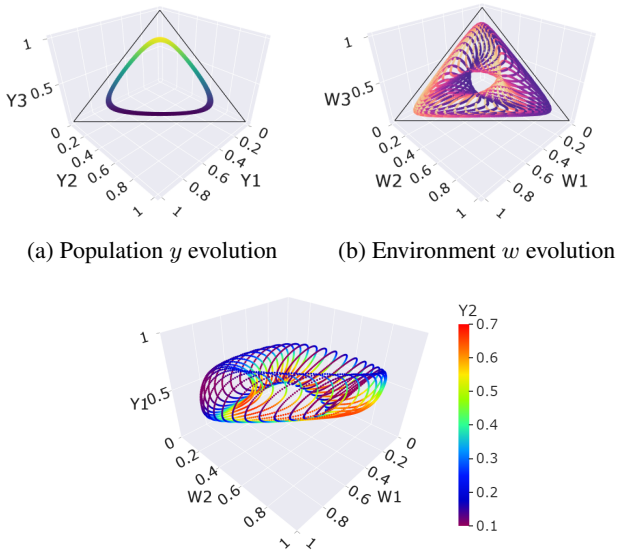
Evolutionary Game Theory Squared: Evolving Agents in Endogenously Evolving Zero-Sum Games

Stratis Skoulakis,[1] **Tanner Fiez,**[2] **Ryann Sim,**[1] **Georgios Piliouras,**[1*] **Lillian Ratliff** [2*]
[1] Singapore University of Technology and Design
[2] University of Washington
{efstratios, georgios}@sutd.edu.sg, {fiezt, ratliffl}@uw.edu, ryann_sim@mymail.sutd.edu.sg

Abstract

The predominant paradigm in evolutionary game theory and more generally online learning in games is based on a clear distinction between a population of *dynamic agents* that interact given a *fixed, static game*. In this paper, we move away from the artificial divide between dynamic agents and static games, to introduce and analyze a large class of competitive settings where both the agents and the games they play evolve strategically over time. We focus on arguably the most archetypal game-theoretic setting—zero-sum games (as well as network generalizations)—and the most studied evolutionary learning dynamic—replicator, the continuous-time analogue of multiplicative weights. Populations of agents compete against each other in a zero-sum competition that itself evolves adversarially to the current population mixture. Remarkably, despite the chaotic coevolution of agents and games, we prove that the system exhibits a number of regularities. First, the system has *conservation laws* of an information-theoretic flavor that couple the behavior of all agents and games. Secondly, the system is *Poincaré recurrent*, with effectively all possible initializations of agents and games lying on recurrent orbits that come arbitrarily close to their initial conditions infinitely often. Thirdly, the *time-average agent behavior and utility converge* to the Nash equilibrium values of the *time-average* game. Finally, we provide a polynomial time algorithm to efficiently predict this time-average behavior for any such coevolving network game.



(a) Population y evolution (b) Environment w evolution

(c) Coevolution of y and w

Figure 1: Poincaré recurrence in a time-evolving generalized Rock-Paper-Scissors model.

1 Introduction

The problem of analyzing evolutionary learning dynamics in games is of fundamental importance in several fields such as evolutionary game theory (Sandholm 2010), online learning in games (Cesa-Bianchi and Lugosi 2006; Nisan et al. 2007), and multi-agent systems (Shoham and Leyton-Brown 2008). The dominant paradigm in each area is that of evolutionary agents adapting to each others behavior. In other words, the dynamism of the environment of each agent is driven by the other agents, whereas the rules of interaction between the agents, that is, the game, is static. This separation between *evolving agents* and a *static game* is so standard that it typically goes unnoticed, however, this fundamental restriction does not allow us to capture many applications of interest. In artificial intelligence (Wang et al. 2019; Garciarena, Santana, and Mendiburu 2018; Costa et al. 2019; Miikkulainen et al. 2019; Wu et al. 2019; Stanley and Miikkulainen 2002) as well as biology, sociology, and economics (Stewart and Plotkin 2014; Tilman, Plotkin, and Akçay 2020; Tilman, Watson, and Levin 2017; Bowles, Choi, and Hopfensitz 2003; Weitz et al. 2016), the rules of interaction can themselves adapt to the collective history of the agent behavior. For example, in adversarial learning and curriculum learning (Huang et al. 2011; Bengio et al. 2009), the difficulty of the game can increase over time by exactly focusing on the settings where the agent has performed the weakest. Similarly, in biology or economics, if a particular advantageous strategy is used exhaustively by agents, then its relative advantages typically dissipate over time (negative frequency-dependent selection, see Heino, Metz, and Kaitala 1998), which once again drives the need for innovation and exploration.

In all these cases, the game itself stops being a passive

*Joint last authors

object that the agents act upon, but instead is best thought of as an algorithm itself. Similar to online learning algorithms employed by agents, the game itself may have a memory/state that encodes history. However, unlike online learning algorithms that receive a history or sequence of payoff vectors and output the current behavior (e.g., a probability distribution over actions), an algorithmic game receives as input a history or sequence of agents' behavior and outputs a new payoff matrix. Hence, learning and games are "dual" algorithmic objects which are coupled in their evolution (Figure 1).

How does one even hope to analyze evolutionary learning in time-evolving games? Once we move away from the safe haven of static games, we lose our prized standard methodology that roughly consists of two steps: i) compute/understand the equilibria of the given game (e.g., Nash, correlated, etc.; see Nash 1951; Aumann 1974) and their properties; ii) connect the behavior of learning dynamics to a target class of equilibria (e.g., convergence). Indeed, the only prior work to ours, namely by Mai et al. (2018), which considers games larger than 2×2, focused on a specific payoff matrix structure based on Rock-Paper-Scissors (RPS) and argued recurrent behavior via a tailored argument that was explicitly designed for the dynamical system in question with no clear connections to game theory. We revisit this problem and find a new systematic game-theoretic analysis that generalizes to arbitrary network zero-sum games.

Contributions. We provide a general framework for analyzing learning agents in time-evolving zero-sum games as well as rescaled network generalizations thereof. To begin, we develop a novel *reduction* that takes as input time-evolving games and reduces them to a game-theoretic graph that generalizes both graphical zero-sum games and evolutionary zero-sum games. In this generalized but static game, evolving agents and evolving games represent different types of nodes (nodes with and without self-loops) in a graph connected by edge games. The bridge we form between time-evolving games and static network games makes the latter far more interesting than previously thought: *our reduction proves they are sufficiently expressive to capture not only multiple pairwise interactions, but time-varying environments as well.* Moreover, by providing a path back to the familiar territory of evolving agents interacting in a static game, the mathematical tools of game theory and dynamical systems theory become available. This allows us to perform a general algorithmic analysis of commonly studied systems from machine learning and biology previously requiring individualized treatment.

From an algorithmic learning perspective, we focus on the most studied evolutionary learning dynamic: replicator, the continuous-time analogue of the multiplicative weights update. Remarkably, despite the chaotic coevolution of agents and games that forces agents to continually innovate, the system can be shown to exhibit a number of regularities. We prove the system is *Poincaré recurrent*, with effectively all initializations of agents and games lying on recurrent orbits that come arbitrarily close to their initial conditions infinitely often (Figure 1). As a crucial component of this result, we demonstrate the dynamics obey information-theoretic *conservation laws* that couple the behavior of all agents and games (Figure 3). Moreover, while the system never equilibrates, the conservation laws allow us to prove the *time-average behavior and utility of the agents converge* to the time-average Nash of their evolving games with bounded regret. Finally, we provide a *polynomial time algorithm* that predicts these time-average quantities. Some proofs along with further experiments have been moved to an accompanying technical report (Skoulakis et al. 2020) due to space constraints.

Related Work and Technical Novelty. Our work relates with the rich previous literature studying the emerging recurrent behavior of replicator dynamics in (network) zero-sum games (Piliouras et al. 2014; Piliouras and Shamma 2014; Boone and Piliouras 2019; Mertikopoulos, Papadimitriou, and Piliouras 2018; Nagarajan, Balduzzi, and Piliouras 2020; Perolat et al. 2020). Unfortunately, this proof technique is an immediate dead-end for time-evolving zero-sum games since the KL divergence between the (evolving) strategies and (evolving) Nash equilibrium need not be a constant of motion. In particular, it is not even clear what the static concept of a Nash equilibrium means in this context. Despite this fact, Mai et al. (2018) managed to prove recurrence via constructing an invariant function for a specific evolving RPS game. However, their invariant function relies on the symmetries of the game and has no deeper interpretation or obvious generalization. A key contribution of our work is the development of a novel characterization of a general class of time-evolving games that possess a number of regularities including recurrence, which we demonstrate by deriving an information theoretic invariant. In particular, this allows us to not only generalize the recurrence results of time-evolving games to a class with much richer and complex interactions than the one studied in Mai et al. (2018), but also provides a naturally interpretable invariant in such time-evolving games.

2 Preliminaries and Definitions

In this section, we formalize the concept of polymatrix games, define the replicator dynamics for this class of games, and provide background material on dynamical systems that is relevant to our results.

Polymatrix Games. An N-player *polymatrix game* is defined using an undirected graph $G = (V, E)$ where V corresponds to the set of agents (or players) and E corresponds to the set of edges between agents in which a *bimatrix game* is played between the endpoints (Cai and Daskalakis 2011). Each agent $i \in V$ has a set of actions $\mathcal{A}_i = \{1, \ldots, n_i\}$ that can be selected at random from a distribution x_i called a *mixed strategy*. The set of mixed strategies of player $i \in V$ is the standard simplex in $\mathbb{R}^{n_i}$ and is denoted $\mathcal{X}_i = \Delta^{n_i-1} = \{x_i \in \mathbb{R}^{n_i}_{\geq 0} : \sum_{\alpha\in\mathcal{A}_i} x_{i\alpha} = 1\}$ where $x_{i\alpha}$ denotes the probability mass on action $\alpha \in \mathcal{A}_i$. The state of the game is then defined by the concatenation of the strategies of all players. We call the set of all possible strategies profiles the *strategy space*, and denote it by $\mathcal{X} = \prod_{i\in V} \mathcal{X}_i$.

The bimatrix game on edge (i, j) is described using a pair of matrices $A^{ij} \in \mathbb{R}^{n_i\times n_j}$ and $A^{ji} \in \mathbb{R}^{n_j\times n_i}$. An entry $A^{ij}_{\alpha\beta}$ for $(\alpha, \beta) \in \mathcal{A}_i \times \mathcal{A}_j$ represents the reward player i

obtains for selecting action α given that player j chooses action β. We note that the graph G may also contain *self-loops*, meaning that an agent $i \in V$ plays a game defined by A^{ii} against itself. The *utility* or *payoff* of agent $i \in V$ under the strategy profile $x \in \mathcal{X}$ is denoted by $u_i(x)$ and corresponds to the sum of payoffs from the bimatrix games the agent participates in. The payoff is equivalently expressed as $u_i(x_i, x_{-i})$ when distinguishing between the strategy of player i and all other players $-i$. More precisely,

$$u_i(x) = \sum_{j:(i,j)\in E} x_i^\top A^{ij} x_j. \tag{1}$$

We further denote by $u_{i\alpha}(x) = \sum_{j:(i,j)\in E}(A^{ij}x_j)_\alpha$ the utility of player $i \in V$ under the strategy profile $x = (\alpha, x_{-i}) \in \mathcal{X}$ for $\alpha \in \mathcal{A}_i$. The game is called *zero-sum* if $\sum_{i\in V} u_i(x) = 0$ for all $x \in \mathcal{X}$. Moreover, if there are positive coefficients $\{\eta_i\}_{i\in V}$ such that $\sum_{i\in V} \eta_i u_i(x) = 0$ for all $x \in \mathcal{X}$ and the self-loops are antisymmetric (meaning $A^{ii} = -(A^{ii})^\top$), the game is called *rescaled zero-sum*.

A common notion of equilibrium behavior in game theory is that of a Nash equilibrium, which is defined as a mixed strategy profile $x^* \in \mathcal{X}$ such that for each player $i \in V$,

$$u_i(x_i^*, x_{-i}^*) \geq u_i(x_i, x_{-i}^*),\ \forall x_i \in \mathcal{X}_i. \tag{2}$$

We denote the support of $x_i^* \in \mathcal{X}_i$ by $\mathrm{supp}(x_i^*) = \{\alpha \in \mathcal{A}_i : x_{i\alpha} > 0\}$. A Nash equilibrium is said to be an *interior* or *fully mixed* Nash equilibrium if $\mathrm{supp}(x_i^*) = \mathcal{A}_i\ \forall i \in V$.

Replicator Dynamics. In polymatrix games, *replicator dynamics* (Sandholm 2010) for each $i \in V$ are given by

$$\dot{x}_{i\alpha} = x_{i\alpha}(u_{i\alpha}(x) - u_i(x)),\ \ \forall \alpha \in \mathcal{A}_i. \tag{3}$$

We suppress the explicit dependence on time t in the system and do so throughout where clear from context to simplify notation. Moreover, we consider initial conditions on the interior of the simplex. The replicator dynamics are equivalently given in vector form for each $i \in V$ by the system

$$\dot{x}_i = x_i \cdot \big(\textstyle\sum_{j:(i,j)\in E} A^{ij}x_j - \big(\sum_{j:(i,j)\in E} x_i^\top A^{ij} x_j\big) \cdot \mathbf{1}\big), \tag{4}$$

where $\mathbf{1}$ is an n_i–dimensional vector of ones and the operator $(\cdot)$ denotes elementwise multiplication.

For the purpose of analysis, the replicator dynamics in (3) are often translated by a diffeomorphism from the interior of $\mathcal{X}$ to the cumulative payoff space $\mathcal{C} = \prod_{i\in V} \mathbb{R}^{n_i - 1}$, which is defined by a mapping such that $x_i = (x_{i1}, \ldots, x_{in_i}) \mapsto (\ln \frac{x_{i2}}{x_{i1}}, \ldots, \ln \frac{x_{in_i}}{x_{i1}})$ for each player $i \in V$.

Review of Topology of Dynamical Systems. We now review some concepts from dynamical systems theory that will help us prove Poincaré recurrence. Further background material can be found in the book of Alongi and Nelson (2007).

Flows: Consider a differential equation $\dot{x} = f(x)$ on a topological space X. The existence and uniqueness theorem for ordinary differential equations guarantees that there exists a unique continuous function $\phi : \mathbb{R} \times X \to X$, which is termed the *flow*, that satisfies (i) $\phi(t, \cdot) : X \to X$—often denoted $\phi^t : X \to X$—is a homeomorphism for each $t \in \mathbb{R}$, (ii) $\phi(t+s, x) = \phi(t, \phi(s, x))$ for all $t, s \in \mathbb{R}$ and all $x \in X$, and (iii) for each $x \in X$, $\frac{d}{dt}|_{t=0}\phi(t, x) = f(x)$. Since the replicator dynamics are Lipschitz continuous, a unique flow ϕ of the replicator dynamics exists.

Conservation of Volume: The flow ϕ of a system of ordinary differential equations is called *volume preserving* if the volume of the image of any set $U \subseteq \mathbb{R}^d$ under ϕ^t is preserved. More precisely, for any set $U \subseteq \mathbb{R}^d$, $\mathrm{vol}(\phi^t(U)) = \mathrm{vol}(U)$. Whether or not a flow preserves volume can be determined by applying *Liouville's theorem*, which says the flow is volume preserving if and only if the divergence of f at any point $x \in \mathbb{R}^d$ equals zero—that is, $\mathrm{div} f(x) = \mathrm{tr}(Df(x)) = \sum_{i=1}^d \frac{df(x)}{dx_i} = 0$.

Poincaré Recurrence: If a dynamical system preserves volume and every orbit remains bounded, almost all trajectories return arbitrarily close to their initial position, and do so infinitely often (Poincaré 1890). Given a flow ϕ^t on a topological space X, a point $x \in X$ is *nonwandering* for ϕ^t if for each open neighborhood U containing x, there exists $T > 1$ such that $U \cap \phi^T(U) \neq \emptyset$. The set of all nonwandering points for ϕ^t, called the *nonwandering set*, is denoted $\Omega(\phi^t)$.

Theorem 2.1 (Poincaré Recurrence (Poincaré 1890))**.** *If a flow preserves volume and has only bounded orbits, then for each open set almost all orbits intersecting the set intersect it infinitely often: if ϕ^t is a volume preserving flow on a bounded set $Z \subset \mathbb{R}^d$, then $\Omega(\phi^t) = Z$.*

3 Studying Doubly Evolutionary Processes via Polymatrix Games

Numerous applications from artificial intelligence (AI) and machine learning (ML) to biology cast competition between populations (e.g., neural networks/algorithms or species/agents) and the environment (e.g., hyperparameters/network configurations or resources) as a time-evolving dynamical system. The basic abstraction takes the form of a population y of *species* which evolve dynamically in time as a function of itself and some *environment* parameters w whose evolution, in turn, depends on y. We now review models from each application and then connect a broad class of time-evolving dynamical systems to static polymatrix games. This reduction provides a path toward analyzing complex non-stationary dynamics using tools developed for the typical static game formulation.

Doubly Evolutionary Behavior in AI and ML. Evolutionary game theory methods for training generative adversarial networks commonly exhibit time-evolving dynamic behavior and there is a pair of predominant doubly evolutionary process models (Costa et al. 2020; Wang et al. 2019; Garciarena, Santana, and Mendiburu 2018; Costa et al. 2019; Miikkulainen et al. 2019). In the first formulation, Wang et al. (2019) describe training the generator network, with parameters y, via a gradient-based algorithm composed of *variation*, *evaluation*, and *selection*. The discriminator network, with parameters w updated via gradient-based learning, is modeled as the environment operating in a feedback loop with y. The second model is such that the generator and discriminator are different species (or *modules*) in the population y which follows evolutionary dynamics, and network hyperparameters (or *chromosomes*) w evolve in time as a function

of y (Garciarena, Santana, and Mendiburu 2018; Costa et al. 2019; Miikkulainen et al. 2019). We connect further to AI and ML applications in the discussion where we highlight exciting future directions (Section 7).

Doubly Evolutionary Behavior in Biology. There are also two common formulations emerging in biology. In the first, the focus is on the level of coordination in a population as a function of evolving environmental variables. The prevailing model is comprised of replicator dynamics $\dot{y} = y(1-y)((A(w)y)_1 - (A(w)y)_2)$ in which a population of two species y plays a prisoner's dilemma (PD) game against themselves in a setting where the payoff matrix $A(w)$ depends on an environment variable w which, in turn, depends on the population via $\dot{w} = w(1-w)G(y)$ where $G(y)$ is a feedback mechanism describing when environmental degradation or enhancement occurs as a function of y (Weitz et al. 2016; Tilman, Plotkin, and Akçay 2020; Tilman, Watson, and Levin 2017; Lade et al. 2013); e.g., in Weitz et al. (2016), $G(y)$ takes the form $\theta y - (1-y)$ for some $\theta > 0$ which represents the ratio of the enhancement rate to degradation rate of 'cooperators' and 'defectors' in the time-evolving PD game. In the second formulation, the focus is on studying how competition among species is modulated by resource availability. Indeed, from a biological perspective, Mai et al. (2018) argue that the environment parameters w on which a population y of n antagonistic species depend are not constant, but rather evolve over time. Since the species fitness depends on the environment, the game among the species is also time-varying. The adopted model of the dynamic behavior with initial conditions on the interior of the simplex for both w and y is given for each $i \in \{1, \dots, n\}$ by

$$\begin{aligned} \dot{w}_i &= w_i \textstyle\sum_{j=1}^{n} w_j(y_j - y_i) \\ \dot{y}_i &= y_i\big((P(w)y)_i - y^\top P(w)y\big) \end{aligned} \tag{5}$$

where $P(w) = P + \mu W$ for $\mu > 0$ with P defined as the generalized RPS payoff matrix

$$P = \begin{pmatrix} 0 & -1 & 0 & \cdots & 0 & 0 & 1 \\ 1 & 0 & -1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 & -1 \\ -1 & 0 & 0 & \cdots & 0 & 1 & 0 \end{pmatrix},$$

and the environmental variations matrix

$$W = \begin{pmatrix} 0 & w_1 - w_2 & \cdots & w_1 - w_n \\ w_2 - w_1 & 0 & \cdots & w_2 - w_n \\ \vdots & \vdots & \vdots & \vdots \\ w_n - w_1 & w_n - w_2 & \cdots & 0 \end{pmatrix}.$$

Reducing Time-Evolving RPS to a Polymatrix Game. Mai et al. (2018) studied the dynamical system in (5) and showed it exhibits a special type of cyclic behavior: *Poincaré recurrence*. By capturing the evolution of the environment (dynamics of the payoff matrix) as additional players that dynamically change their strategies, we reduce the coevolution of w and y to a *static polymatrix game* of greater dimensionality (greater number of players). Given this reduction, Theorem 4.1, which establishes the Poincaré recurrence of replicator dynamics in rescaled zero-sum polymatrix games, immediately captures the results of Mai et al. (2018) (see Corollary 4.1).

Proposition 3.1. *The time-evolving generalized rock-paper-scissors game from* (5) *is equivalent to replicator dynamics in a two-player rescaled zero-sum polymatrix game.*

Proof Sketch. The initial condition $w(0)$ is on the interior of the simplex and $\sum_{i=1}^{n} \dot{w}_i = 0$. Consequently, $\sum_{i=1}^{n} w_i(0) = \sum_{i=1}^{n} w_i(t) = 1$, and we obtain

$$\dot{w}_i = w_i \textstyle\sum_{j=1}^{n} w_j(y_j - y_i) = w_i\big(-y_i + \sum_{j=1}^{n} w_j y_j\big),$$

which is the replicator equation of a node w in a polymatrix game with payoff matrix $A^{wy} = -I$. Using a similar decomposition, we reformulate the y dynamics:

$$\dot{y}_i = y_i\big((Py)_i - y^\top Py\big) + y_i\big(\mu w_i - \mu \textstyle\sum_{j=1}^{n} w_j y_j\big).$$

This corresponds to the replicator equation of node y playing against itself with $A^{yy} = P$ and against w with $A^{yw} = \mu I$. The game is rescaled zero-sum with $\eta_y = 1$ and $\eta_w = \mu$. □

Generalized Reduction. The previous reduction generalizes to a class of time-evolving games defined by a set of populations $y = (y_1, \dots, y_{n_y})$ and environments $w = (w_1, \dots, w_{n_w})$, where $y_\ell \in \Delta^{n-1}$ for each $\ell \in \{1, \dots, n_y\}$ and $w_k \in \Delta^{n-1}$ for each $k \in \{1, \dots, n_w\}$. Environments coevolve with only populations and not other environments, while any population coevolves only with environments and itself. Let $\mathcal{N}_k^w$ be the set of populations which coevolve with w_k and $\mathcal{N}_\ell^y$ be the set of environments which coevolve with y_ℓ. The time-evolving dynamics for each environment k and population ℓ are given componentwise by

$$\dot{w}_{k,i} = w_{k,i} \sum_{\ell \in \mathcal{N}_k^w} \sum_{j} w_{k,j}\left((A^{k,\ell} y_\ell)_i - (A^{k,\ell} y_\ell)_j\right), \tag{6}$$

$$\dot{y}_{\ell,i} = y_{\ell,i}\left((P_\ell(w) y_\ell)_i - y_\ell^\top P_\ell(w) y_\ell\right), \tag{7}$$

where $P_\ell(w) = P_\ell + \sum_{k \in \mathcal{N}_\ell^y} W^{\ell,k}$ with $P_\ell \in \mathbb{R}^{n \times n}$ and $W^{\ell,k} \in \mathbb{R}^{n \times n}$ is defined such that the (i,j)–th entry is $(A^{\ell,k} w_k)_i - (A^{\ell,k} w_k)_j$.

Despite the complex nature of this dynamical system, we can show that it is equivalent to replicator dynamics in a polymatrix game.

Theorem 3.1. *Any time-evolving system defined by the dynamics in (6-7) is equivalent to replicator dynamics in a polymatrix game.*

The expressive power we gain from this reduction permits us to efficiently describe and characterize coevolutionary processes of higher complexity than past work since we can return to the familiar territory of analyzing dynamic agents in static games. In what follows we focus on providing theoretical results for the subclass of time-evolving systems which reduce to a rescaled zero-sum game. However, this reduction is of independent interest since it can prove useful for future work analyzing the class of general-sum games after the behavior of network zero-sum games and rescaled generalizations are well understood.

4 Poincaré Recurrence

In this section, we show that the replicator dynamics are Poincaré recurrent in N-player rescaled zero-sum polymatrix games with interior Nash equilibria. In particular, for almost all initial conditions $x(0) \in \mathcal{X}$, the replicator dynamics will return arbitrarily close to $x(0)$ an infinite number of times.

Theorem 4.1. *The replicator dynamics given in* (3) *are Poincaré recurrent in any N-player rescaled zero-sum polymatrix game that has an interior Nash equilibrium.*

Boone and Piliouras (2019), the closest known result, prove replicator dynamics are Poincaré recurrent in N-player *pairwise zero-sum* polymatrix games with an interior Nash equilibria, which requires $A^{ij} = -(A^{ji})^\top$ for every $(i,j) \in E$. Our extension to N-player *rescaled* zero-sum polymatrix games is a far more general characterization of the Poincaré recurrence of replicator dynamics since there are no explicit restrictions on the edge games and the polymatrix game itself need not even be strictly zero-sum. The significance of this result is further enhanced by the connection developed in Section 3 between a class of time-evolving games and N-player rescaled zero-sum polymatrix games. As a concrete example, given the reduction of Proposition 3.1, Theorem 4.1 recovers the work of Mai et al. (2018).

Corollary 4.1. *The time-evolving generalized rock-paper-scissors game in* (5) *is Poincaré recurrent.*

It is worth noting that the technical results we prove in order to show the system is Poincaré recurrent, namely volume preservation and the bounded orbits property, are themselves independently important as they provide conservation laws that couple the behavior of agents. In fact, they are fundamental to showing that while the system never equilibrates, the time-average dynamics and utility converge to the Nash equilibrium and its utility.

Overview of Proof Methods. To prove Poincaré recurrence, we need to show the flow corresponding to the system of ordinary differential equations in (3) is volume preserving and has bounded orbits (cf. Theorem 2.1). Notice that the flow of (3) always has bounded orbits since $x_{i\alpha} \geq 0$ and $\sum_{\alpha \in \mathcal{A}_i} x_{i\alpha}(t) = 1 \ \forall \ i \in V$, however proving the volume preserving property is not as straightforward. To show volume preservation, we transform the dynamics via a *canonical transformation*. Indeed, we prove Poincaré recurrence of the flow of a system of ordinary differential equations that is diffeomorphic to the flow of the replicator equation. Given $x \in \mathcal{X}$, consider the transformed variable $z \in \mathbb{R}^{n_1+\cdots+n_N-N}$ defined by

$$z_i = \left(\ln \tfrac{x_{i2}}{x_{i1}}, \ldots, \ln \tfrac{x_{in_i}}{x_{i1}} \right), \ \forall i \in V. \tag{8}$$

Given the vector z_i, the components of x_i are given by $x_{i\alpha} = e^{z_{i\alpha}} / (\sum_{\ell=1}^{n_i} e^{z_{i\ell}})$. Under this transformation, $\dot{z} = F(z)$ is given componentwise for each $\alpha \in \mathcal{A}_i$ and all $i \in V$ by

$$\begin{aligned} \dot{z}_{i\alpha} = F_{i\alpha}(z) &= \frac{\dot{x}_{i\alpha}}{x_{i\alpha}} - \frac{\dot{x}_{i1}}{x_{i1}} \\ &= \textstyle\sum_{j \in V} \sum_{\beta \in \mathcal{A}_j} (A^{ij}_{\alpha\beta} - A^{ij}_{1\beta}) e^{z_{j\beta}} / \sum_{\ell=1}^{n_j} e^{z_{j\ell}}. \end{aligned} \tag{9}$$

Observe that $F_{i1} = 0$, meaning $z_{i1} = 0$ for all time. To show Poincaré recurrence of (3), we prove two key properties: *(i)* the flow of $\dot{z}$ is volume preserving, meaning the trace Jacobian of the respective vector field $\dot{z} = F(z)$ is zero, and, *(ii)* $\dot{z}$ has bounded orbits from any interior initial condition. Then, the Poincaré recurrence of $\dot{z}$, and consequently $\dot{x}$, follows from Theorem 2.1.

Conservation of Volume. We show that the trace of the vector field $F(z)$ is zero, which then from Liouville's theorem guarantees $\dot{z}$, as defined in (9), is volume preserving.

Lemma 4.1. *For any N-player rescaled zero-sum polymatrix game,* $\operatorname{tr}(DF(z)) = \sum_{i \in V} \sum_{\alpha \in \mathcal{A}_i} \frac{d}{dz_{i\alpha}} F_{i\alpha}(z) = 0$.

The proof of Lemma 4.1 crucially relies on the fact the self-loops are antisymmetric, $(A^{ii})^\top = -A^{ii}$.

Bounded Orbits. In order to prove that the orbits from any initial interior point $z(0)$ are bounded, we show that for any initial interior point $x(0)$, the orbit produced by the replicator dynamics stays on the interior of the simplex, that is, there exists a fixed parameter $\epsilon > 0$ such that for any agent $i \in V$ and strategy $\alpha \in \mathcal{A}_i$, $\epsilon \leq x_{i\alpha} \leq 1 - \epsilon$. Then, $|z_{i\alpha}|$ is clearly bounded since $z_{i\alpha} = \ln(x_{i\alpha}/x_{1\alpha})$.

Lemma 4.2. *Consider an N-player rescaled zero-sum polymatrix game such that for positive coefficients $\{\eta_i\}_{i \in V}$, $\sum_{i \in V} \eta_i u_i(x) = 0$ for $x \in \mathcal{X}$. If the game admits an interior Nash Equilibrium x^*, then $\Phi(t) = \sum_{i \in V} \sum_{\alpha \in \mathcal{A}_i} \eta_i x^*_{i\alpha} \ln x_{i\alpha}$ is time-invariant, meaning $\Phi(t) = \Phi(0)$ for $t \geq 0$. Hence, orbits from any interior initial condition $x(0)$ remain on the interior of the simplex.*

From the preceding discussion, Lemma 4.2 guarantees orbits from any interior initial condition $z(0)$ remain bounded. The proof of Lemma 4.2 is the primary novelty in the proof of Theorem 4.1 and the techniques may be of independent interest. To show $\Phi(t)$ is time-invariant, we prove that the time derivative of the function is equal to zero. From the given form of the replicator dynamics and the rescaled zero-sum property of the polymatrix game, we obtain $\dot{\Phi}(t) = \sum_{i \in V} \sum_{j:(i,j) \in E} \eta_i (x^*_i)^\top A^{ij} (x_j - x^*_j)$ nearly immediately, where the sum over edges describes how the rescaled utility of agent $i \in V$ changes at her equilibrium strategy when the rest of the players are allowed to deviate. To continue, we draw a key connection to a fascinating result regarding the payoff structure of zero-sum polymatrix games.

Cai and Daskalakis (2011) proved there exists a payoff preserving transformation from any zero-sum polymatrix game to a pairwise constant-sum polymatrix game. We translate this result to rescaled zero-sum polymatrix games. The primary implication is that the change in player i's rescaled utility at equilibrium when all other players connected to i deviate is equal to the change in player j's rescaled utility from deviating while all other players connected to j remain in equilibrium. This is a direct consequence of the fact that the game is equivalent to a pairwise constant-sum game. Explicitly, we prove that $\dot{\Phi}(t) = \sum_{j \in V} \sum_{i:(j,i) \in E} \eta_j (x^*_j - x_j)^\top A^{ji} x^*_i$ and conclude $\dot{\Phi}(t) = 0$ since x^* is an interior Nash equilibrium, which means $u_{j\alpha}(x^*) = u_j(x^*)$ for $\alpha \in \mathcal{A}_j$ and any linear combination.

Proof of Theorem 4.1. The proof follows directly from Lemma 4.1, Lemma 4.2, and Theorem 2.1. Indeed, the dynamics in (9) are Poincaré recurrent since from Lemma 4.1 they are volume preserving and from Lemma 4.2 the orbits are bounded. This property in the cumulative payoff space carries over to the dynamics in the strategy space from (3) since the transformation is a diffeomorphism. □

5 Time-Average Behavior, Equilibrium Computation, & Bounded Regret

In this section, we transition away from analyzing the dynamic behavior of replicator dynamics and focus on characterizing the long-term behavior along with its connections to notions of equilibrium and regret. We prove that the enduring system behavior is guaranteed to satisfy a number of desirable game-theoretic metrics of consistency and optimality. Moreover, we design a polynomial time algorithm able to predict this behavior.

While the replicator dynamics exhibit complex dynamics and never equilibriate in rescaled zero-sum polymatrix games with interior Nash equilibrium, the time-average behavior of the dynamics is closely tied to the equilibrium. The following result shows that given the existence of a unique interior Nash equilibrium, the time-average of the replicator dynamics converges to the equilibrium and the time-average utility converges to the utility at the equilibrium.

Theorem 5.1. *Consider an N-player rescaled zero-sum polymatrix game that admits a unique interior Nash equilibrium x^*. The trajectory $x(t)$ produced by replicator dynamics given in* (3) *is such that* ***i)*** $\lim_{t\to\infty} \frac{1}{t}\int_0^t x(\tau)d\tau = x^*$ *and* ***ii)*** $\lim_{t\to\infty} \frac{1}{t}\int_0^t u_i(x(\tau))d\tau = u_i(x^*)$.

The preceding result provides a broad generalization of past results that show the time-average of replicator dynamics converges to the unique interior Nash equilibrium in zero-sum bimatrix games (Hofbauer, Sorin, and Viossat 2009). We remark that our proof crucially relies on Lemma 4.2 since the trajectory of the dynamics must remain on the interior of the simplex to guarantee there exists a bounded sequence which admits a subsequence that converges to a limit corresponding to the time-average.

We now provide a polynomial time algorithm that efficiently predicts the time-average quantities even for an arbitrary networks of players. Linear programming formulations for computing and characterizing the set of Nash equilibria for zero-sum polymatrix games are known (Cai et al. 2016). The following result extends this formulation to rescaled zero-sum polymatrix games.

Theorem 5.2. *Consider an N-player rescaled zero-sum polymatrix game such that for positive coefficients $\{\eta_i\}_{i\in V}$, $\sum_{i=1}^N \eta_i u_i(x) = 0$ for $x \in \mathcal{X}$. The optimal solution of the following linear program is a Nash equilibrium of the game:*

$$\min_{x\in\mathcal{X}} \{\textstyle\sum_{i=1}^n \eta_i v_i |\ v_i \geq u_{i\alpha}(x),\ \forall\, i \in V,\ \forall\, \alpha \in \mathcal{A}_i\}$$

It cannot be universally expected that an interior equilibrium exists or that players are fully rational and obey a common learning rule. Similarly, players may not always be able to determine an equilibrium strategy *a priori* depending on the information available. This motivates an evaluation of the trajectory of a player who is oblivious to opponent behavior. We consider a notion of *regret* for a player. That is, the time-averaged utility difference between the mixed strategies selected along the learning path $t \geq 0$ and the fixed strategy that maximizes the utility in hindsight. Even in polymatrix games (with self-loops), the regret of replicator dynamics stays bounded.

Proposition 5.1. *Any player following the replicator dynamics* (3) *in an N-player polymatrix game (with self-loops) achieves an $\mathcal{O}(1/t)$ regret bound independent of the rest of the players. Formally, for every trajectory $x_{-i}(t)$, the regret of player $i \in V$ is bounded as follows for a player-dependent positive constant Ω_i,*

$$\mathrm{Reg}_i(t) := \max_{y\in\mathcal{X}_i} \frac{1}{t}\int_0^t \left[u_i(y, x_{-i}(s)) - u_i(x(s))\right] ds \leq \frac{\Omega_i}{t}.$$

The proof mirrors closely more general arguments by Mertikopoulos, Papadimitriou, and Piliouras (2018).

6 Simulations

The goal of this section is to experimentally verify some of the key results, and to highlight other empirically observed properties outside the established theoretical results.[1]

Theorem 4.1 states that any population/environment dynamics which can be captured via a *rescaled zero-sum game* (no matter the complexity of such a description) exhibit a type of *cyclic behavior* known as Poincaré recurrence. Indeed, the trajectories shown in Figure 1 from the time-evolving generalized RPS game of Section 3 are cyclic in nature. Specifically, Figure 1c shows the coevolution of the system for a fixed initial condition. We plot the joint trajectory of the first two strategies for both the population y and environment w, which creates a 4D space where the color legend acts as the final dimension. The simulation demonstrates that as the initial conditions move closer to the interior equilibrium, the trajectories themselves remain bounded within a smaller region around the equilibrium, which confirms the bounded regret property of the dynamics from Proposition 5.1.

Lemma 4.2 shows that for any *rescaled zero-sum game* there is a constant of motion, namely $\Phi(t)$. It is easy to see from the definition of $\Phi(t)$ that a weighted sum of KL-divergences between the strategy vectors produced by replicator dynamics and an interior Nash Equilibrium is also a constant of motion. We simulated an extension to the game depicted in Figure 2 in which many 'butterfly' clusters are joined in a toroid shape. Figure 3 depicts our claim: although each agent specific divergence term $\eta_i \mathrm{KL}(x_i^* || x_i(t))$ fluctuates, the weighted sum $\sum_{i\in V} \eta_i \mathrm{KL}(x_i^* || x_i(t))$ is constant.

To generate Figure 4, we scale-up the game structure from Mai et al. (2018) to 64 nodes. This is a relatively dense graph, where the initial condition of each player informs the RGB value of a corresponding pixel on a grid. If the system exhibits Poincaré recurrence, we should eventually see

[1] Code is available at github.com/ryanndelion/egt-squared

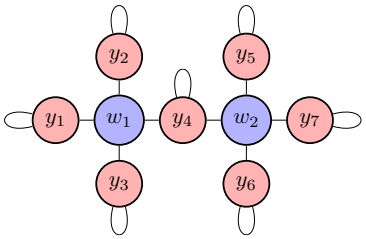


Figure 2: Two clusters of nodes that join together to form a 'butterfly' structure. Self-loops represent RPS self-play games, while edges between nodes represent $(I, -I)$. The *red* nodes denote a population of species, while the *blue* nodes stand for an environment.

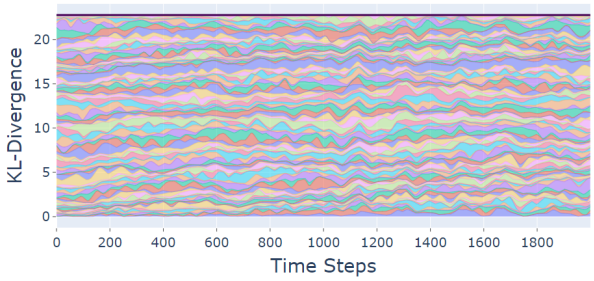


Figure 3: Weighted KL divergence for 25 cluster (100 player) time-evolving zero-sum game.

similar patterns emerge as the pixels change color over time (i.e., as their corresponding strategies evolve). In general, an upper bound on the expected time to see recurrence in such a system is exponential in the number of agents. As observed in Figure 4, the system returns near the initial image in the first several hundred iterations, but takes more than 100k iterations for a clearer Pikachu to reappear.

7 Discussion

We show that systems in which populations of dynamic agents interact with environments that evolve as a function of the agents themselves can equivalently be modeled as polymatrix games. For the class of rescaled zero-sum games, we prove replicator dynamics are Poincaré recurrent and converge in time-average to equilibrium, while experiments show the complexity of these systems. An interesting direction for future research is the study of games dynamics when the games evolve exogenously, instead of only endogenously.

Moreover, there are several exciting applications where our theory has relevance. Google DeepMind trains populations of artificial intelligence agents against each other and computes win probabilities in heads-up competition resulting in a symmetric constant-sum game (Czarnecki et al. 2020; Balduzzi et al. 2019). Up to a shift by an all 0.5 matrix, these are exactly anti-symmetric self-loop games connecting a population of users (programs) to itself as the programs are trying to out-compete each other. The game always remains (anti)-symmetric, but the payoff entries change as stronger

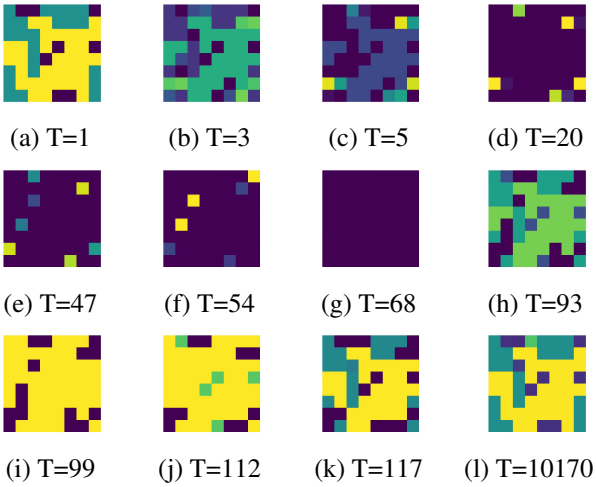


Figure 4: Sequence of Pikachu images showing approximate recurrence in an 8×8 zero-sum polymatrix game, where the changing color of each pixel on the grid represents the strategy of the player over time.

agents replace old agents. While we cannot capture the system fully, we can create the following abstract model of it. The self-loop zero-sum game is the initialization of the system and is equal to the original anti-symmetric empirical zero-sum game. There is another zero-sum game between the population and a meta-agent which simulates the reinforcement policy that chooses which programs get replaced and thus generates a new empirical zero-sum payoff matrix. We can mimic this randomized choice of the policy as a mixed strategy that chooses a convex combination from a large number of possible empirical zero-sum payoff matrices. One of these payoff matrices is the all zero matrix, and the initial strategy of the reinforcement policy chooses that game with high probability at time zero, so that the population is at the start of the process effectively playing just their original empirical game. For such systems, our results provide some theoretical justification for the preservation of diversity and for the satisfying empirical performance.

To conclude, we briefly touch on the connection to progressive training of GANs (Karras et al. 2018). The basic idea is to start the training process with small generator and discriminator networks and, over time, periodically add layers to the networks of higher dimension to grow the resolution of the generated images. This process causes the zero-sum game (between generator and discriminator) to evolve with time. Importantly, as a consequence, the equilibrium is not fixed in the game. We can capture behavior of this process as a time evolving game in our model: the base game matrix P is sparse and of high dimension; as the environment w changes in time the nonzero values in the time-evolving payoff $P(w)$ 'turn on', progressively making the matrix dense. Despite the critical nature of the above artificial intelligence architectures, which are both based on the guided evolution of zero-sum games, no model of them exists in the literature and thus offer exciting possibilities for future work.

Acknowledgments

Stratis Skoulakis gratefully acknowledges NRF 2018 Fellowship NRF-NRFF2018-07. Tanner Fiez acknowledges support from the DoD NDSEG Fellowship. Ryann Sim gratefully acknowledges support from the SUTD President's Graduate Fellowship (SUTD-PGF). Lillian Ratliff is supported by NSF CAREER Award number 1844729 and an Office of Naval Research Young Investigor Award. Georgios Piliouras gratefully acknowledges support from grant PIE-SGP-AI-2020-01, NRF2019-NRF-ANR095 ALIAS grant and NRF 2018 Fellowship NRF-NRFF2018-07.

References

Alongi, J. M.; and Nelson, G. S. 2007. *Recurrence and topology*, volume 85. American Mathematical Society.

Aumann, R. J. 1974. Subjectivity and correlation in randomized strategies. *Journal of mathematical Economics* 1(1): 67–96.

Balduzzi, D.; Garnelo, M.; Bachrach, Y.; Czarnecki, W.; Pérolat, J.; Jaderberg, M.; and Graepel, T. 2019. Open-ended learning in symmetric zero-sum games. In *International Conference on Machine Learning*, volume 97, 434–443.

Bengio, Y.; Louradour, J.; Collobert, R.; and Weston, J. 2009. Curriculum learning. In *International Conference on Machine Learning*, 41–48.

Boone, V.; and Piliouras, G. 2019. From Darwin to Poincaré and von Neumann: Recurrence and Cycles in Evolutionary and Algorithmic Game Theory. In *International Conference on Web and Internet Economics*, 85–99.

Bowles, S.; Choi, J.-K.; and Hopfensitz, A. 2003. The coevolution of individual behaviors and social institutions. *Journal of Theoretical Biology* 223(2): 135–147.

Cai, Y.; Candogan, O.; Daskalakis, C.; and Papadimitriou, C. H. 2016. Zero-Sum Polymatrix Games: A Generalization of Minmax. *Mathematics of Operations Research* 41(2): 648–655.

Cai, Y.; and Daskalakis, C. 2011. On minmax theorems for multiplayer games. In *Symposium of Discrete Algorithms*, 217–234.

Cesa-Bianchi, N.; and Lugosi, G. 2006. *Prediction, learning, and games*. Cambridge university press.

Costa, V.; Lourenço, N.; Correia, J.; and Machado, P. 2019. COEGAN: evaluating the coevolution effect in generative adversarial networks. In *Genetic and Evolutionary Computation Conference*, 374–382.

Costa, V.; Lourenço, N.; Correia, J.; and Machado, P. 2020. Using Skill Rating as Fitness on the Evolution of GANs. In *International Conference on the Applications of Evolutionary Computation*, 562–577. Springer.

Czarnecki, W.; Gidel, G.; Tracey, B.; Tuyls, K.; Omidshafiei, S.; Balduzzi, D.; and Jaderberg, M. 2020. Real World Games Look Like Spinning Tops. In *Advances in Neural Information Processing Systems*.

Garciarena, U.; Santana, R.; and Mendiburu, A. 2018. Evolved GANs for generating Pareto set approximations. In *Genetic and Evolutionary Computation Conference*, 434–441.

Heino, M.; Metz, J. A.; and Kaitala, V. 1998. The enigma of frequency-dependent selection. *Trends in Ecology & Evolution* 13(9): 367–370.

Hofbauer, J.; Sorin, S.; and Viossat, Y. 2009. Time average replicator and best-reply dynamics. *Mathematics of Operations Research* 34(2): 263–269.

Huang, L.; Joseph, A. D.; Nelson, B.; Rubinstein, B. I.; and Tygar, J. D. 2011. Adversarial machine learning. In *ACM workshop on Security and artificial intelligence*, 43–58.

Karras, T.; Aila, T.; Laine, S.; and Lehtinen, J. 2018. Progressive Growing of GANs for Improved Quality, Stability, and Variation. In *International Conference on Learning Representations*.

Lade, S. J.; Tavoni, A.; Levin, S. A.; and Schlüter, M. 2013. Regime shifts in a social-ecological system. *Theoretical Ecology* 6(3): 359–372.

Mai, T.; Mihail, M.; Panageas, I.; Ratcliff, W.; Vazirani, V.; and Yunker, P. 2018. Cycles in Zero-Sum Differential Games and Biological Diversity. In *ACM Conference on Economics and Computation*, 339–350.

Mertikopoulos, P.; Papadimitriou, C.; and Piliouras, G. 2018. Cycles in adversarial regularized learning. In *Symposium of Discrete Algorithms*, 2703–2717.

Miikkulainen, R.; Liang, J.; Meyerson, E.; Rawal, A.; Fink, D.; Francon, O.; Raju, B.; Shahrzad, H.; Navruzyan, A.; Duffy, N.; et al. 2019. Evolving deep neural networks. In *Artificial Intelligence in the Age of Neural Networks and Brain Computing*, 293–312. Elsevier.

Nagarajan, S. G.; Balduzzi, D.; and Piliouras, G. 2020. From Chaos to Order: Symmetry and Conservation Laws in Game Dynamics. In *International Conference on Machine Learning*, 7186–7196.

Nash, J. 1951. Non-cooperative games. *Annals of Mathematics* 286–295.

Nisan, N.; Roughgarden, T.; Tardos, E.; and Vazirani, V. 2007. *Algorithmic Game Theory*. Cambridge university press.

Perolat, J.; Munos, R.; Lespiau, J.-B.; Omidshafiei, S.; Rowland, M.; Ortega, P.; Burch, N.; Anthony, T.; Balduzzi, D.; De Vylder, B.; Piliouras, G.; Lanctot, M.; and Tuyls, K. 2020. From Poincaré Recurrence to Convergence in Imperfect Information Games: Finding Equilibrium via Regularization. *arXiv preprint arXiv:2002.08456* .

Piliouras, G.; Nieto-Granda, C.; Christensen, H. I.; and Shamma, J. S. 2014. Persistent patterns: Multi-agent learning beyond equilibrium and utility. In *International Conference on Autonomous Agents and Multi-Agent Systems*, 181–188.

Piliouras, G.; and Shamma, J. S. 2014. Optimization despite chaos: Convex relaxations to complex limit sets via Poincaré recurrence. In *Symposium of Discrete Algorithms*, 861–873.

Poincaré, H. 1890. Sur le problème des trois corps et les équations de la dynamique. *Acta mathematica* 13(1).

Sandholm, W. H. 2010. *Population games and evolutionary dynamics*. MIT press.

Shoham, Y.; and Leyton-Brown, K. 2008. *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press.

Skoulakis, S.; Fiez, T.; Sim, R.; Piliouras, G.; and Ratliff, L. 2020. Evolutionary Game Theory Squared: Evolving Agents in Endogenously Evolving Zero Sum Games. *arXiv preprint arXiv:2012.08382* .

Stanley, K. O.; and Miikkulainen, R. 2002. Evolving neural networks through augmenting topologies. *Evolutionary Computation* 10(2): 99–127.

Stewart, A. J.; and Plotkin, J. B. 2014. Collapse of cooperation in evolving games. *Proceedings of the National Academy of Sciences* 111(49): 17558–17563.

Tilman, A. R.; Plotkin, J. B.; and Akçay, E. 2020. Evolutionary games with environmental feedbacks. *Nature communications* 11(1): 1–11.

Tilman, A. R.; Watson, J. R.; and Levin, S. 2017. Maintaining cooperation in social-ecological systems. *Journal of Theoretical Biology* 10(2): 155–165.

Wang, C.; Xu, C.; Yao, X.; and Tao, D. 2019. Evolutionary generative adversarial networks. *IEEE Transactions on Evolutionary Computation* 23(6): 921–934.

Weitz, J. S.; Eksin, C.; Paarporn, K.; Brown, S. P.; and Ratcliff, W. C. 2016. An oscillating tragedy of the commons in replicator dynamics with game-environment feedback. *National Academy of Sciences* 113(47).

Wu, Y.; Donahue, J.; Balduzzi, D.; Simonyan, K.; and Lillicrap, T. 2019. LOGAN: Latent Optimisation for Generative Adversarial Networks. *arXiv preprint arXiv:1912.00953* .