Poincaré-Bendixon Limit Sets in Multi-Agent Learning

Aleksander Czechowski
Delft University of Technology
The Netherlands
a.t.czechowski@tudelft.nl

Georgios Piliouras
Singapore University of Technology and Design
Singapore
georgios@sutd.edu.sg

ABSTRACT
A key challenge of evolutionary game theory and multi-agent learning is to characterize the limit behavior of game dynamics. Whereas convergence is often a property of learning algorithms in games satisfying a particular reward structure (e.g., zero-sum games), even basic learning models, such as the replicator dynamics, are not guaranteed to converge for general payoffs. Worse yet, chaotic behavior is possible even in rather simple games, such as variants of the Rock-Paper-Scissors game. Although chaotic behavior in learning dynamics can be precluded by the celebrated Poincaré-Bendixson theorem, it is only applicable to low-dimensional settings. Are there other characteristics of a game that can force regularity in the limit sets of learning? We show that behavior consistent with the Poincaré-Bendixson theorem (limit cycles, but no chaotic attractor) can follow purely from the topological structure of the interaction graph, even for high-dimensional settings with an arbitrary number of players and arbitrary payoff matrices. We prove our result for a wide class of follow-the-regularized leader (FoReL) dynamics, which generalize replicator dynamics, for binary games characterized by interaction graphs where the payoffs of each player are only affected by one other player (i.e., interaction graphs of indegree one). Since chaos occurs already in games with only two players and three strategies, this class of non-chaotic games may be considered maximal. Moreover, we provide simple conditions under which such behavior translates into efficiency guarantees, implying that FoReL learning achieves time-averaged sum of payoffs at least as good as that of a Nash equilibrium, thereby connecting the topology of the dynamics to social-welfare analysis.

KEYWORDS
Replicator Dynamics; Follow-the-Regularized Leader; Polymatrix Games; Poincaré-Bendixson Theorem; Regret Minimization

1 INTRODUCTION
Dynamical systems and evolutionary game theory have been instrumental in modern research on multi-agent learning [8, 11, 20, 46, 52, 53, 55]. In particular, characterizing the convergence and limit sets of learning trajectories is vital for understanding the long-term behavior of multi-agent systems. However, even in simple games, such as Rock-Paper-Scissors [44, 51], models of evolution and learning are not guaranteed to converge; even beyond cycles, long-term behavior may lead to chaotic behavior, known to the dynamical systems community from, e.g., weather models [35]. Not only does chaos manifest itself even in simple games with two players, but moreover, a string of recent results suggests that such chaotic, unpredictable behavior may indeed be the norm across a variety of simple low-dimensional game dynamics [4–6, 12, 15, 17, 19, 42, 49, 56]. Importantly, these results are persistent even for the well-known class of Follow-the-Regularized-leader (FoReL) dynamics [13, 38], despite the fact that FoReL dynamics include some of the most widely studied learning dynamics such as replicator dynamics [26, 54], which is the continuous-time analogue of the Multiplicative Weights Update meta-algorithm [3], well known for its optimal regret properties. Finally, the emergence of chaotic behavior has been connected with increased social inefficiency, which shows that chaotic dynamics can lead to highly inefficient outcomes [14, 47]. Such profoundly negative results raise the following questions:

• Do simple, robust conditions exist under which learning behaves well?
• Which types of games lie at the "edge of chaos"?
• Does dynamic simplicity translate to high-efficiency and social welfare?

Traditionally, a lot of work has focused on showing that, in specific classes of games (e.g., zero-sum or potential games), learning dynamics can lead to convergence and equilibration, see [11, 18, 50, 59] and references therein. Few results span over to general sum games and games of arbitrary payoff structures; however, such general approaches are arguably essential in modern research on multi-agent learning. For instance, unstructured payoffs can occur naturally when stochastic extensive form games are used to create empirical normal form games, by averaging payoffs from simulations for combinations of strategies [34, 39, 57]. Unstructured payoffs also arise in many real-world applications, such as, e.g., modeling the impact of investing strategies of large funds on the stock market. While equilibration may not always be possible in such cases, one can still wish to ensure a regularity of sorts in the learning outcomes of the multi-agent system. In particular, the famous Poincaré–Bendixson theorem (Theorem 1) ensures that two-dimensional continuous learning and adaptation dynamics never form truly chaotic outcomes. However, this comes at a cost: although no specific payoff structure is needed, the underlying learning dynamics must be at most two dimensional.

Our approach and results. Rather than by making assumptions on the reward structure or on the dimensionality, we explore a different type of constraint in games. We show that the limit behavior of
learning can be determined solely by the topological-combinatorial structure of the game, regardless of the number of players, or algebraic correlations between the payoffs (e.g., zero-sum). Firstly, we restrict ourselves to binary games [9, 37, 60], where players have two strategies. Secondly, we assume that every player can be affected by the behavior of up to one other player. Finally, we add a technical restriction that the game is connected, meaning that it cannot be decomposed into two subgames that are completely independent of each other. Such games encompass, among others, all $2 \times 2$ games [22], Jordan’s game [21, 24, 28], and easily identifiable subclasses of real-world systems where the graph structure is evident, such as certain traffic networks [1, 33], supply chains [10], or problems of water allocation in deltas [2, 30]. Under these assumptions, we prove in Section 3 our main contribution in the form of Theorems 3 and 4, which say that the limit behavior of FoReL learning of these games is always consistent with the Poincaré–Bendixson theorem.

Having excluded the presence of chaos, we further analyze quantitative properties of binary games, which admit cyclic interaction graphs. In Section 4 we show that, under additional but structurally robust assumptions on the payoff matrices (i.e., assumptions that remain valid after small perturbations of the payoff matrices and so are suitable, for example, for empirical payoff matrices), one can derive positive results about the efficiency of the time-averaged behavior of the dynamics regardless of whether they are convergent. As is typically the case in the price of anarchy (PoA) literature [32], we focus on the measure of social welfare, which is the sum of individual payoffs. Whereas the typical PoA literature argues that regret-minimizing dynamics (such as FoReL) are at most a constant factor worse than the behavior of the worst-case Nash equilibrium [47, 48], we instead show that FoReL dynamics are always at least as efficient as the worst-case Nash equilibrium. Finally, Section 5 provides examples of games satisfying our assumptions and their possible limit behavior, as well as a counterexample in the form of a simple binary game that breaks our assumptions and induces chaotic learning dynamics.

Related work. First of all, we consider several papers containing complementary results in the form of examples of simple FoReL systems with chaotic dynamics. In addition to the papers mentioned in the introduction, we highlight a chaotic example of Sato et al. [51] that involves a two-player, three-action game and two complex/chaotic examples in three-player binary games without structured interactions [43, 45]. Comparing these with our assumptions (i.e., binary games and previous-neighbor interactions), we see that our results establish a maximal class of games for which such regularity results on limit sets are possible.

Research that considers non-convergence but focuses on non-chaoticity is scarce. In the closest works to ours, [16, 40, 41], the authors leverage the Poincaré–Bendixson theorem to show that the limit behavior of bounded learning trajectories in certain learning systems can be either convergent or cyclic, and in particular no chaotic attractor is possible. However, they do so by assuming low dimensionality (three-player limit) or a nongeneric structure on the set of allowable games, which allows for dimensionality reduction (i.e., a network of $2 \times 2$ zero-sum, or coordination games). In terms of connections between cyclic behavior and the efficiency of learning dynamics, [31] shows that, for a class of three players, two strategy games with a cyclic attractor can result in social welfare (sum of payoffs) that can be better than the Nash equilibrium payoff; however, the result is once again constrained to the exact game theoretic model.

2 PRELIMINARIES

2.1 Normal form games

A finite game in normal form consists of a set of $N$ players, each with a finite set of strategies $A_i$. The preferences of each player are represented by the payoff function $u_i : \prod_i A_i \to \mathbb{R}$. To model the behavior at scale or probabilistic strategy choices, one assumes that players use mixed strategies, namely, probability distributions $(x_{i\alpha})_{\alpha \in A_i} \in \Delta(A_i) =: X_i$. With a slight abuse of notation, the expected payoff of player $i$ in the profile $(x_{i\alpha})_{\alpha \in A_i}$ is denoted $u_i$ and given by

$$u_i(x) = \Sigma_{\alpha \in A_i \cdots A_N \in A_N} u_i(\alpha_1, \ldots, \alpha_N) x_{1\alpha_1} \cdots x_{N\alpha_N}. \tag{1}$$

A mixed strategy $\hat{x}$ is a Nash equilibrium iff $i$ and $\forall x : x_j = \hat{x}_j$, $j \neq i$ we have $u_i(x) \leq u_i(\hat{x})$. In other words, no player can unilaterally increase their payoff by changing their strategy distribution. The minimax value for player $i$ is given by $\min_{x_i} \max_{x_{-i}} u_i(x)$, where $x_{-i} := (x_j)_{j \neq i}$. This is the smallest possible value that player $i$ can be forced to attain by other players, without them knowing the strategy of player $i$. We call a game binary iff $|A_i| = 2$ for all $i$.

2.2 Graphical polymatrix games

To model the topology of interactions between players, we restrict our attention to a subset of normal form games, where the structure of interactions between players can be encoded by a graph of two-player normal form subgames, leading us to consider so-called graphical polymatrix games (GPGs) [27, 29, 58]. A simple directed graph is a pair $(V, E)$, where $V = \{1, \ldots, N\}$ is a finite set of vertices (representing the players), and $E$ is a set of ordered vertex pairs (edges), where the first element is called the predecessor, and the second is called the successor. Each edge $(i, k)$ has an associated two-player normal form game, where only the successor $k$ is assigned payoffs. These are represented by a matrix $A^{i,k}$ with rows enumerating the strategies of player $k$, and columns enumerating the strategies of player $i$. For a given strategy profile $s = (s_i)_{i \in V} \in \prod_i A_i$, the payoffs for player $k$ in the full game are then determined as the sum

$$u_k(s) = \sum_{i \in (k) \in E} A^{i,k}(s_i, s_k). \tag{2}$$

The payoffs can be extended to mixed strategies in a standard multilinear fashion:

$$u_k(x) = \sum_{i \in (k) \in E} \sum_{x_i, x_k} A^{i,k}(s_i, s_k) x_i x_k. \tag{3}$$

A situation where both the successor $k$ and the predecessor $i$ obtain a reward can be modeled by including both edges $(i, k)$ and $(k, i)$ in the graph.

We say that a simple directed graph is weakly connected if any two vertices can be connected by a set of edges, where the direction of the edges is not considered. This is a weaker condition than
strong connectedness, where each pair of vertices must be connected by a path (i.e., a sequence of edges together with associated vertices, where the successor in one edge is the predecessor in the next). The indegree of a vertex is the number of edges for which the vertex is the successor (i.e., the number of predecessors). The outdegree is the number of edges for which the vertex is the predecessor (i.e., the number of successors). A cycle is a path where the predecessor in the first edge is the successor in the last edge. For our exposition we identify cycles modulo shifts, i.e., if two paths consist of the same edges in shifted order, then they form the same cycle. In this paper we consider two types of weakly connected GPGs:

(1) First, cyclic games, where the interaction between the players forms a cycle, where each player interacts only with the previous neighbor. We observe that in such a cyclic game the indegree and outdegree of each vertex is one. For simplicity, we label the nodes of such N-player games by natural numbers \( i = 0, 1, \ldots, N \) and use the convention that node \( i \) is the successor to node \( i - 1 \) and that node 0 is identified with node \( N \).

(2) Second, a more general class of graphical games, where each player’s payoffs depend on at most one other player (i.e., the indegree of each vertex is at most one). For a vertex \( i \in \mathcal{V} \), we denote the predecessor vertex by \( i \), if it exists. For cyclic games we have \( i = i - 1 \).

Below, we state and prove a simple lemma that characterizes the one-predecessor assumption in terms of graph topology and clarifies the relation between cyclic and indegree-one graphs (cf. Figure 1).

Lemma 1. Let \((\mathcal{V}, \mathcal{E})\) be a weakly connected, simple, directed graph. If the indegree of each vertex is at most one, then the graph can have at most one cycle. If the graph has no cycle, then it has at most one root vertex (i.e., a vertex of indegree zero), such that all other vertices are connected to it by a unique directed path.

Proof. For the first part of the lemma, we assume the contrary: that \( a_1, a_2 \) are nodes of two distinct cycles within the same weakly connected component. The edges between \( a_1 \) and \( a_2 \) must form a path (otherwise there would be a vertex with two predecessors). Assume the path leads from \( a_1 \) to \( a_2 \) and let \( a_0 \) be the first vertex which is both on the path and on the cycle of \( a_2 \). Then \( a_0 \) has two predecessors, which leads to a contradiction.

For the second part of the lemma we argue as follows. If any vertex has a sequence of predecessors that does not form a cycle, and does not have a root node, then by backtracking through the predecessors we could identify an infinite collection of distinct vertices. Therefore, there must be at least one root node for each vertex. The path from such a root node to the given vertex must be unique, otherwise one could identify a vertex along the path with two predecessors. Finally, it is impossible to have two distinct root nodes, as connectedness imposes that there would have to exist a node with two predecessors between them.

Remark 1. Under the assumptions of Lemma 1, if the graph has a cycle, then the cycle enjoys properties similar to those of a root node: no paths go from outside the cycle to the cycle (otherwise one vertex in the cycle would have two predecessors), and all vertices outside the cycle must be connected by a path from one of the vertices of the cycle (a unique path, up to the starting point within the cycle). Later, we shall refer to such cycle as the root cycle.

2.3 Follow-the-regularized-leader equations

Denote by \( \psi_{ia}(x):= u_i(a_i; x_{-i}) \) and \( \psi_i(x) = (\psi_{ia}(x))_{a_i \in A_i} \). To model the dynamics of learning we use a class of learning systems known as follow-the-regularized-leader systems (FoRel) [11, 52]. This class encompasses a variety of models ranging from gradient to replicator dynamics, and allows for natural description of learning as regularized maximization of individual payoffs.

FoRel dynamics for player \( i \) are defined by evolution of utilities \( y_i = (y_{ia})_{a_i \in A_i} \in \mathbb{R}^{|A_i|} \) that is real numbers representing a score each player assigns to each respective strategy – by the integral equation

\[
y_i(t) = y_i(0) + \int_0^t \psi_i(x(s))ds,
\]

where the choice map \( Q = (Q_1, \ldots, Q_N) \), \( Q_i : \mathbb{R}^{|A_i|} \rightarrow X_i \), which determines the evaluated strategy profile \( x(t) \) is given on each coordinate by:

\[
Q_i(y_i) = \text{argmax}_{x_i \in X_i} \langle y_i, x_i \rangle - h_i(x_i).
\]

In the above \( h_i : X_i \rightarrow \mathbb{R} \cup \{\infty\} \) is a convex regularizer function, representing a regularization/exploration term. The equation (4) represents how players adapt their mixed strategies to changing utility values. Observe, that without the regularization term, the map \( Q_i \) would simply put all weight on the strategy with the highest utility.

In binary games, each player has only two strategies at his disposal, say \( a_0, a_1 \). The variable \( x_i \) denotes then the proportion of time player \( i \) plays strategy \( a_0 \), and the proportion of \( a_1 \) is given by \( 1 - x_i \). Following [38], we introduce new variables \( z_i := y_{ia_0} - y_{ia_1} \in \mathbb{R} \), representing the difference in utilities between playing strategy \( a_0 \) and \( a_1 \). It is intuitively clear, and it was proved formally e.g. in [38] that \( Q_i(z_i + c, c) \) is constant in \( c \), and therefore, without loss of generality, we can set \( c := 0 \), and restrict our considerations to a \( z \)-dependent choice map \( \hat{Q}_i(z_i) := Q_i(z_i, 0) \). Provided that \( Q_i \) is sufficiently regular (e.g. continuous), the integral equation (4) can be converted to a system of differential equations

\[
\dot{z} = V(z),
\]
A necessary condition for maximum to be attained in \( h \) and yields the following equations for a binary GPG with up to one node that influence the payoffs of \( i \). In particular, such regularizer satisfies the assumptions of Lemma 2, from steepness, continuity and strict convexity it follows that equation (10) has unique solution \( x(t) \) for all \( t \geq 0 \) and it is a connecting orbit between equilibria \( x_1 \) and \( x_2 \) (allowing \( x_1 = x_2 \)), if \( x(t) \to x_1 \) as \( t \to \infty \) and \( x(t) \to x_2 \) as \( t \to -\infty \). A set \( \omega(x_0) \subseteq \Omega \) is a limit set for an initial condition \( x_0 \in \Omega \), if \( \forall \delta > 0 \), there exists \( T \) such that \( x(t) \) is in the \( \delta \)-neighborhood of \( \omega(x_0) \) for all \( t \geq T \). From steepness and strict convexity it follows that equation (10) has a unique solution \( x_i := \hat{Q}_i (z_i) \) for any \( z_i \in \mathbb{R} \). From the inverse function theorem we have

\[
\frac{\partial x_i}{\partial z_i} = \hat{Q}_i'(z_i) = 1 / h_i''(x_i) > 0,
\]

which translates to the following system in original (x) coordinates:

\[
x_i = x_i(1 - x_i) \sum_{j,k \in \{0,1\}} (-1)^{(j+k)} A^{ij}(x_j, a_k) x_j - x_i(1 - x_i) \left( A^{ii}(x_i, a_i) + A^{0i}(a_i, a_0) \right), \quad i = 1, \ldots, N.
\]

2.4 Limit sets, periodic orbits and chaos

A differential equation \( \dot{x} = F(x) \) given by a \( C^1 \) vector field \( F : \Omega \to \mathbb{R}^n \) on a domain \( \Omega \subseteq \mathbb{R}^n \) admits a unique solution on a maximal open interval \( I = (I_l, I_r), \ I_l, I_r \in \mathbb{R} \cup \{\pm\infty\}, \) denoted by \( x(t) : I \to \mathbb{R}^n \), for any initial condition \( x(0) = x_0 \in \Omega \). Among possible solutions to such equation, we distinguish particular types of solutions defined by their qualitative properties: we say that a solution \( x(t) \) is an equilibrium iff \( x(t) = \text{const} \) for all \( t \in I \). A solution is periodic iff \( x(t) = x(t + T) \) for some \( T > 0 \) and all \( t \in I \), and it is a connecting orbit between equilibria \( x_1 \) and \( x_2 \) (allowing \( x_1 = x_2 \)), if \( x(t) \to x_1 \) as \( t \to \infty \) and \( x(t) \to x_2 \) as \( t \to -\infty \). A set \( \omega(x_0) \subseteq \Omega \) is a limit set for an initial condition \( x_0 \in \Omega \), if \( \forall \delta > 0 \), there exists \( T \) such that \( x(t) \) is in the \( \delta \)-neighborhood of \( \omega(x_0) \) for all \( t \geq T \). From steepness, continuity and strict convexity it follows that equation (10) has a unique solution \( x_i := \hat{Q}_i (z_i) \) for any \( z_i \in \mathbb{R} \). From the inverse function theorem we have

\[
\frac{\partial x_i}{\partial z_i} = \hat{Q}_i'(z_i) = 1 / h_i''(x_i) > 0,
\]

which translates to the following system in original (x) coordinates:

\[
x_i = x_i(1 - x_i) \sum_{j,k \in \{0,1\}} (-1)^{(j+k)} A^{ij}(x_j, a_k) x_j - x_i(1 - x_i) \left( A^{ii}(x_i, a_i) + A^{0i}(a_i, a_0) \right), \quad i = 1, \ldots, N.
\]
system of differential equations
\[ x_i = f_i(x_{i-1}, x_i), \quad i = 1, \ldots, n, \quad x \in \Omega, \] (15)
has the Poincaré-Bendixson property.

The above theorem is key to proving our further results.

3 THE POINCARÉ-BENDIXSON THEOREM FOR GAMES

In this section we state and prove our main results on the topology of limit sets in Follow-the-regularized-Leader learning. We will first state and prove the Poincaré-Bendixson theorem for cyclic games:

Theorem 3. Let \( \dot{z} = V(z) \) be a system of differential equations given by the vector field (7) — the follow-the-regularized-leader learning dynamics — for a binary, cyclic game. For any smooth, steep, strictly convex collection of regularizers \( \{h_i\}_i \) such system possesses the Poincaré-Bendixson property.

Proof. Since \( u_i \) depends only on \( Q_i \) and \( Q_{i-1} \), we have
\[
V_i(\hat{Q}(z)) = V_i(\hat{Q}_{i-1}(z_{i-1})) = \psi_{\text{int}}(Q_{i-1}(z_{i-1})), \quad i = 1, \ldots, n,
\] (16)
for all \( i \). We have:
\[
\frac{\partial V_i}{\partial z_{i-1}} \neq 0.
\] (17)
Hence, we can proceed with the proof of Theorem 4.

First, we state the following lemma on inheritance of the Poincaré Bendixson property for augmented systems.

Lemma 3. Consider the following \( y \)-augmented system of differential equations
\[
\dot{x} = f(x),
\]
\[
\dot{y} = g(x),
\]
\[ x = \{x_1, \ldots, x_n\} \in \mathbb{R}^n, \quad y \in \mathbb{R}. \] (21)
for smooth \( f, g \). If the original system
\[
\dot{x} = f(x)
\] (22)
has the Poincaré-Bendixson property, then the augmented system (21) also has the Poincaré-Bendixson property.

Proof. Let \( Z \) be an \( \omega \)-limit set corresponding to some solution \( (x(t), y(t)) \) to the system (21). Consider \( X \) — an \( \omega \)-limit set to solution \( x(t) \) of (22).

From invariance of \( \omega \)-limit sets it follows set \( Z \) consists of a union of solutions of (21). For any solution \( \{x^*(t), y^*(t) : t \in \mathbb{R}\} \subset Z \), we have \( \{x^*(t)\} \subset X \). By the Poincaré-Bendixson property of the original system, we can distinguish three cases:

1. \( x^*(t) \) is an equilibrium of (22).
2. \( x^*(t) \) is a periodic orbit of (22).
3. \( x^*(t) \) is a connecting orbit of (22) — a part of a cycle of connecting orbits.

In the rest of the proof we will frequently use the integral form of solutions \( y(t) \) to (21), given by \( y(t) = y(0) + \int_0^t g(x(s))ds \).

Case (1): We prove that \( (x^*(t), y^*(t)) \) is stationary for (21). It is enough to show \( g(x^*_t) = 0 \). Assume otherwise. Then \( |y^*(t)| = |y(0) + \int_0^t g(x^*_s)ds| \rightarrow \infty \) as \( t \rightarrow \pm \infty \). This contradicts the boundedness of an \( \omega \)-limit set.

Case (2) Let \( T \) be the period of \( x^*(t) \). We show that \( (x^*(t), y^*(t)) \) is a periodic solution of (21) of the same period. We have:
\[
\frac{d}{dt}(y^*(t + T) - y^*(t)) = \frac{d}{dt}\int_t^{t+T} g(x^*_s)ds
\]
\[
= g(x^*_t(T + t)) - g(x^*_t(t)) = 0,
\] (23)
hence \( y^*(t + T) - y^*(t) = \text{const} \). If this quantity would be non-zero, the diameter of the set \( \{y^*(t) : t \in \mathbb{R}\} \) would be infinite. However, the set \( Z \) is bounded, and therefore \( y^*(t + T) = y^*(t) \).

Case (3): We show that \( (x^*(t), y^*(t)) \) is a connecting orbit between two equilibria for the full system (21). We shall only prove convergence with \( t \rightarrow \infty \), the very same argument holds for \( t \rightarrow -\infty \) and \( \alpha \)-limit sets. The orbit \( (x^*(t), y^*(t)) \) is bounded and therefore it has an accumulation point as \( t \rightarrow \infty \) given by \( (x^{**}, y^{**}) \in \omega(x^*(0), y^*(0)) \). The point \( x^{**} \) is an equilibrium for (22). We will show that \( (x^{**}, y^{**}) \) is an equilibrium. It is enough to show that \( g(y^{**}) = 0 \). Assume otherwise. Then \( y^{**}(t) = y^{**} + tg(x^{**}) \) which is unbounded. However, it is also a part of \( \omega((x^*(0), y^*(0)) \), since \( \omega \)-limit sets are invariant. Boundedness of \( \omega((x^*(0), y^*(0)) \) leads to a contradiction. The same process, repeated for all connecting orbits of (22), creates a cycle of connecting orbits for (21).

Now, we can proceed to the proof of Theorem 4.

We are now ready to state and prove the theorem for GPGs with nodes of indegree at most one.

Theorem 4. Let \( \dot{z} = V(z) \) be a system of differential equations given by the follow-the-regularized leader dynamics of a binary, weakly connected, graphical polymatrix game, where each player has up to one predecessor. Then, for any smooth, steep, strictly convex collection of regularizers \( \{h_i\}_i \), such system possesses the Poincaré-Bendixson property.
Proof. By Lemma 1, and Remark 1, we know that the graph of the system has either a root vertex or a root cycle. We will first address the case of a root vertex. We will see that this case is somewhat degenerate. Without loss of generality let us assume that it is labelled as the 1st vertex, and that the other vertices are numbered in order of increasing path distance from vertex 1 (i.e. \( j < i \) implies that the path from 1 to \( j \) is shorter than the path from 1 to \( i \)) – this is possible by Lemma 1.

The payoffs of the root node are only affected by its own choice of strategy. Therefore, we can write \( z_1 = u_1(a_0) - u_1(a_1) \) and, consequently, \( z_1(t) = t(u_1(a_0) - u_1(a_1)) + z_1(0) \). This system constitutes an autonomous ODE, which trivially has the Poincaré-Bendixson property (as it is either completely stationary, or is divergent). From then on, we can add nodes, starting from vertices connected to the root vertex, and then continuing in an inductive fashion. Then, either one of the nodes diverges, or they are all stationary, and trivially satisfy the Poincaré-Bendixson property. It should be noted that “divergence” in practice means that \( z_1(t) \)’s approach in the limit \( t \to \infty \) to either \( \infty \) or \(-\infty\); the former implies that the player \( i \) is placing almost all probability mass on strategy \( a_0 \), and the latter on \( a_1 \).

The more interesting scenario arises for the root cycle, where periodic limit sets are possible. Enumerate these vertices by \( 1, \ldots, N_0 \), with \( N_0 \leq N \), and assume that the vertices from \( N_0 + 1 \) to \( N \) are arranged in the order of increasing path distance from vertices of the cycle (possible by Remark 1). Observe that the system

\[
\begin{align*}
\dot{z}_i &= V(z)_i, \\
i &= 1, \ldots, N_0,
\end{align*}
\]

is an autonomous system of differential equations (as there are no edges with successors in \( \{1, \ldots, N_0\} \), and predecessors outside of this set), and forms a binary, cyclic game in the sense of Theorem 3. As such, this subsystem possesses the Poincaré-Bendixson property. From then on, the proof continues similarly as for the root vertex. We add a vertex \( N_0 + 1 \) which has an incoming edge from the root cycle, and, by Lemma 3 observe that the system

\[
\begin{align*}
\dot{z}_i &= V(z)_i, \\
i &= 1, \ldots, N_0 + 1,
\end{align*}
\]

again has the Poincaré-Bendixson property. The proof continues inductively w.r.t. the vertices, until we conclude that the full system \( \dot{z} = V(z) \) has the Poincaré-Bendixson property.

Remark 3. Theorems 3, 4 apply to dynamics of fully mixed initial strategy profiles bounded away from pure strategies, as FoRel learning (4) is ill-defined for pure strategies. For some learning models such as as the replicator equations (14) the theorems can be applied to subsystems arising when certain players assume a pure strategy profile, as in these models pure strategy profiles define invariant learning spaces.

4 FROM GEOMETRY TO EFFICIENCY: SOCIAL WELFARE ANALYSIS

The following result shows that for cyclic, binary games, under additional but structurally robust assumptions on the payoff matrices (i.e., assumptions that remain valid after small perturbations of the payoff matrices), the time-average social welfare of our FoRel dynamics is at least as high, as the social welfare \( SW = \sum_i u_i \) of the worst Nash equilibrium. The proof crucially relies on the interplay of the optimal regret properties of FoRel dynamics combined with structural characterizations of the set of Nash equilibria of these games.

Theorem 5. In any binary, cyclic game with the property that for any player \( i \), the payoff entries are distinct and

\[
[A^{i-1,i}(a_0, a_0) - A^{i-1,i}(a_1, a_0)][A^{i-1,i}(a_0, a_1) - A^{i-1,i}(a_1, a_1)] < 0,
\]

the time-average of the social welfare of FoRel dynamics is at least that of the social welfare of the worst Nash equilibrium. Formally,

\[
\lim \inf \frac{1}{T} \int_0^T \sum_i u_i(x(t)) dt \geq \sum_i u_i(x_{NE}).
\]

where \( x_{NE} \) the worst case Nash equilibrium, i.e., a Nash equilibrium that minimizes the sum of utilities of all players.

In other words, the Nash equilibrium is the worst imaginable outcome for all players; and the dynamical, regret minimization approach yields superior payoffs.

Proof. Let’s consider the payoff matrix of each player \( i \). Recall, that by the cyclicity assumption, there is at most one player \( k \) such that \( A^{i,k} \) is a non-zero matrix, i.e., the unique predecessor of \( i \), that for simplicity of notation we call \( i - 1 \). By assumption, the four entries will be considered distinct. Next, we break down the analysis into two cases. As a first case, we consider the scenario where there exists at least one player with a strictly dominant strategy. The FoRel dynamics of that player strategy profile will trivially converge to playing the strictly dominant strategy with probability one. Similarly, all players reachable from player \( i \) will similarly best respond to it. This is clearly the unique NE for the binary cyclic game, so in this case the limit behavior of FoRel dynamics exactly corresponds to the unique Nash behavior and the theorem follows immediately.

Next, let’s consider the case where no player has a strictly dominant strategy. In this case, we will construct a specific Nash equilibrium for the cyclic game (although it may have more than one). In this Nash equilibrium each player \( i - 1 \) plays the unique mixed strategy that makes its successor (player \( i \)) indifferent between its two strategies. Such a strategy exists for each player, because otherwise there would exist a player with a strictly dominant strategy. In fact by the assumption \( [A^{i-1,i}(a_0, a_0) - A^{i-1,i}(a_1, a_0)][A^{i-1,i}(a_0, a_1) - A^{i-1,i}(a_1, a_1)] < 0 \) such a strategy would be the \( i - 1 \) player’s minmax strategy if they participated in a zero-sum game with player \( i \) defined by the payoff matrix of player \( i \). Indeed, this assumption, along with the fact that player \( i \) does not have a dominant strategy, exactly encodes that the zero-sum game (defined by payoff matrix \( A^{i-1,i} \)) has an interior Nash. Given this assumption, player \( i \) will be receiving exactly its max-min payoff no matter which strategy they select, therefore this strategy profile where each player \( i - 1 \) just plays the strategy that makes player \( i \) indifferent between their two options is a Nash equilibrium, where each player receives exactly their max-min payoffs. However, by [38] (Lemma C.1), continuous-time FoRel dynamics are no-regret with their time-average regret converging to zero at an optimal rate of
O(1/T), i.e. there exists an $\Omega_i > 0$, such that for all players $i$ we have:
\[
\max_{p_i \in \mathcal{X}_i} \frac{1}{T} \int_0^T (u_i(p_i; x_{-i}(t)) - u_i(x(t))) \, dt \leq \frac{\Omega_i}{T}.
\] (27)

However, the left hand side is greater or equal to
\[
u_i(x_{NE}) - \frac{1}{T} \int_0^T u_i(x(t)) dt,
\] (28)
since the mixed Nash equilibrium consists of max-min strategies. Therefore, the sum over $i$ of the time-average performance is at least the sum of the max-min utilities minus a quickly vanishing term $O(1/T)$ and the theorem follows.

\section{Examples}

To illustrate our theoretical results, we analyze the replicator dynamics (14) of two classes multidimensional binary cyclic games that exhibit non-convergence and therefore non-trivial limit behavior. The goal of the example is to show that all possible limit sets indicated in the Poincaré–Bendixon property (i.e., an equilibrium, a periodic solution, and a cycle of connecting solutions) are attainable for systems satisfying our assumptions. In addition, we plot the social welfare of simulated trajectories, relating them to the results of Theorem 5. Finally, we provide a counterexample in the form of a three-dimensional replicator system that violates the assumptions of our theorems and exhibits chaos. To determine the limit sets, we numerically integrate the initial-value problems with varied starting conditions via the Isoda differential equation integrator [25].

\subsection{5.1 Matched-mismatched pennies game}

First, we analyze a four-dimensional game of matched-mismatched pennies. Each player has a choice of two strategies, $a_0$ and $a_1$. The payoffs for players 0 and 2 are given by
\[
A^{0,1} = A^{1,2} = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}
\] (29)
and the payoffs for players 1 and 3 are given by
\[
A^{0,1} = A^{2,3} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.
\] (30)

Simply put, players 0 and 2 try to mismatch the strategy with players 1 and 3, and players 1 and 3 try to match them.

The system possesses three Nash equilibria, which correspond to the following strategy profiles: $(0, 0, 1, 1)$, $(1, 1, 0, 0)$, $(0.5, 0.5, 0.5, 0.5)$, out of which the pure Nash equilibria are attracting, and the mixed Nash equilibrium has two center directions: one repelling and one attracting. We denote the mixed Nash equilibrium by $x_{NE}$. Given the symmetry of the system, the plane $\{(t, s, t, s), t, s \in [0,1]\}$ is invariant, consists purely of periodic orbits, and forms the center manifold to the mixed Nash equilibrium.

The numerical results are consistent with Theorems 3 and 4. The only limit sets observed by the numerical simulations are the mixed Nash equilibrium $x_{NE}$ itself (along a single-dimensional attracting set) and the limit cycles around it, which also appear to be of saddle nature and have a single attracting direction, see Figure 2. Most crucially, more complicated behavior, such as chaos or invariant tori, does not emerge, despite the system being nontrivially embedded in four dimensions.

The mixed Nash equilibrium yields the minimax payoff vector $(0, 0, 0, 0)$ for each player and the social welfare of 0. The payoff matrices satisfy the assumptions of Theorem 5, and the average payoffs along solutions are therefore at least non-negative. In fact, almost all (a set of full measure) initial conditions appear to converge to the pure equilibria at the boundary, with their time-average payoffs exceeding that of the Nash equilibrium and converging to the maximal welfare of 4, see Figure 3.

\subsection{5.2 Asymmetric N-penny game}

Our second system is a system of $N$-player asymmetric mismatched pennies, previously introduced in [31]. There are three players, and each can choose between two strategies: $a_0$ and $a_1$. The payoffs for player $i$ with respect to player $i - 1$ are given by the matrix
\[
A_i = \begin{bmatrix} 0 & 1 \\ p & 0 \end{bmatrix}.
\] (31)

with $p > 0$.

For odd $N$, there is no Nash equilibrium in pure strategies. In the replicator system, the pure strategy profiles are saddle-type stationary points of the ordinary differential equation, linked by connecting orbits of mixed strategies. The system has a unique mixed Nash equilibrium defined by $x_i = \frac{1}{p+1}$, $i \in \{1, \ldots, N\}$, where each player obtains payoff of $\frac{p}{p+1}$.

The system was thoroughly analyzed in [31], and the main result given therein was that, for $N = 3$ and $p > 7$, all mixed strategies except for the diagonal converge to a sequence of orbits connecting boundary stationary points. Moreover, the social welfare attained close to the boundary exceeds the social welfare at the Nash equilibrium. We extend these results. From Theorem 3 we deduce that, for all $N$ and for all $p \neq -1$, the only limit sets in the interior are equilibria, periodic orbits, and cycles of connecting orbits to equilibria. The payoff matrices satisfy the assumptions of Theorem 5, and, in particular, for all $p > 0$, the mixed equilibrium yields the minimax payoff for each player, and time averages of payoffs along other orbits must exceed the minimax payoffs. For almost all initial
After some transformations (for details, see [43]), we arrive at the persistent strange (chaotic) attractor for a range of parameter values \( \mu \in [1.4645, 9.5055] \). We replicate their findings by integrating a sample trajectory and observing its approach to the chaotic attractor for \( \mu = 2.8 \), see Figure 4. Due to lack of cyclicity, the game does not guarantee the payoff structure given by Theorem 5.

### 6 CONCLUSIONS

Numerous recent results regarding learning in games have established a clear separation between the idealized behavior of equilibria and the erratic, unpredictable, and typically chaotic behavior of learning dynamics even in simple games and domains. At a first glance, this realization might seem to be a setback, but when viewed from the correct perspective it unveils a new way of understanding learning dynamics, namely, by examining solution concepts from the topology of dynamical systems. Our results showcase the possibility of establishing links between the topological-combinatorial structure of multi-agent games (e.g., game graph, number of actions) to understand and constrain the topological complexity of game dynamics (Poincaré-Bendixson property) and finally link back to more traditional game theoretic analyses, such as calculating the efficiency of the system via social welfare. These connections showcase the promising advantages of this approach, which we hope will lead to more work along these lines in the future.

**ACKNOWLEDGMENTS**

AC was supported by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. 758824 – INFLUENCE).

GP was supported in part by the National Research Foundation, Singapore under NRF 2018 Fellowship NRF-NRF2018-07, AI Singapore Program (AISG Award No: AISG2-CP-2020-016, NRF2019-NRF-ANR095 ALIAS grant, Technology and Research (A*STAR), AME Programmatic Fund (Grant No. A20H6b0151) from the Agency for Science, grant PIE-SGP-AL-2018-01 and Provost’s Chair Professorship grant RGEPPV2101.

We would like to thank prof. Frans A. Oliehoek for his support, and helpful advice.

Figure 3: Time-average payoffs and social welfare of a sample learning trajectory in the matched-mismatched pennies game (top), and in the asymmetric 5-penny game with \( p = 3 \) (bottom, projection onto first three variables).

Figure 4: A learning trajectory approaching a chaotic attractor in the polymatrix replicator [43] (left) and a plot of values of its coordinates (right). The game is characterized by unstructured interactions between payoffs and therefore breaks the assumptions of Theorems 3 and 4.
