, H. P. (1993) "The evolution of conventions," *Econometrica: Journal of the Econometric Society*, Vol. 61, pp. 57–84.

A Payoff parameters and classes of games

To prove Proposition 1 for the generic 2×2 payoff matrix (1), with 8 degrees of freedom, is cumbersome. However, we show that there are only 4 effective degrees of freedom in the payoff matrix, for what concerns the NE and the dynamical properties of EWA learning.

Consider the payoff matrix (1); then consider

$$\Pi' = \begin{pmatrix} H, L & 0, 0 \\ 0, 0 & K, M \end{pmatrix}, \tag{22}$$

where $H = a - c$, $K = d - b$, $L = e - g$ and $M = h - f$.

Finally, consider

$$\Pi'' = \begin{pmatrix} 0, 0 & 0, L' \\ H', 0 & K', M' \end{pmatrix}, \tag{23}$$

where $H' = -H$, $L' = -L$, $K' = K$ and $M' = M$.

Proposition A.1. *(i) The payoff matrices* Π *and* Π'*, defined by* (1) *and* (22) *respectively, have the same pure and mixed strategy NE.*

(ii) The EWA dynamics (6) *is identical in the two cases, and so is any learning dynamics where the propensities are mapped to the probabilities using a logit function and the expected payoff enters as an additive term in the update of the propensities.*

(iii) Any other payoff matrix Π'' *where the elements* H'*,* K'*,* L' *and* M' *are either identical or opposite to* H*,* K*,* L *and* M*, and are in the same position if identical and on the opposite position if opposite (up rather than down for Row, left rather than right for Column) is equivalent to* Π *and* Π'*. An example of such payoff matrix is* Π''*, defined in* (23)*.*

Therefore, we set the off-diagonal elements to zero and prove Proposition 1 using payoff matrix (22). We then prove Proposition A.1.

Proof of proposition 1. In terms of the payoff matrix (22), the parameters A, B, C and D are defined as

$$\begin{aligned} A &= \frac{1}{4}(H + K), \\ B &= \frac{1}{4}(H - K), \\ C &= \frac{1}{4}(L + M), \\ D &= \frac{1}{4}(L - M). \end{aligned} \tag{24}$$

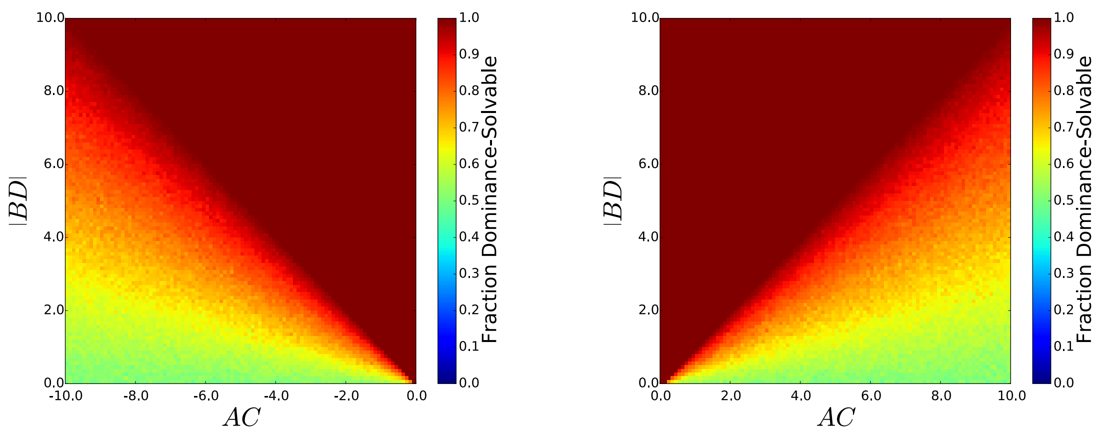
As we are interested in their relative magnitudes, we drop the 1/4 multiplicative factors and write coordination and dominance respectively as

$$\begin{aligned} AC &= (H + K)(L + M), \\ |BD| &= |(H - K)(L - M)|. \end{aligned} \tag{25}$$

We start proving *(i)*. The game is symmetric, so $H = L$ and $K = M$. So $|A| = |H + K|$ and $|B| = |H - K|$. Moreover, the conditions $H, K > 0$ or $H, K < 0$ respectively describe coordination and anticoordination games, whereas if either H or K are negative the game is dominance-solvable. So, if H and K have the same sign, the payoff matrix describes a coordination game and the sum of H and K (in absolute value) is larger than their difference, so that coordination is larger than dominance; if the signs of H and K are different, the game is dominance-solvable and the difference between H and K is larger (in absolute value) than their sum: dominance is larger than coordination.

We then consider *(ii)*. If ($|BD| > |AC|$), either $|B| > |A|$, or $|D| > |C|$, or both. Therefore, either H and K do not have the same sign, or L and M do not have the same sign, or both. All of these cases represent dominance-solvable games (the profile of pure strategies which is the NE of the game depends on the relative signs). On the contrary, the condition $|BD| < |AC|$ does not necessarily imply that both $|B| < |A|$ and $|D| < |C|$.[13] However, if that is the case, the sums of $H + K$ and $L + M$ are larger than the differences $H - K$ and $L - M$, which means that H, K and L, M have the same sign. If $AC > 0$, also A and C have the same sign, so either H, K, L, M are all positive, or they are all negative. If $H, K, L, M > 0$, the payoff matrix describes a coordination game; if $H, K, L, M < 0$, the payoff matrix describes an anticoordination game. If $AC < 0$, A and C have different signs. Suppose without loss of generality that $A > 0$, $C < 0$. Then $H, K > 0$, $L, M < 0$. The payoff matrix represents a discoordination game.

We still have to show that, the larger the value of coordination (compared to dominance), the more likely the payoff matrix describes a coordination or anticoordination game (rather than a dominance-solvable game). This is not obvious. Coordination may be large because $A \gg B$, but it could still be that $C \lesssim D$. An extreme example is that $B = 0$ (so dominance is null), whereas $A, C \neq 0$: it is always $|AC| > |BD| = 0$, but this condition imposes no restrictions on whether $|C| > |D|$ or $|C| < |D|$. The intuition is, if we randomly choose the payoff elements, it is not likely to generate such a specific payoff matrix. We verify this conjecture by running extensive numerical simulations. For each (AC,$|BD|$) point, we generate 1000 random realizations of the payoff matrix with specified AC and $|BD|$; we then compute the fraction of dominance-solvable games (the other fractions are coordination or discoordination games, according to whether we are in the positive or negative AC semiplane). The results are in Figure A.1. As expected, if $|BD| > AC$, all games are dominance-solvable. Viceversa, the larger the absolute value of AC, the more likely the payoff matrix may represent (anti)coordination or discoordination games. Interestingly, the fraction of dominance-solvable games never drops to zero. Finally, notice the consistency between Figure A.1 and Figure 1 (net of the fact there is not a neat separation between the dominance-solvable and the (anti)coordination and discoordination regions).



(a) (b)

Figure A.1: Fraction of dominance-solvable games for randomly generated payoff matrices, as a function of the coordination (AC) and dominance ($|BD|$) parameters. The larger is coordination compared to dominance, the more likely the payoff matrix describes a coordination (if $AC > 0$) or discoordination (if $AC < 0$) game.

[13] For instance, consider the payoff matrix (1), with $a = 3$, $e = 1$, $d = -1$, $h = 2$, $b = c = f = g = 0$: this is a dominance-solvable game, but $|AC| = 3/8 > |BD| = 2/8$. Note that $|D| = 1/4 < |C| = 3/4$, but $|B| = 1 > |A| = 1/2$

□

Proof of proposition A.1. We start proving *(i)*. The pure strategy NE are only determined by the ordinal properties of the payoffs. Consider player Row. Her contribution in determining the pure strategy Nash Equilibria depends on whether $a > c$ or $d > b$, so it is unchanged if we consider $H = a - c > 0$ and $K = d - b > 0$. The same argument applies to player Column: his contribution in determining the pure strategy Nash Equilibria depends on whether $e > g$ or $h > f$, so it is unchanged if we consider $L = e - g > 0$ and $M = h - f > 0$. The same is true for all other positive/negative combinations.

In the 2×2 game described in (1), the mixed strategies NE $(p, 1-p)$, $(q, 1-q)$ for players Row and Column respectively are given by

$$\begin{aligned} p &= \frac{h-f}{e-g+h-f}, \\ q &= \frac{d-b}{a-c+d-b}. \end{aligned} \tag{26}$$

Again, we can rewrite the above equations without loss of generality in terms of H, K, L and M, namely

$$\begin{aligned} q &= \frac{K}{H+K}, \\ p &= \frac{M}{L+M}. \end{aligned} \tag{27}$$

We consider *(ii)*. We only focus on player Row (the proof for Column is identical). If, at time t, Column plays a mixed strategy given by $(y(t), 1-y(t))$, the expected payoff for Row for playing pure strategy 1 is $\overline{\Pi_1^R}(y(t)) = ay(t) + b(1-y(t))$ and the expected payoff for strategy 2 is $\overline{\Pi_2^R}(y(t)) = cy(t) + d(1-y(t))$. Now, the ratio $\frac{x(t+1)}{1-x(t+1)}$ fully determines $x(t+1)$. Using (6) we find

$$\frac{x(t+1)}{1-x(t+1)} \propto e^{\beta(1-(1-\alpha)(1-\kappa))\left(\overline{\Pi_1^R}(y(t)) - \overline{\Pi_2^R}(y(t))\right)}, \tag{28}$$

where $\overline{\Pi_1^R}(y(t)) - \overline{\Pi_2^R}(y(t)) = (a-c)y(t) + (d-b)(1-y(t)) = Hy(t) + Ky(t)$. Note that the same argument applies for any other learning algorithm where the expected payoffs are in the argument of an exponential and can be separated from the past propensities (e.g. do not enter multiplicatively).

Finally, *(iii)* follows simply from the above results. If we consider $H' = -H$ at the bottom left in the payoff matrix, it is $H' = c - a$, so $a > c$ implies that $H' < 0$ (the payoff element at the top left is still larger than that at the bottom left), the mixed strategies are identical upon the transformation $H \leftrightarrow -H'$ and the learning dynamics depends on H as well, as $\overline{\Pi_1^R}(y(t)) - \overline{\Pi_2^R}(y(t)) \propto (0 - H')y(t) = Hy(t)$. □

Apart from the above properties, we stress that the transformed payoff matrix (22) is not fully equivalent to (1). For instance, consider the following Prisoner Dilemma (PD):

$$\Pi_{PD} = \begin{pmatrix} 2,2 & 0,3 \\ 3,0 & 1,1 \end{pmatrix}, \tag{29}$$

where strategy 1 is Cooperate and strategy 2 is Defect. The transformed payoff matrix is

$$\Pi'_{PD} = \begin{pmatrix} -1,-1 & 0,0 \\ 0,0 & 1,1 \end{pmatrix}. \tag{30}$$

The payoff matrices (30) and (29) are not equivalent, in that the property that the NE and the Pareto Equilibrium do not coincide is lost, and so is the dilemma between cooperation and defection.

In a similar manner, consider the Stag-Hunt (SH) game:

$$\Pi_{SH} = \begin{pmatrix} 2,2 & 0,1 \\ 1,0 & 1,1 \end{pmatrix}, \tag{31}$$

where strategy 1 is Stag (S) and strategy 2 is Hare (H). Here (S,S) is the payoff-dominant NE, while (H,H) is the risk-dominant NE. If we apply the transformation we find

$$\Pi'_{SH} = \begin{pmatrix} 1,1 & 0,0 \\ 0,0 & 1,1 \end{pmatrix}. \tag{32}$$

The above is a pure coordination game, and the properties of payoff and risk-dominance no longer hold.

However, note that both in (29) and (30) and in (31) and (32) the NE are the same and so are all differences in payoffs, holding the strategy of the other player fixed. A learning algorithm that bases its learning properties on the performance of one pure strategy compared to the other one, should be invariant under the payoff matrix transformation which we described: this is probably the most intuitive explanation of why Proposition 1 holds.

B Proof of Proposition 2

We first consider assertion *(i)*. Since $B = 0$, there is always a fixed point $\tilde{x}^\star = 0$. It is stable if (from Eq. (13))

$$\frac{\beta}{\alpha}|A| \leq 1. \tag{33}$$

So, as long as $\tilde{x}^\star = 0$ is the unique fixed point, it is stable.

We then consider assertion *(ii)*, and in particular the lower bound, $\frac{\beta}{\alpha}|A| \to 1^+$. There are two fixed points $\tilde{x}^\star = \pm\epsilon$, where ϵ is an arbitrarily small number. Thanks to the symmetry of the game, we focus on a profile of mixed strategies given by $(\tilde{x}^\star, \tilde{x}^\star)$. To second order, $\cosh\tilde{x}^\star \approx 1 + (\tilde{x}^\star)^2/2$. The stability condition becomes

$$\frac{\alpha}{\beta}\left(1 + \frac{(\tilde{x}^\star)^2}{2}\right)\left(1 + \frac{(\tilde{x}^\star)^2}{2}\right) - |A| \geq 0, \tag{34}$$

i.e.

$$(\tilde{x}^\star)^2 \geq \frac{\beta}{\alpha}|A| - 1. \tag{35}$$

Now, we Taylor expand $\Psi(\tilde{x}^\star)$ (defined in Section 4.1.1) to third order (first order would just yield $\tilde{x}^\star = 0$) and solve $\tilde{x}^\star = \Psi(\tilde{x}^\star)$. Apart from the null solution, we get

$$(\tilde{x}^\star)^2 = \frac{3\left(\frac{\beta^2 A^2}{\alpha^2} - 1\right)}{\frac{\beta^2 A^2}{\alpha^2}\left(1 + \frac{\beta^2 A^2}{\alpha^2}\right)}. \tag{36}$$

It is easily checked that for $\frac{\beta}{\alpha}|A| \to 1^+$, the condition (35) is satisfied: the fixed points whose components are the "lateral solutions" are stable. Therefore, there is a supercritical pitchfork bifurcation at the value $\frac{\beta}{\alpha}|A| = 1$.

The upped bound, namely $\frac{\beta}{\alpha}|A| \to \infty$, is easily dealt with. Indeed, because we are searching for the intersection with the $\tilde{x}^\star$ line, the fixed point is approximately the height of the hyperbolic tangent itself: $\tilde{x}^\star \approx \pm\frac{\beta}{\alpha}|A|$. Now, for $\frac{\beta}{\alpha}|A| \to \infty$ the hyperbolic cosine can be approximated by

$$\cosh\left(\frac{\beta}{\alpha}|A|\right) \approx \exp\left(\frac{\beta}{\alpha}|A|\right)/2. \tag{37}$$

We can rewrite the stability condition as:

$$\frac{4\beta}{\alpha}|A|\exp\left(-2\frac{\beta}{\alpha}|A|\right)\leq 1. \tag{38}$$

For $\frac{\beta}{\alpha}|A|\to\infty$, the LHS of the above equation goes to zero, so the inequality obviously holds.

Finally, the proof of *(iii)* is identical to the proof of the upper bound for $\frac{\beta}{\alpha}|A|$, in that the same arguments apply to sufficiently large values of $\frac{\beta}{\alpha}|B|$, for which the only fixed point will be far enough from zero to be stable.

C Proof of Proposition 3

In order to study the properties of the pure strategy NE we need to consider the learning dynamics in the original coordinates (the pure strategies map into infinite elements in the transformed coordinates). The EWA dynamics reads (using (6) and the payoff matrix (1)):

$$\begin{aligned} x(t+1) &= \frac{x(t)^{1-\alpha}e^{\beta(ay(t)+b(1-y(t))}}{x(t)^{1-\alpha}e^{\beta(ay(t)+b(1-y(t))}+(1-x(t))^{1-\alpha}e^{\beta(cy(t)+d(1-y(t))}}, \\ y(t+1) &= \frac{y(t)^{1-\alpha}e^{\beta(ex(t)+f(1-x(t))}}{y(t)^{1-\alpha}e^{\beta(ex(t)+f(1-x(t))}+(1-y(t))^{1-\alpha}e^{\beta(gx(t)+h(1-x(t))}}. \end{aligned} \tag{39}$$

From Eq. (39) we can see that the pure strategies $(x,y)\in\{(0,0),(0,1),(1,0),(1,1)\}$ are all fixed points of the dynamics. Let us study their stability properties. We get a Jacobian

$$J=\begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix}, \tag{40}$$

with

$$\begin{aligned} J_{11} &= \frac{(1-\alpha)(x-x^2)^{\alpha}e^{\beta(y(a-b-c+d)+b-d)}}{\left(x(1-x)^{\alpha}e^{\beta(y(a-b-c+d)+b-d)}-(x-1)x^{\alpha}\right)^2}, \\ J_{12} &= \frac{\beta(x-x^2)^{\alpha+1}(a-b-c+d)e^{\beta(y(a-b-c+d)+b-d)}}{\left(x(1-x)^{\alpha}e^{\beta(y(a-b-c+d)+b-d)}-(x-1)x^{\alpha}\right)^2}, \\ J_{21} &= \frac{\beta(y-y^2)^{\alpha+1}(e-f-g+h)e^{\beta(x(e-f-g+h)+f-h)}}{\left(y(1-y)^{\alpha}e^{\beta(x(e-f-g+h)+f-h)}-(y-1)y^{\alpha}\right)^2}, \\ J_{22} &= \frac{(1-\alpha)(y-y^2)^{\alpha}e^{\beta(x(e-f-g+h)+f-h)}}{\left(y(1-y)^{\alpha}e^{\beta(x(e-f-g+h)+f-h)}-(y-1)y^{\alpha}\right)^2}. \end{aligned} \tag{41}$$

As it can be seen by taking the appropriate limits in Eqs. (41), for all pure strategies the Jacobian has infinite elements along the main diagonal and null elements along the antidiagonal. This means that the NE in pure strategies are infinitely unstable, and may be the reason for the extreme nonlinearities observed in Galla and Farmer (2013) near the faces of the simplex.

The only case where the elements of the Jacobian for the NE in pure strategies would not be infinite is that of no memory loss, $\alpha=0$, as it is possible to see if one computes the eigenvalues with this parameter restriction. In fact, the NE in pure strategies become stable fixed points of the learning dynamics.

D Effect of stochasticity on learning

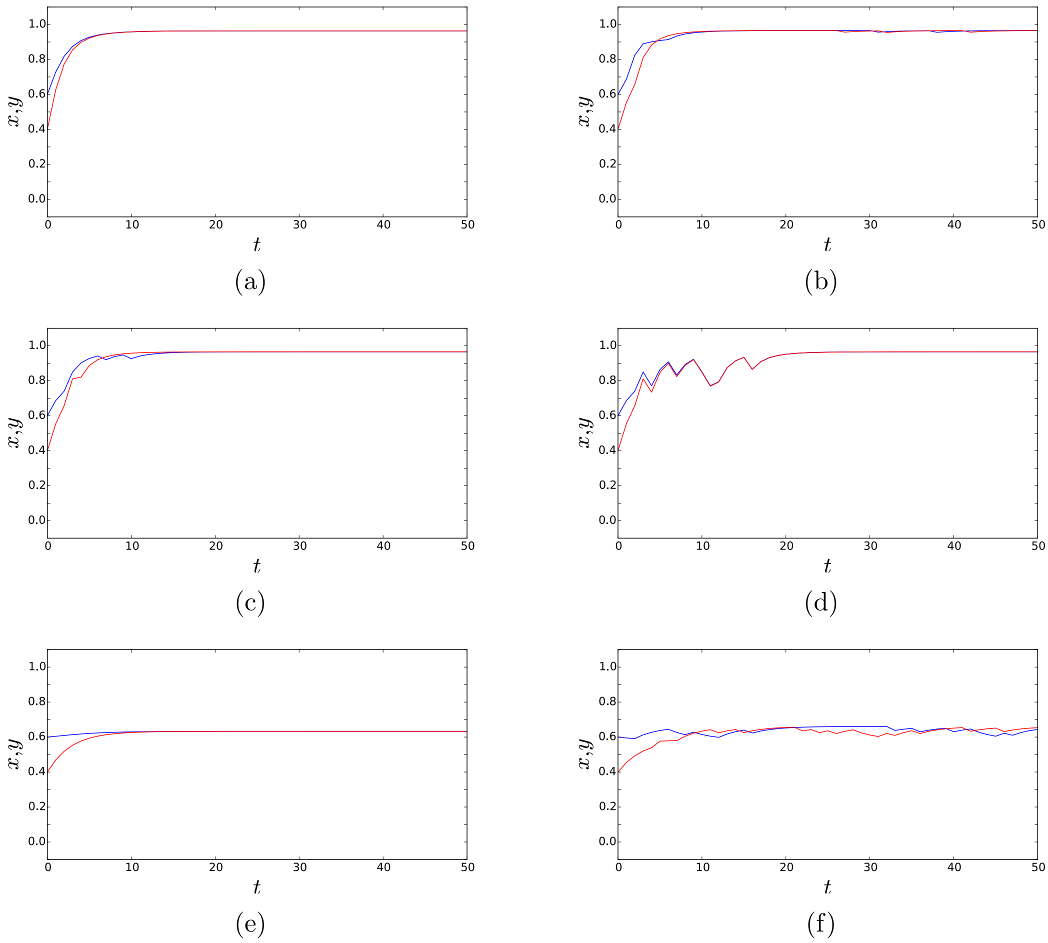


Figure D.1: Time series of the probabilities x (in blue) and y (in red). Values of the parameters: $\alpha = 0.3$, $b = c = f = g = 0$, $a = e = 2$, $d = h = -1$, $\beta = 0.5$ in (a)-(d), $\beta = 0.1$ in (e)-(f). (a) Deterministic learning; (b) Stochastic learning with $T = 2$; (c) Stochastic learning with $T = 1$ and $\delta = 1$; (d) Stochastic learning with $T = 1$ and $\delta = 0$; (e) Deterministic learning; (f) Stochastic learning with $T = 1$ and $\delta = 1$. Deterministic and stochastic learning are largely similar. See Section 5.1 for further comments.

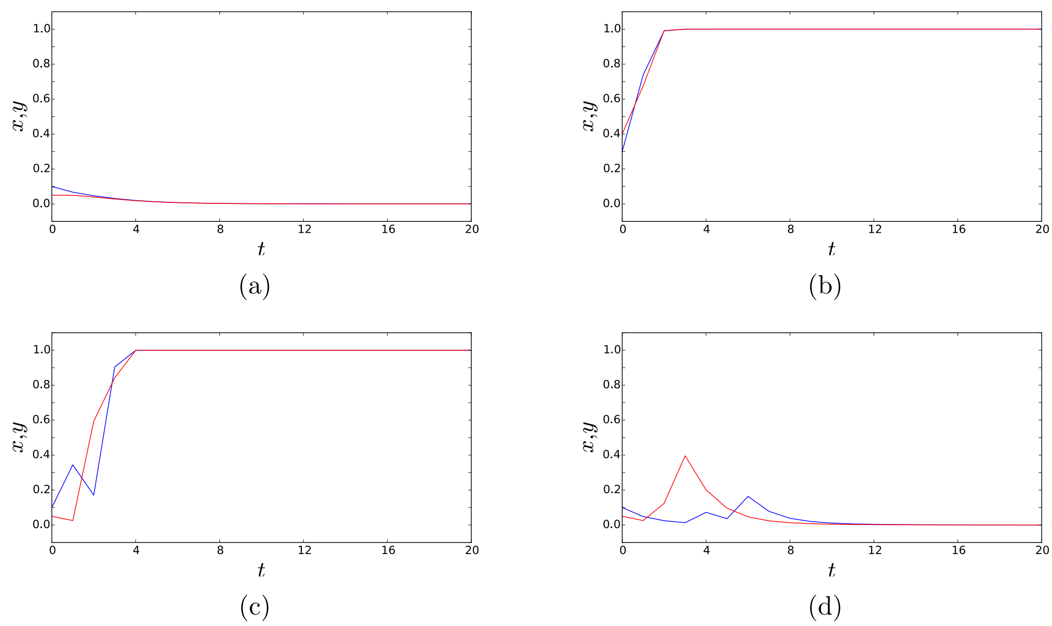


Figure D.2: Time series of the probabilities x (in blue) and y (in red). Values of the parameters: $\alpha = 0.1$, $\beta = 1$, $b = c = f = g = 0$, $a = e = 6$, $d = h = 1$. (a) Deterministic learning starting from the initial conditions $x(0) = 0.1$ and $y(0) = 0.05$, close to the Pareto-dominated NE; (b) Deterministic learning starting from the initial conditions $x(0) = 0.3$ and $y(0) = 0.4$, farther to the Pareto-dominated NE; (c)-(d) Single realizations of stochastic learning, starting from the same initial condition as (a). Noise can help reaching the most efficient fixed point only in the early stage of the dynamics. See Section 5.1 for further comments.

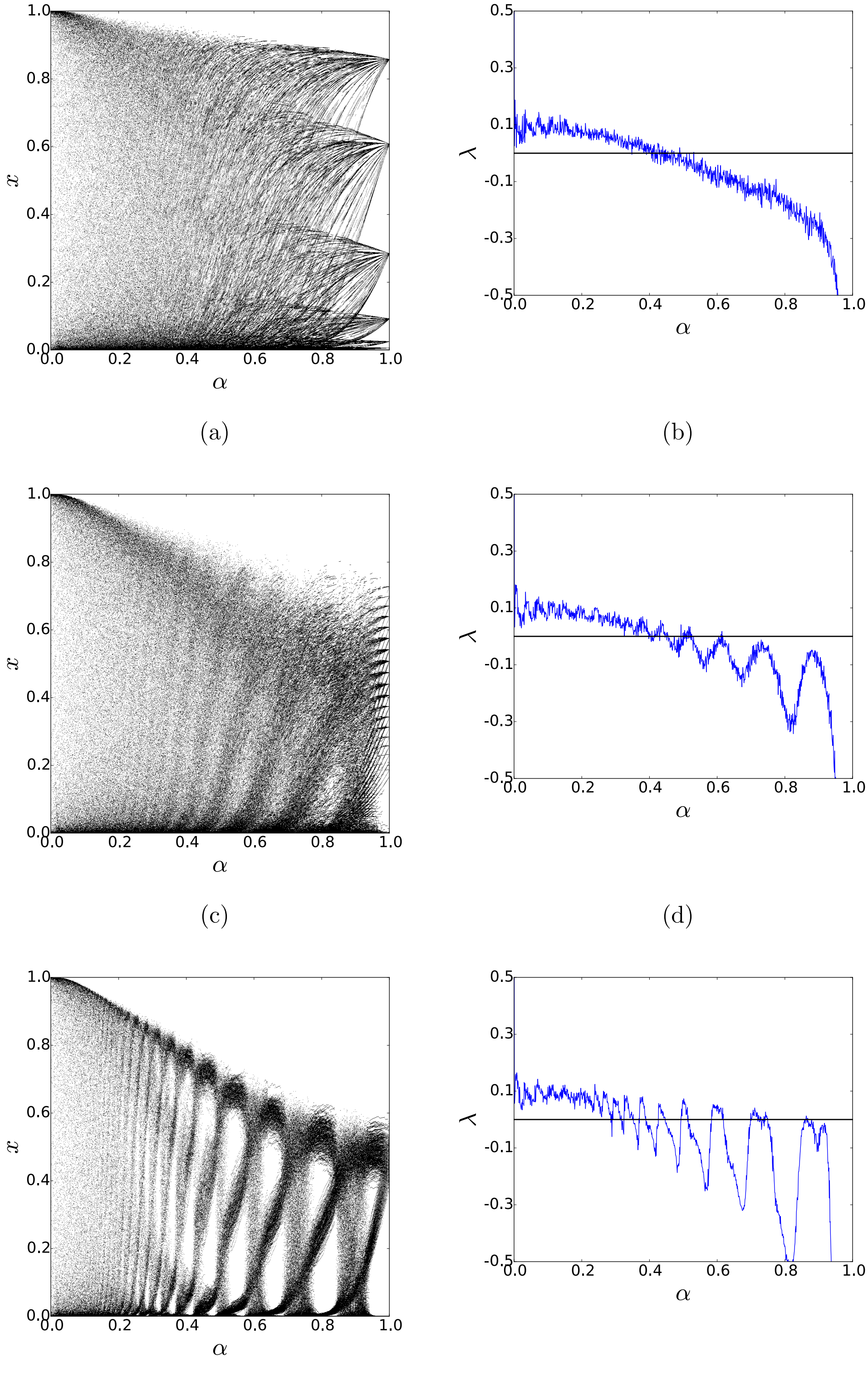


Figure D.3: Bifurcation diagram and largest Lyapunov exponent as a function of α for several values of T. The payoff parameters are $b = c = f = g = 0$, $a = -11.8$, $d = -1.8$, $e = 11.8$ and $h = 1.8$ and the intensity of choice is $\beta = 1$. Panels (a)-(b): $T = 10$; Panels (c)-(d): $T = 100$; Panels (e)-(f): $T = 1000$. The stronger the level of noise, the more blurred are the dynamical properties. See Section 5.1 for further comments.