Sink equilibria and the attractors of learning in games

Oliver Biggar OLIVER.BIGGAR@COLUMBIA.EDU

Christos Papadimitriou CHRISTOS@COLUMBIA.EDU

Columbia University, New York, USA

Editors: Matus Telgarsky and Jonathan Ullman

Abstract

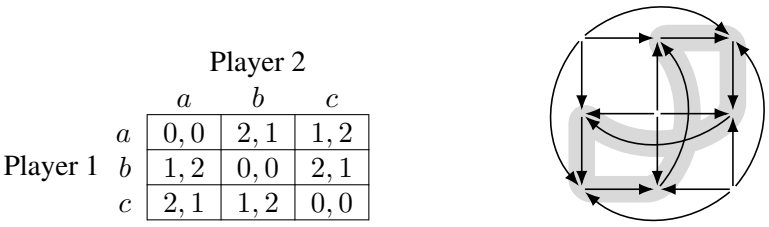
Characterizing the limit behavior—that is, the attractors—of learning dynamics is one of the most fundamental open questions in game theory. In recent work on this front, it was conjectured that the attractors of the replicator dynamic are in one-to-one correspondence with the *sink equilibria* of the game—the sink strongly connected components of a game's preference graph—, and it was established that they do stand in at least one-to-many correspondence with them. Here, we show that the one-to-one conjecture is false. We disprove this conjecture over the course of three theorems: the first disproves a stronger form of the conjecture, while the weaker form is disproved separately in the two-player and N-player ($N > 2$) cases. By showing how the conjecture fails, we lay out the obstacles that lie ahead for characterizing attractors of the replicator, and introduce new ideas with which to tackle them. All three counterexamples derive from an object called a *local source*—a point lying within the sink equilibrium, and yet which is 'locally repelling'; we prove that the absence of local sources is necessary, but not sufficient, for the one-to-one property to be true. We complement this with a sufficient condition: we introduce a local property of a sink equilibrium called *pseudoconvexity*, and establish that when the sink equilibria of a two-player game are pseudoconvex then they precisely define the attractors. Pseudoconvexity generalizes the previous cases—such as zero-sum games and potential games—where this conjecture was known to hold, and reformulates these cases in terms of a simple graph property.

1. Introduction

What are the possible outcomes of a collection of jointly-learning rational agents? This is a fundamental—arguably *the* fundamental—problem in the study of learning in games, with consequences for machine learning, economics and evolutionary biology. The question has received decades of study by mathematicians, economists and computer scientists (see, for example, Zeeman, 1980; Milgrom and Roberts, 1991; Hofbauer, 1996; Sandholm, 2010; Papadimitriou and Piliouras, 2019) and yet it remains broadly unanswered. Excluding some special cases (such as *zero-sum* and *potential games*, and slight generalizations), we do not know how to compute these outcomes (Sandholm, 2010).

One reason for the lack of progress has been the historical focus on Nash equilibria as the outcome of a game (Papadimitriou and Piliouras, 2019; Myerson, 1997). Over time this approach was found to be problematic (Kleinberg et al., 2011)—not only do learning algorithms fail to converge to Nash equilibria in general games (Benaïm et al., 2012; Milionis et al., 2023), but Nash equilibria are also intractable to compute (Daskalakis et al., 2009).

Non-convergence to a Nash equilibria has been traditionally viewed as a limitation of game dynamics. In a departure, Papadimitriou and Piliouras (2019) argued the opposite: non-convergence of learning to Nash equilibria should be viewed as yet another limitation of the Nash equilibrium concept. Instead, the outcomes of learning—whatever they may be—should be the fundamental



		Player 2		
		a	b	c
Player 1	a	$0,0$	$2,1$	$1,2$
	b	$1,2$	$0,0$	$2,1$
	c	$2,1$	$1,2$	$0,0$

(a) (b)

Figure 1: The preference graph of *Shapley's game* (Shapley, 1964), and a typical payoff matrix representation. It has a unique sink equilibrium, which is the highlighted 6-cycle.

objects of interest in game theory. In other words, *the meaning of a game G should be understood as a function μ_G mapping a prior on the space of mixed strategies to a posterior distribution on game outcomes*—where the game outcomes are the attractors of the learning dynamic. To fix a particular well-motivated and widely-used dynamic, we focus on the *replicator dynamic* (Taylor and Jonker, 1978; Smith and Price, 1973)—the continuous-time analog of the multiplicative weights algorithm (Arora et al., 2012) and the flagship dynamic of evolutionary game theory (Hofbauer and Sigmund, 2003; Sandholm, 2010). The replicator dynamic is a most natural learning behavior ("move in the direction of the utility gradient"), which is invariant under the addition of positive constants to the players' utilities, and is qualitatively invariant under positive scaling of the utilities—qualities that are *sine qua non* in economic modeling.

The replicator can have extremely complex behavior (Sato et al., 2002) due to the emergence of chaos and the sensitive dependence on initial conditions, and initially it was not even known whether its attractors are finite in number. However, Papadimitriou and Piliouras (2019) suggested a possible path forward: they hypothesized that the behavior of the replicator (and possibly of far more general dynamics) can be approximated by a simple combinatorial tool called here the *preference graph* of the game (Biggar and Shames, 2023a, 2024, 2025). Specifically, they suggested that the *sink strongly connected components* of this graph—the '*sink equilibria*' (Goemans et al., 2005)—are good proxies for the attractors of the replicator.

This hypothesis is a plausible and hopeful one, first two reasons. First, the preference graph is a natural and insightful graphical representation of the structure of the utilities in a normal-form game (Biggar and Shames, 2023a, 2025) (see Figure 1). Second, the sink equilibria can be viewed as a combinatorial generalization of *pure Nash equilibria* (PNE). Recall that, among the dozens of solution concepts in game theory, only the PNE is largely uncontroversial; its only recognized downside being that not all games have one. Both the sink equilibrium and the mixed Nash equilibrium are generalizations of the PNE which guarantee existence in all games. This observation tempts one to rethink the sink equilibrium as a novel solution concept, a generalization of the PNE in a different direction than Nash's—and possessing the advantages of being both computable in almost linear time in the description of the game, as well as seemingly compatible with learning dynamics.

Unfortunately, proving connections between the sink equilibria and the attractors of the replicator is very difficult. The first significant progress appeared in Biggar and Shames (2023b), where it was established that every attractor of the replicator must contain one or more sink equilibria, and therefore the sink equilibria are in *many-to-one* correspondence with the attractors—implying, for

the first time, that the replicator has finitely many attractors. They also articulated the hypothesis of Papadimitriou and Piliouras (2019) as a conjecture:

Conjecture 1.1 *Each attractor of the replicator contains* exactly one *sink equilibrium, and each sink equilibrium is contained in an attractor.*

This conjecture proposes a strong mathematical correspondence between sink equilibria and attractors. Biggar and Shames also stated a strictly stronger conjecture, hypothesizing one precise form for this correspondence:

Conjecture 1.2 *The attractors of the replicator in any game are exactly the* content *of the sink equilibria.*

The "content" of a set of profiles is the union of all subgames spanned by these profiles (see Definition 1.3). Were these two conjectures proven true, the quest for the outcomes of learning in games would have come to a triumphant conclusion. Both conjectures *are* known to hold in some special cases of sink equilibria (Biggar and Shames, 2023b) such as *attracting subgames* (those where the sink equilibrium profiles are precisely the pure profiles of a subgame). Some other classes of games, such as ordinal potential games (Monderer and Shapley, 1996) and weakly acyclic games (Young, 1993) have only PNEs as sink equilibria, so all these games also satisfy the conjectures. It was subsequently proved that these conjectures also hold for the sink equilibria of zero-sum games (Biggar and Shames, 2024).

Our contributions

We prove that both Conjecture 1.1 and Conjecture 1.2 *fail* in general games. Our results rely on the key concept of a *local source*, which is a mixed profile within a sink equilibrium on the boundary of the strategy space, with the property that nearby trajectories move into the interior of the strategy space, resulting in paths that are not captured by the preference graph. The existence of a local source suffices to disprove Conjecture 1.2; for Conjecture 1.1, an explicit local source argument suffices to disprove it for *three-player games*. Disproving Conjecture 1.1 in the case of two-player games requires a more complex argument. In Section 2.3 we analyze a particular 2×3 game (Figure 4(*c*)) possessing a local source, and establish additionally the existence of a trajectory between two fixed points. We use this game as a gadget, and by composing several copies we are able to arrive at a two-player game with two sink equilibria, but only one replicator attractor, disproving Conjecture 1.1.

Finally, we use the insights gained from the three counterexamples to identify a simple local property called *pseudoconvexity*[1], which suffices to ensure that the content of a sink equilibrium is an attractor (so Conjecture 1.2 holds in this case). Pseudoconvexity is a property of the 2×2 subgames which intersect a sink equilibrium, and so it can be easily established computationally. Roughly speaking, pseudoconvexity requires that the net inflow to the sink equilibrium in the neighborhood of a 2×2 subgame that appears "concave" is actually positive. This property generalizes all the previous classes of games where Conjecture 1.2 was known to hold—*potential games* (Monderer and Shapley, 1996), games where the sink equilibrium is a subgame (Biggar and Shames, 2023b; Ritzberger and Weibull, 1995) and zero-sum games (Biggar and Shames, 2024)—as well as some new classes, like *uniform-weighted cycles* (see Section 3.1 and Figure 1).

1. Not related to pseudoconvex *functions*.

We believe that our main contribution is not the fact that the conjectures are false, but rather the tools and concepts through which we are able to disprove them and the new avenues for progress that they reveal. Understanding of local sources and stability conditions like pseudoconvexity are the conceptual obstacles that lie squarely in the path towards a full characterization of the attractors of the replicator dynamic. We note that while the correspondence between attractors and sink equilibria fails in the form articulated in the two conjectures, a broad correspondence between attractors and the combinatorial structure of the preference graph does seem to hold. In Section 4 we discuss problems inspired and left open by our work, and hypothesize how further developments of the concepts introduced here can build our understanding of learning in games.

Related work

Learning in games has a long and complex history (Fudenberg and Levine, 1998; Cesa-Bianchi and Lugosi, 2006). In this paper we focus on the replicator dynamic (Taylor and Jonker, 1978), which is the continuous-time equivalent (Sorin, 2020) of the *multiplicative weights algorithm* (Arora et al., 2012; Fudenberg and Levine, 1998; Freund and Schapire, 1999), the flagship method in this field. The replicator dynamic arose from the work of Maynard Smith on evolutionary game theory (Smith and Price, 1973), being named and formalized in (Taylor and Jonker, 1978). Since then, it has retained its central role in evolutionary game theory (Sandholm, 2010; Hofbauer and Sigmund, 2003). Finding its attractors is a central goal of the study of the replicator, both in evolutionary game theory (Zeeman, 1980; Sandholm, 2010) and more recently in learning (Papadimitriou and Piliouras, 2016, 2018). The preference graph (and related *best-response graph*) have been sporadically rediscovered in game theory (Goemans et al., 2005; Candogan et al., 2011; Pangallo et al., 2019; Papadimitriou and Piliouras, 2019; Biggar and Shames, 2023a), see Biggar and Shames (2025) for a survey. Sink equilibria originate with the work of Goemans et al. (2005), who used them to study the Price of Anarchy (Koutsoupias and Papadimitriou, 1999). Since the work of Papadimitriou and Piliouras (2019), a line of research has developed relating the replicator and the sink equilibria. Recently, Omidshafiei et al. (2019) used the sink equilibria as an approximation of attractors for the purpose of ranking the strength of game-playing algorithms. Similarly, Omidshafiei et al. (2020) used the preference graph as a tool for representing the space of games for the purposes of learning. Later, Biggar and Shames (2023a,b, 2024) wrote a series of papers on the preference graph and its relationship to the attractors of the replicator dynamic. Another recent work (Hakim et al., 2024) explored the problem of computing the limit distributions over sink equilibria, given a prior over profiles. Our work extends the frontier of this line of investigation.

Preliminaries

We study normal-form games with a finite number N of players and finite strategy sets $S_1, S_2, \ldots, S_N$ for each player (Myerson, 1997). The payoffs in the game are defined by a utility function $u : \prod_{i=1}^N S_i \to \mathbb{R}^N$. $Z := \prod_{i=1}^N S_i$ is the set of *strategy profiles* of the game. We often treat the strategy sets S_i implicitly, and define a game simply by the function u. We use p_{-i} to denote an assignment of strategies to all players other than i. A *subgame* is the game resulting from restricting each player to a subset of their strategies.

A *mixed strategy* is a distribution over a player's pure strategies, and a *mixed profile* is an assignment of a mixed strategy to each player. We often refer to strategies as *pure strategies* and profiles as *pure profiles* to distinguish them from the mixed kinds. For a mixed profile x, we write x^i for the

distribution over player i's strategies, and x_s^i for the s-entry of player i's distribution, where $s \in S_i$. Similar to p_{-i}, we use x_{-i} to denote an assignment of mixed strategies to all players other than i. The *strategy space* of the game is the set of mixed profiles, which is given by $\prod_{i=1}^{N} \Delta_{|S_i|}$ where $\Delta_{|S_i|}$ are the simplices in $\mathbb{R}^{|S_i|}$. We denote this product by X, the mixed analog of Z. The utility function can be naturally extended to mixed profiles, by taking the expectation over strategies. We denote the utility of a mixed profile x by $\mathbb{U}$. Given a subgame y, we write X_y (resp. Z_y) to denote the set of mixed (resp. pure) profiles in y, where (in a slight abuse of notation) we treat X_y as a subset of X.

A *Nash equilibrium* of a game is a mixed profile where no player can increase their expected payoff by a unilateral deviation of strategy. More precisely, a Nash equilibrium x is a point where for each player i and strategy $s \in S_i$, $\mathbb{U}_i(x) \geq \mathbb{U}_i(s; x_{-i})$. A Nash equilibrium is *quasi-strict* if the payoff for deviating to strategies outside the support of the equilibrium is *strictly worse*. That is, if $\hat{x}$ is a Nash equilibrium, it is quasi-strict if $\hat{x}_s^i = 0$ implies that $\mathbb{U}_i(\hat{x}) > \mathbb{U}_i(s; \hat{x}_{-i})$.

Definition 1.3 *(Content of a set of pure profiles, Biggar and Shames (2023b).) Let H be a set of pure profiles in a game. The content of H, denoted* $\mathrm{content}(H)$*, is the set of all mixed strategy profiles x where all pure profiles in the support of x are in H. Equivalently, it is the union $\bigcup_y X_y$ for all subgames y where $Z_y \subseteq H$.*

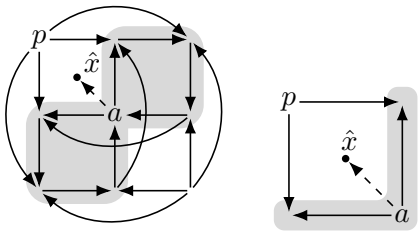
The *replicator dynamic* is a continuous-time dynamical system (Sandholm, 2010; Hofbauer and Sigmund, 2003) in X defined as the solution of the following ordinary differential equation:

$$\dot{x}_s^i = x_s^i \left(\mathbb{U}_i(s; x_{-i}) - \sum_{r \in S_i} x_r^i \mathbb{U}_i(r; x_{-i}) \right)$$

where x is a mixed profile, i a player and s a strategy. It is known to be the limit of the multiplicative weights update algorithm (Arora et al., 2012) when the time step goes to zero. The solution to this equation defines a *flow* (Alongi and Nelson, 2007) $\phi : X \times \mathbb{R} \to X$, which is a continuous group action of the reals on X. Informally, the flow $\phi(t, x)$ (commonly written $\phi^t(x)$) maps $x \in X$ to the point reached after time $t \in \mathbb{R}$. An *orbit* or *trajectory* of a point x_0 is the set $\{\phi^t(x_0) : t \in \mathbb{R}\}$.

A central notion in dynamical systems is the concept of an attractor (Strogatz, 2018). First, fix a dynamic. An *attracting set* under that dynamic is a set S of points with these two properties: (1) there is a neighborhood $U \supset S, U \neq S$ that is *invariant* under the dynamic (if the dynamic starts in U it will stay there), and (2) all points of U converge to S under the dynamic. An attracting set is an *attractor* if it is minimal, that is, no proper subset of it is attracting. We will also need the notion of the *forward* and *backward* limit sets. Given a point x, its *(forward) limit set* $\omega(x)$ under a given dynamic is the set of accumulation points of the orbit starting at x. Formally, a point y is an ω-limit point of x_0 if there exists a sequence $t_n \in \mathbb{R}$ with $t_n \to \infty$ as $n \to \infty$ with $y = \lim_{n\to\infty} \phi^{t_n}(x_0)$. The forward limit set $\omega(x)$ is the set of all ω-limit points. Similarly, the *backward limit set* $\alpha(x)$ of x is the set of all α-limit points, defined analogously as those y where there exists a sequence $t_n \in \mathbb{R}$ with $t_n \to \infty$ as $n \to \infty$ with $y = \lim_{n\to\infty} \phi^{-t_n}(x_0)$ (Alongi and Nelson, 2007).

The *preference graph* of a game (Biggar and Shames, 2023a) is a directed graph whose nodes are the profiles of the game. Two profiles are *i-comparable* if they differ in the strategy of player i only, and they are *comparable* if they are i-comparable for some i. There is an arc between two profiles if they are comparable, and the arc is directed towards the profile where that player receives



(a) (b)

		P_2		
		a	b	c
P_1	a	$0,0$	$2,1$	$1,2$
	b	$1,2$	$1,1$	$2,0$
	c	$2,1$	$0,2$	$0,0$

(c)

Figure 2: A preference graph (Fig. 2(*a*)) whose sink equilibrium (highlighted in gray) has a local source a (Def. 2.1). The point $\hat{x}$ represents the interior fixed point of the 2×2 subgame in the top left, shown separately in Fig. 2(*b*). The presence of the local source at a implies that any replicator attractor of the game which contains a must also contain $\hat{x}$, disproving Conjecture 1.2. Fig. 2(*c*) shows a payoff matrix generating this preference graph.

higher payoff. The arcs in the preference graph can be given non-negative weights representing the difference in utility (Biggar and Shames, 2023a, 2025), which gives us the *weighted preference graph*. Specifically, if p and q are i-comparable, then the arc $p \longrightarrow q$ is weighted by $u_i(q) - u_i(p) \geq 0$, which we write as $W_{q,p} := u_i(q) - u_i(p)$. Finally, the *sink equilibria* of a game are the sink strongly connected components of the preference graph of the game. For a recent summary of results related to preference graphs see Biggar and Shames (2025).

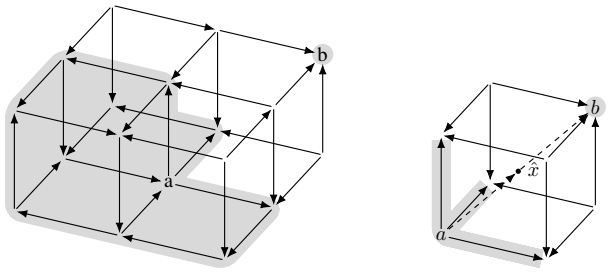
Missing proofs can be found in the appendix.

2. Refuting the Conjectures

Our starting point is the result of Biggar and Shames (2023b) that every attractor of the replicator dynamic contains some sink equilibrium, implying that the attractors of the replicator dynamic are finite in number. Moreover, this result is not too specific to the replicator, as it relies only on two of the dynamic's key properties: *volume conservation* and *subgame-independence* (Hofbauer and Sigmund, 1998; Sandholm, 2010). Volume conservation prohibits any asymptotically stable set from being in the interior of the strategy space. Subgame-independence asserts that (1) each subgame is invariant under the dynamic and (2) the dynamic in a subgame is unaffected by strategies outside the subgame. We note that these properties extend to a broad range of variants of the replicator and more complex dynamics, see for example (Vlatakis-Gkaragkounis et al., 2020).

2.1. Local sources and Conjecture 1.2

All three counterexamples are based on a feature a sink equilibrium may or may not contain, called a *local source*. Consider the game in Figure 2. This game has a unique sink equilibrium H (highlighted in gray), which has an interesting property: despite being *globally* the unique sink connected component of the game, it contains a profile, namely a, which is a *source* in the upper left subgame (Fig. 2(*b*)). Such profiles are called *local sources* of the sink equilibrium, a concept that is a basic ingredient of all three counterexamples.



(*a*) (*b*)

Figure 3: A 3-player counterexample. We show that a replicator trajectory exists from a to $\hat{x}$ (a fully-mixed Nash equilibrium of the subgame in 3(*b*)) and also from $\hat{x}$ to b, implying that any attractor containing a must also contain b.

Definition 2.1 (Local sources) *Let $x \in X$ be a mixed profile, H a sink equilibrium, and $Y \subseteq X$ a subgame. Then x is a* local source of H in Y *if, (1) $x \in \text{content}(H) \cap Y$, (2) $Y \not\subseteq \text{content}(H)$ and (3) x is a quasi-strict Nash equilibrium of the negated game $-u$ restricted to Y.*

In English, this definition can be read as: a local source x of H in Y is a mixed profile in Y which is contained in the sink equilibrium $\text{content}(H)$, yet 'locally' (in the subgame Y) looks like a *source*, in the sense that all players can strictly improve their payoff by deviating to any currently unplayed strategy in Y. To illustrate, consider Figure 2. Let Y be the 2×2 subgame in the upper left, shown separately in Fig. 2(*b*). The sink equilibrium H intersects Y, but $Y \not\subseteq \text{content}(H)$. The profile a lies in $\text{content}(H) \cap Y$ and is a source there, as it is a PNE of this 2×2 subgame if the utility is negated. Hence a is a local source of H in Y. The second requirement ($Y \not\subseteq \text{content}(H)$) is important to eliminate trivial cases; without it, any pure profile in a sink equilibrium that is not a PNE would be a local source, because it is a source in some 1×2 subgame defined by an outgoing arc of the preference graph.

Armed with the concept of a local source, we return to Conjecture 1.2. In Figure 2, the node a is a local source of H in the subgame Y (Fig. 2(*b*)). This subgame is a 2×2 coordination game, where we know a trajectory exists[2] from each source (in this case a) to the Nash equilibrium in the interior of this subgame ($\hat{x}$) (Hofbauer and Sigmund, 1998). However $\hat{x}$ is not in the content of the sink equilibrium, because its support includes the profile p. The existence of this trajectory implies that any attracting set containing a must also contain $\hat{x}$, contradicting Conjecture 1.2. This example is one instance of a more general fact: any sink equilibrium possessing a local source causes Conjecture 1.2 to fail. We prove this in the appendix.

Lemma 2.2 *If H is a sink equilibrium, and H has a local source, then $\text{content}(H)$ is not an attractor.*

2.2. Games with (at least) three players

Consider a game possessing the preference graph in Figure 3(*a*). This game has three players[3] with 3, 3 and 2 strategies respectively. Its preference graph has two sink equilibria H_a and H_b, which we have each highlighted in gray. We have named two nodes a and b, the former in H_a and the latter in H_b. The critical features of the example lie in one $2 \times 2 \times 2$ subgame, which we isolate in Figure 3(*b*). The remainder of the graph serves to ensure that H_a is a sink equilibrium of the game.

In this subgame, node a is a source and node b is a sink. That is, a is again a local source of H_a! Because b is in a different sink equilibrium to a, there are necessarily no paths from a to b in this subgraph. However, it is not possible for an attractor to contain a and not b.

Lemma 2.3 *In Figure 3, any attracting set of the replicator containing a must contain b.*

Proof We will focus on the $2 \times 2 \times 2$ subgame in Figure 3(*b*) (recall that the replicator is subgame-independent). We define the payoffs in this subgame as one for all players in every sink profile, and zero for all players in every source profile. Each player has exactly two pure strategies, and we will represent their mixed strategies by the variables x_1, x_2, x_3. Expressed in these variables, we assume w.l.o.g. that $a = (0, 0, 0)$ and $b = (1, 1, 1)$. This subgame also contains a Nash equilibrium at $\hat{x} = (0.5, 0.5, 0.5)$, which is the only fixed point of the replicator in the interior of the strategy space. Because of the unit payoffs, the replicator equation is given by:

$$\dot{x}_1 = x_1(1 - x_1)\left((1 - x_2)(1 - x_3) - x_2(1 - x_3) - (1 - x_2)x_3 + x_2x_3\right)$$

The equations for $\dot{x}_2$ and $\dot{x}_3$ follow by symmetry. Next, consider the one-dimensional diagonal subspace which contains a, b and $\hat{x}$, defined by $x_1 = x_2 = x_3$. Because of symmetry, this subspace is closed—if we start in this subspace, we remain there. Hence, we can express the replicator on this subspace by a one-dimensional dynamical system, with $w = x_1 = x_2 = x_3$:

$$\dot{w} = w(1 - w)\left((1 - w)^2 - w(1 - w) - (1 - w)w + w^2\right)$$

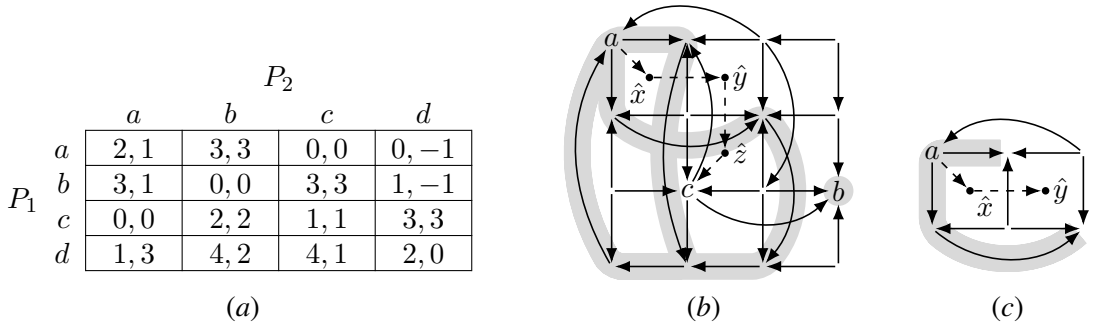
This factorizes to $\dot{w} = w(1 - w)(2w - 1)^2$. This equation is always non-negative, with fixed points at 0 (a), 0.5 ($\hat{x}$) and 1 (b). Thus there is a trajectory from any neighborhood of a to $\hat{x}$, and similarly there is a trajectory from any neighborhood of $\hat{x}$ to b. Any attracting set containing a must also contain $\hat{x}$, and hence also b. ■

2.3. Games with two players

The previous technique does not work in two-player games. The reason is that every two-player game containing a source and a sink necessarily has a path in the preference graph from the source to the sink. In the example above, our construction used a subgame (Figure 3(*b*)) containing both a source (a) and sink (b) but with no path between them.

It turns out that the conjecture still fails in two-player games, though the argument is more subtle. We will use the graph in Figure 4. Like before, we have a game containing two nodes a and

2. Note that, when we say a trajectory exists between two fixed points x and y, we mean there is a *heteroclinic orbit* between them—a trajectory where $\alpha(z) = \{x\}$ and $\omega(z) = \{y\}$ for some point z.
3. Note that this counterexample also serves as a counterexample for N-player games with $N \geq 3$, because we can embed a copy of this game in a game with more players.



P_1 \ P_2	a	b	c	d
a	$2,1$	$3,3$	$0,0$	$0,-1$
b	$3,1$	$0,0$	$3,3$	$1,-1$
c	$0,0$	$2,2$	$1,1$	$3,3$
d	$1,3$	$4,2$	$4,1$	$2,0$

(*a*) (*b*) (*c*)

Figure 4: A two-player counterexample (Fig. 4(*a*)) and its preference graph (Fig. 4(*b*)). We show that replicator trajectories exist from a to $\hat{x}$ to $\hat{y}$ to c to b, meaning that b must be in any attractor containing a.

b, where there are no paths from a to b in the graph. The node a is contained in a sink equilibrium H_a and b is contained in a different sink equilibrium H_b. We have highlighted these in gray. For clarity we have omitted some arcs from the graph, where they are implied by the fact that H_a and H_b are sink equilibria.

Despite the apparent complexity, the key step can be reduced to reasoning about local sources in a much smaller subgame, shown in Figure 4(*c*). The points $\hat{x}$, $\hat{y}$ and $\hat{z}$ are fixed points of the replicator dynamic in 2×2 subgames—that is, they are Nash equilibria in their respective subgames. First, consider the 2×2 subgame containing a and $\hat{x}$. The profile a is a local source of this subgame, and just as we argued in the previous subsection, there is a trajectory from a to $\hat{x}$, implying that any attracting set contain a must contain $\hat{x}$. We will now show that, for some choices of arc weights, $\hat{x}$ is itself a local source of the 2×3 subgame depicted in Fig. 4(*c*).

Lemma 2.4 *Let u be a (generic) 2×3 game whose preference graph is isomorphic to that in Figure 4(*c*). There exist two fixed points $\hat{x}$ and $\hat{y}$ whose supports are both 2×2, and exactly one of these is a Nash equilibrium. There is a trajectory between these points, beginning at the non-Nash fixed point and ending at the Nash equilibrium.*

Using this lemma, we return to Figure 4. By choosing appropriate payoffs (such as those in Fig. 4(*a*)), we can make $\hat{y}$ be a Nash equilibrium in the 2×3 subgame containing $\hat{x}$ and $\hat{y}$, and make $\hat{z}$ a Nash equilibrium in the 2×3 subgame containing $\hat{y}$ and $\hat{z}$. By Lemma 2.4, there is a trajectory from $\hat{x}$ to $\hat{y}$ and similarly there is a trajectory from $\hat{y}$ to $\hat{z}$. By the same argument as the case of a and $\hat{x}$, there is a trajectory from $\hat{z}$ to c. Finally, we complete the argument by observing that these trajectories form a sequence of *heteroclinic orbits* (see a previous footnote) from a to $\hat{x}$ to $\hat{y}$ to $\hat{z}$ to c and finally to b. This implies that any attracting set which contains a must necessarily contain b. Each attractor contains at least one sink equilibrium, and distinct attractors are disjoint. Because any attractor containing H_a must also contain H_b, we conclude that this game has only a single attractor, despite having two sink equilibria.

3. Pseudoconvex sink equilibria are attractors

Our understanding of stability under the replicator dynamic is becoming clearer. When the sink equilibrium has a very simple structure, such as when it is exactly the profiles of a subgame (Biggar

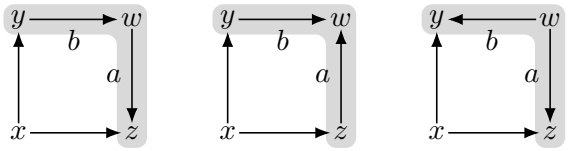


Figure 5: A *cavity* (Def. 3.1) of a sink equilibrium H, where $x \notin H$ and $w, y, z \in H$. Because $x \notin H$, the arcs at x are necessarily directed towards y and z respectively. Up to symmetry, there are three cases, shown in Figs. 5(*a*), 5(*b*) and 5(*c*).

and Shames, 2023b), then its content is an attractor. PNEs are a special case of this. These sinks have no local sources. Sink equilibria of zero-sum games can have non-trivial structure, albeit subject to significant constraints, and they too define attractors (Biggar and Shames, 2024). On the other hand, the counterexamples above demonstrate that the presence of local sources can cause some sink equilibria to not be attractors. What special properties of the sink equilibria in these games prevent the occurrence of local sources, and ensure stability? To better understand this, we will examine the structure of their 2×2 subgames, which will lead us to the concept of *pseudoconvexity*.

3.1. Understanding pseudoconvexity

In a 2×2 subgame, the presence of a local source requires that exactly three of the profiles are within the sink equilibrium.

Definition 3.1 *Let Y be a 2×2 subgame of a game. If* exactly three *of the profiles in Y are contained in a sink equilibrium H, we call this subgame a* cavity *of H.*

Observe that a sink equilibrium H is a subgame *if and only if* it has no cavities. Equivalently, H is a subgame if and only if $\mathrm{content}(H)$ is convex. Cavities are 2×2 subgames where the sink equilibrium is "locally concave." It is not hard to see that the absence of cavities is sufficient to guarantee that $\mathrm{content}(H)$ is an attractor, but it is a very strong requirement, because it implies the sink equilibrium is a subgame. Can a weaker requirement suffice? Examining Figure 5, we note that in case (c) the sink equilibrium contains a local source, and so the sink cannot be an attractor (Lemma 2.2). In case (b), by contrast, the profile w is a local sink—all points converge uniformly to the sink equilibrium. Finally, in case (a), all points approach the sink equilibrium but may or may not do so uniformly, depending on the relative sizes of b and a.

Definition 3.2 *Let $w = (\alpha, \beta)$ be a profile in a sink equilibrium H, and $x = (\gamma, \delta)$ be another profile outside H such that the subgame $Y = \{\alpha, \gamma\} \times \{\beta, \delta\}$ is a cavity of H. We say Y is* pseudoconvex *if $W_{(\gamma,\beta),(\alpha,\beta)} + W_{(\alpha,\delta),(\alpha,\beta)} \leq 0$.*

That is, pseudoconvexity means that the concavity of Y is 'not too severe'—the sum of weights of the arcs into the profile w is positive. Using the language of Figure 5, pseudoconvexity implies that Y is either of type 5(*b*), or it is of type 5(*a*) *and* $a \leq b$. This prevents local sources (type 5(*c*)) in 2×2 subgames. However, quite remarkably, this is also sufficient to guarantee that the entirety of $\mathrm{content}(H)$ is an attractor (Theorem 3.6 below).

Definition 3.3 (Pseudoconvex sink equilibria) *A sink equilibrium is pseudoconvex if every cavity is pseudoconvex, and a game is pseudoconvex if every sink equilibrium is pseudoconvex.*

Note that because pseudoconvexity depends only on 2×2 subgames, there is a very natural polynomial-time algorithm for checking pseudoconvexity of a sink equilibrium. One must simply examine each cavity and check it satisfies Definition 3.2.

As we mentioned earlier, when a sink equilibrium is a subgame it is trivially pseudoconvex, because there are no cavities. It is also true—though much less obvious, see the appendix—that the sink equilibria of zero-sum games are pseudoconvex.

Lemma 3.4 *Two-player zero-sum games are pseudoconvex.*

Hence for all the classes of sink equilibria where we know the content is an attractor, these sink equilibria are pseudoconvex. It is natural to therefore conjecture that pseudoconvexity is *sufficient* for the content to be an attractor. This turns out to be true for two-player games, as we show in Theorem 3.6. However, pseudoconvexity also encompasses games which are quite distinct from these cases. As an example, consider the famous example of *Shapley's game*, from Shapley (1964). A payoff matrix representation is given in Figure 1(*a*), with its preference graph in Figure 1(*b*). Shapley demonstrated that *fictitious play* (FP) (Brown, 1949; Robinson, 1951) converged to a 6-cycle on the boundary—which is exactly the content of the unique sink equilibrium. This cycle is obviously far from being a subgame, and further it is also far from being the sink equilibrium of a zero-sum game (see Theorem 4.10 of Biggar and Shames (2023a), which proves that such sink equilibria are '*near-subgames*'). However, it is pseudoconvex! First, observe that by Fig. 1(*a*) each arc on the 6-cycle has the same weight, which is one. Because the sink equilibrium is exactly a cycle, each cavity (Def. 3.1) is of the form in Figure 5(*a*). Because the two weights a and b in Fig. 5(*a*) are equal (both one), this sink equilibrium satisfies Definition 3.2. More generally, the whole class of *uniformly weighted cycles* is pseudoconvex—these are sink equilibrium that are simple cycles where every arc on the cycle has the same weight. Being pseudoconvex, all such cycles turn out to be replicator attractors.

3.2. Stability of pseudoconvex sink equilibria

To prove Theorem 3.6, we will need some new concepts. A key idea is to shift perspective from the mixed strategy space into the *correlated space* of distributions over profiles. We show that the replicator can be given a simple presentation in this space, in terms of a matrix we call the *product matrix* of the game. This construction derives from Biggar and Shames (2024), who introduced a similar idea in two-player zero-sum games—here we generalize it to n-player general-sum games. While straightforward, this transformation is critical for our result and we believe it is a useful fact for analyzing the replicator in other contexts as well.

Lemma 3.5 *Let u be a N-player $m_1 \times m_2 \times \cdots \times m_N$ game. The* product matrix *of u is the following matrix $M \in \mathbb{R}^{(\prod_i m_i)\times(\prod_i m_i)}$, indexed by profiles p and q in Z:*

$$M_{q,p} = \sum_{i=1}^{N} \left(u_i(q_i; p_{-i}) - u_i(p)\right) = \sum_{i=1}^{N} W_{(q_i;p_{-i}),p} \tag{1}$$

Next, let $p = (s_1, s_2, \ldots, s_N)$ be a pure profile. Given a mixed profile x, the product distribution $z_p := \prod_{i=1}^{N} x^i_{s_i}$ defines a distribution over profiles. Then, under the replicator dynamic:

$$\dot{z}_p = z_p(Mz)_p$$

In other words, the distribution z induced over profile in Z by a mixed strategy x evolves by a simple formula in terms of the product matrix—simpler even than the original definition of the replicator! Using this, we can prove the main theorem of this section:

Theorem 3.6 *If a sink equilibrium H of a two-player game u is pseudoconvex, then its content is an attractor of the replicator.*

Proof [*Sketch—full proof in appendix*] We prove this using a *Lyapunov argument* on the cumulative total mass distributed over the profiles in H. In correlated space, this is simply the sum $z_H := \sum_{h \in H} z_h$ where z is the mass on a single profile, as in Lemma 3.5. The total z_H is equal to one exactly when a point lies in $\mathrm{content}(H)$. We show that this is increasing near $\mathrm{content}(H)$ (where z_H is sufficiently close to one). Using linearity and Lemma 3.5, we can obtain an expression for $\dot{z}_H$ in terms of the product matrix M, where each term corresponds to an arc in the preference graph with at least one endpoint in H. By grouping terms into 2×2 subgames, we end up with a case-by-case analysis on the preference graphs of those subgames. With z_H close to one, we show that the only potentially negative term is caused by a cavity that is not pseudoconvex! Applying pseudoconvexity thereby guarantees the result. ■

We believe that Theorem 3.6 is a major step towards our ultimate goal: a polynomial-time algorithm which, given a two-player game in normal form, identifies its attractors. When the sink equilibria are pseudoconvex, we know the attractors: they are the content of these sink equilibria. But if a local source exists in a sink component, some replicator paths "escape" the component. However finding local sources, or determining if any exist, can be very hard. The power of this theorem is that pseudoconvexity (a 2×2 property) is sufficient to guarantee, non-constructively, that no such escaping trajectories exist. If the sink equilibrium is not pseudoconvex, the proof of the theorem does not seem to provide guidance on where to look for these escaping trajectories. Analyzing the attractors of two-player games beyond pseudoconvexity is an important open research problem left by this work.

4. Conclusions and open problems

Let us return to our original goal: to understand, and ultimately compute, the possible long-run outcomes of game dynamics by exploiting their relationship to sink equilibria. We have made some significant progress on this problem for the replicator dynamic. For games where the sink equilibria are pseudoconvex, they give a precise and efficiently-computable characterization of the attractors. On the other hand, we have also shown how this simple picture does not apply to all games, as the presence of local sources can lead to attractors that are larger than the sink equilibria, sometimes merging two or more sink equilibria.

This work opens up a number of important open problems.

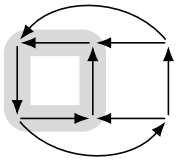


Figure 6: Under the replicator, all interior starting points converge to the highlighted 2×2 subgame. However, this is a strict subset of the unique attractor, which is the whole strategy space (Biggar and Shames, 2025).

Local sources and paths. We know that the presence of local sources can lead to trajectories which escape sink equilibria. Yet local sources too possess a specific graph-theoretic structure—we believe that a broader combinatorial framework generalizing the preference graph could incorporate this case, and thus potentially characterize the ultimate structure of the attractors of a game.

Beyond uniform convergence. The criteria for a set to be an attractor are very strong—in particular they require a neighborhood of the set to exist where all trajectories approach the attractor. However, for the replicator dynamic, there exist cases where this requirement seems too demanding. Figure 6 shows a 2×3 game, where the second column strategy dominates the third. Under the replicator, the third column strategy eventually vanishes from all interior starting points, and so the dynamic converges to the highlighted 2×2 subgame. But the preference graph is strongly connected, so the unique attractor is the whole game. The discrepancy is explained by the fact that the replicator does not converge *uniformly* to this smaller set (Biggar and Shames, 2024, 2025). This lack of uniformity follows as a consequence of subgame-independence. This suggests that stronger prediction of game dynamics could be made if we allowed for weaker notions of convergence.

Algorithmic problems. The ultimate goal of this research program is a polynomial-time algorithm which, given a game in normal form, will output the combinatorial structure of its attractors, and do this for a wide range of learning dynamics. For the case of the replicator dynamic we have made reasonable progress, with pseudoconvexity providing a simple sufficient condition for local stability. However, the possibility of local sources leads to two new problems:

1. We know that the presence of a local source in a sink equilibrium H is sufficient to show that $\mathrm{content}(H)$ is not an attractor. But what of the converse: if a sink equilibrium is not stable, must there exist a local source acting as a 'witness' to this instability? From a computational perspective, given H, can we efficiently verify if $\mathrm{content}(H)$ is an attractor?

2. If a local source does exist, some additional points may need to be added to the sink equilibrium. But which? Is there an iterative procedure for adding the missing points to a sink equilibrium until we find an attractor?

Large games. The limit behavior of learning in *large* multi-player games is of great interest in economics—however, we know that even the most modest algorithmic goals related to sink equilibria are PSPACE-complete for many succinct representations of games (Fabrikant and Papadimitriou, 2008). It would be very interesting to make progress in characterizing and computing the attractor structure of a game for the case of *symmetric* multi-player games.

References

John M Alongi and Gail Susan Nelson. *Recurrence and topology*, volume 85. American Mathematical Soc., 2007.

Sanjeev Arora, Elad Hazan, and Satyen Kale. The multiplicative weights update method: a meta-algorithm and applications. *Theory of computing*, 8(1):121–164, 2012. ISSN 1557-2862. Publisher: Theory of Computing Exchange.

Michel Benaïm, Josef Hofbauer, and Sylvain Sorin. Perturbations of set-valued dynamical systems, with applications to game theory. *Dynamic Games and Applications*, 2:195–205, 2012. ISSN 2153-0785. Publisher: Springer.

Oliver Biggar and Iman Shames. The graph structure of two-player games. *Scientific Reports*, 13 (1):1833, February 2023a. ISSN 2045-2322. doi: 10.1038/s41598-023-28627-8. URL https://www.nature.com/articles/s41598-023-28627-8. Publisher: Nature Publishing Group.

Oliver Biggar and Iman Shames. The Replicator Dynamic, Chain Components and the Response Graph. In *Proceedings of The 34th International Conference on Algorithmic Learning Theory*, pages 237–258. PMLR, February 2023b. URL https://proceedings.mlr.press/v201/biggar23a.html. ISSN: 2640-3498.

Oliver Biggar and Iman Shames. The Attractor of the Replicator Dynamic in Zero-Sum Games. In Claire Vernade and Daniel Hsu, editors, *Proceedings of The 35th International Conference on Algorithmic Learning Theory*, volume 237 of *Proceedings of Machine Learning Research*, pages 161–178. PMLR, 25–28 Feb 2024. URL https://proceedings.mlr.press/v237/biggar24a.html.

Oliver Biggar and Iman Shames. Preference graphs: a combinatorial tool for game theory. *arXiv preprint arXiv:2502.03546*, 2025.

George W Brown. *Some notes on computation of games solutions*. Rand Corporation, 1949.

Ozan Candogan, Ishai Menache, Asuman Ozdaglar, and Pablo A. Parrilo. Flows and Decompositions of Games: Harmonic and Potential Games. *Mathematics of Operations Research*, 36(3): 474–503, August 2011. ISSN 0364-765X. doi: 10.1287/moor.1110.0500. URL https://pubsonline.informs.org/doi/abs/10.1287/moor.1110.0500. Publisher: INFORMS.

Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006.

Constantinos Daskalakis, Paul W. Goldberg, and Christos H. Papadimitriou. The complexity of computing a Nash equilibrium. *Communications of the ACM*, 52(2):89–97, February 2009. ISSN 0001-0782. doi: 10.1145/1461928.1461951. URL https://doi.org/10.1145/1461928.1461951.

Alex Fabrikant and Christos H Papadimitriou. The complexity of game dynamics: Bgp oscillations, sink equilibria, and beyond. In *SODA*, volume 8, pages 844–853. Citeseer, 2008.

Yoav Freund and Robert E Schapire. Adaptive game playing using multiplicative weights. *Games and Economic Behavior*, 29(1-2):79–103, 1999. ISSN 0899-8256. Publisher: Elsevier.

Drew Fudenberg and David K Levine. *The theory of learning in games*, volume 2. MIT press, 1998.

M. Goemans, Vahab Mirrokni, and A. Vetta. Sink equilibria and convergence. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS'05)*, pages 142–151, October 2005. doi: 10.1109/SFCS.2005.68. ISSN: 0272-5428.

Rashida Hakim, Jason Milionis, Christos Papadimitriou, and Georgios Piliouras. Swim till you sink: Computing the limit of a game. In *International Symposium on Algorithmic Game Theory*, pages 205–222. Springer, 2024.

Josef Hofbauer. Evolutionary dynamics for bimatrix games: A Hamiltonian system? *Journal of mathematical biology*, 34:675–688, 1996. ISSN 0303-6812. Publisher: Springer.

Josef Hofbauer and Karl Sigmund. *Evolutionary games and population dynamics*. Cambridge university press, 1998.

Josef Hofbauer and Karl Sigmund. Evolutionary game dynamics. *Bulletin of the American Mathematical Society*, 40(4):479–519, 2003. ISSN 0273-0979, 1088-9485. doi: 10.1090/S0273-0979-03-00988-1. URL https://www.ams.org/bull/2003-40-04/S0273-0979-03-00988-1/.

Josef Hofbauer, Sylvain Sorin, and Yannick Viossat. Time Average Replicator and Best-Reply Dynamics. *Mathematics of Operations Research*, 34(2):263–269, May 2009. ISSN 0364-765X. doi: 10.1287/moor.1080.0359. URL https://pubsonline.informs.org/doi/abs/10.1287/moor.1080.0359. Publisher: INFORMS.

Robert D Kleinberg, Katrina Ligett, Georgios Piliouras, and Éva Tardos. Beyond the Nash Equilibrium Barrier. In *ICS*, pages 125–140, 2011.

Elias Koutsoupias and Christos Papadimitriou. Worst-case equilibria. In *Annual symposium on theoretical aspects of computer science*, pages 404–413. Springer, 1999.

Paul Milgrom and John Roberts. Adaptive and sophisticated learning in normal form games. *Games and Economic Behavior*, 3(1):82–100, February 1991. ISSN 0899-8256. doi: 10.1016/0899-8256(91)90006-Z. URL https://www.sciencedirect.com/science/article/pii/089982569190006Z.

Jason Milionis, Christos Papadimitriou, Georgios Piliouras, and Kelly Spendlove. An impossibility theorem in game dynamics. *Proceedings of the National Academy of Sciences*, 120(41): e2305349120, October 2023. doi: 10.1073/pnas.2305349120. URL https://www.pnas.org/doi/abs/10.1073/pnas.2305349120. Publisher: Proceedings of the National Academy of Sciences.

Dov Monderer and Lloyd S. Shapley. Potential Games. *Games and Economic Behavior*, 14(1): 124–143, May 1996. ISSN 0899-8256. doi: 10.1006/game.1996.0044. URL https://www.sciencedirect.com/science/article/pii/S0899825696900445.

Roger B Myerson. *Game theory: analysis of conflict*. Harvard university press, 1997.

Shayegan Omidshafiei, Christos Papadimitriou, Georgios Piliouras, Karl Tuyls, Mark Rowland, Jean-Baptiste Lespiau, Wojciech M. Czarnecki, Marc Lanctot, Julien Perolat, and Remi Munos. α-Rank: Multi-Agent Evaluation by Evolution. *Scientific Reports*, 9(1):9937, July 2019. ISSN 2045-2322. doi: 10.1038/s41598-019-45619-9. URL https://www.nature.com/articles/s41598-019-45619-9.

Shayegan Omidshafiei, Karl Tuyls, Wojciech M. Czarnecki, Francisco C. Santos, Mark Rowland, Jerome Connor, Daniel Hennes, Paul Muller, Julien Pérolat, Bart De Vylder, Audrunas Gruslys, and Rémi Munos. Navigating the landscape of multiplayer games. *Nature Communications*, 11 (1):5603, November 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-19244-4. URL https://www.nature.com/articles/s41467-020-19244-4. Number: 1 Publisher: Nature Publishing Group.

Marco Pangallo, Torsten Heinrich, and J. Doyne Farmer. Best reply structure and equilibrium convergence in generic games. *Science Advances*, 5(2):eaat1328, February 2019. doi: 10.1126/sciadv.aat1328. URL https://www.science.org/doi/full/10.1126/sciadv.aat1328. Publisher: American Association for the Advancement of Science.

Christos Papadimitriou and Georgios Piliouras. From nash equilibria to chain recurrent sets: Solution concepts and topology. In *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*, pages 227–235, 2016.

Christos Papadimitriou and Georgios Piliouras. From nash equilibria to chain recurrent sets: An algorithmic solution concept for game theory. *Entropy*, 20(10):782, 2018. ISSN 1099-4300. Publisher: MDPI.

Christos Papadimitriou and Georgios Piliouras. Game dynamics as the meaning of a game. *ACM SIGecom Exchanges*, 16(2):53–63, 2019. ISSN 1551-9031. Publisher: ACM New York, NY, USA.

Klaus Ritzberger and Jörgen W. Weibull. Evolutionary Selection in Normal-Form Games. *Econometrica*, 63(6):1371–1399, 1995. ISSN 0012-9682. doi: 10.2307/2171774. URL https://www.jstor.org/stable/2171774. Publisher: [Wiley, Econometric Society].

Julia Robinson. An Iterative Method of Solving a Game. *Annals of Mathematics*, 54(2):296–301, 1951. ISSN 0003-486X. doi: 10.2307/1969530. URL https://www.jstor.org/stable/1969530. Publisher: Annals of Mathematics.

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

Yuzuru Sato, Eizo Akiyama, and J. Doyne Farmer. Chaos in learning a simple two-person game. *Proceedings of the National Academy of Sciences*, 99(7):4748–4751, April 2002. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.032086299. URL https://pnas.org/doi/full/10.1073/pnas.032086299.

Lloyd Shapley. Some topics in two-person games. *Advances in game theory*, 52:1–29, 1964.

J Maynard Smith and George R Price. The logic of animal conflict. *Nature*, 246(5427):15–18, 1973.

Sylvain Sorin. Replicator dynamics: Old and new. *Journal of Dynamics & Games*, 7(4), 2020.

Steven H Strogatz. *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. CRC press, 2018.

Peter D. Taylor and Leo B. Jonker. Evolutionary stable strategies and game dynamics. *Mathematical Biosciences*, 40(1):145–156, July 1978. ISSN 0025-5564. doi: 10.1016/0025-5564(78)90077-9. URL https://www.sciencedirect.com/science/article/pii/0025556478900779.

Emmanouil-Vasileios Vlatakis-Gkaragkounis, Lampros Flokas, Thanasis Lianeas, Panayotis Mertikopoulos, and Georgios Piliouras. No-Regret Learning and Mixed Nash Equilibria: They Do Not Mix. In *Advances in Neural Information Processing Systems*, volume 33, pages 1380–1391. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/hash/0ed9422357395a0d4879191c66f4faa2-Abstract.html.

H. Peyton Young. The Evolution of Conventions. *Econometrica*, 61(1):57–84, 1993. ISSN 0012-9682. doi: 10.2307/2951778. URL https://www.jstor.org/stable/2951778. Publisher: [Wiley, Econometric Society].

E. C. Zeeman. Population dynamics from game theory. In Zbigniew Nitecki and Clark Robinson, editors, *Global Theory of Dynamical Systems*, Lecture Notes in Mathematics, pages 471–497, Berlin, Heidelberg, 1980. Springer. ISBN 978-3-540-38312-3. doi: 10.1007/BFb0087009.

Appendix A. Omitted proofs

Proof of Lemma 2.2 Let x be a local source of H in a subgame Y. Because $Y \not\subseteq \operatorname{content}(H)$, we know that $\operatorname{content}(H) \cap \operatorname{int} Y = \emptyset$. Thus we only need to show that at there is at least one point $z \in \operatorname{int} Y$ with $\alpha(z) = \{x\}$. Because x is a quasi-strict Nash equilibrium of $-u$, this is equivalent to showing that, in $-u$, there is a $z \in \operatorname{int} Y$ with $\omega(z) = \{x\}$. We prove this claim in Lemma A.1, by observing that the stable manifold of x in $-u$ must necessarily intersect the interior of X. ■

Lemma A.1 *Let $\hat{x}$ be a quasi-strict Nash equilibrium on the boundary $\operatorname{bd} X$ of the strategy space X. Then there exists a point $z \in \operatorname{int} X$ with $\omega(z) = \{\hat{x}\}$.*

Proof Let Y be the subgame that is the support of $\hat{x}$, with does not include all strategies because $\hat{x}$ lies in the boundary of X. We begin by constructing the Jacobian of the replicator at $\hat{x}$. Let s_i be a strategy that is not in the support of $\hat{x}^i$, so $\hat{x}^i_{s_i} = 0$. Observe that

$$\partial \dot{x}^i_{s_i} / \partial x^i_{s_i} = u_i(s_i; x_{-i}) - u_i(x) + x^i_{s_i} \partial / \partial x^i_{s_i} (u_i(s_i; x_{-i}) - u_i(x))$$

and so at $\hat{x}$, where $\hat{x}^i_{s_i} = 0$, we have $\partial \dot{x}^i_{s_i} / \partial x^i_{s_i} = u_i(s_i; \hat{x}_{-i}) - u_i(\hat{x})$. Similarly, for any other x^k_ℓ,

$$\partial \dot{x}^i_{s_i} / \partial x^k_\ell = x^i_{s_i}(\dots)$$

so $\partial \dot{x}^i_{s_i} / \partial x^k_\ell = 0$ at $\hat{x}$. Hence the strategy s_i column of the Jacobian of the replicator at $\hat{x}$ is equal to $(u_i(s_i; \hat{x}_{-i}) - u_i(\hat{x}))\mathbf{e}_{s_i}$, where $\mathbf{e}_{s_i}$ is the standard basis vector of the s_i column. We conclude

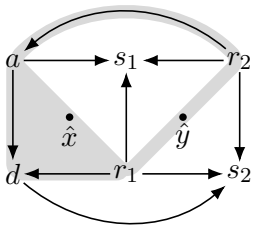


Figure 7: The game from Figure 4(*c*), with the set W highlighted as in the proof of Lemma 2.4.

that $\mathbf{e}_{s_i}$ is an eigenvector of the Jacobian at $\hat{x}$, whose eigenvalue is $u_i(s_i; \hat{x}_{-i}) - u_i(\hat{x})$—the relative payoff of s_i at $\hat{x}$. Because this is a quasi-strict Nash equilibrium, this is strictly negative. These are called the transversal eigenvalues of $\hat{x}$ (Hofbauer and Sigmund, 1998).

We now apply the stable manifold theorem. This tells us that, in the neighborhood of $\hat{x}$, there is an invariant manifold of points which converge to $\hat{x}$, and this manifold is tangent to the space E^s spanned by the eigenvectors with negative eigenvalues. Because all transversal eigenvalues are negative, there is a direction in the stable manifold of $\hat{x}$ which points into the interior of the space. Hence there exists a points on the interior of X which converges to $\hat{x}$. ■

Proof of Lemma 2.4 (*Claim:* generically, exactly one of $\hat{x}$ and $\hat{y}$ is Nash.) Let r be the pure strategy for player 1 which involves playing the top row. Then $\hat{x}_r^1, \hat{y}_r^1 \in (0, 1)$ denote the probability mass on r in $\hat{x}$ and $\hat{y}$ respectively. We will show that $\hat{x}$ is Nash if $\hat{x}_r^1 > \hat{y}_r^1$, $\hat{y}$ is Nash if $\hat{x}_r^1 < \hat{y}_r^1$ and both are Nash in the non-generic case where $\hat{x}_r^1 = \hat{y}_r^1$.

Let c_1, c_2, c_3 be the column strategies, ordered from left to right. Given a fixed mixed strategy $(z, 1-z)$ for the row player, the column player receives a payoff for each c_i that is a linear function of z. At $z = 0$, the column player prefers their strategies in the best-to-worst order c_3, c_1, c_2. At $z = 1$, the column player prefers their strategies in the best-to-worst order c_2, c_1, c_3. At $z = \hat{x}_r^1$, the column player is indifferent between c_1 and c_2 and at $z = \hat{y}_r^1$, they are indifferent between c_2 and c_3. If $\hat{x}_r^1 < \hat{y}_r^1$, then at $z = \hat{x}_r^1$ the column player must prefer c_3 over c_1 and c_2, so $\hat{x}$ is not a Nash equilibrium. By contrast, at $z = \hat{y}_r^1$, the column player prefers both c_2 and c_3 (and hence $\hat{y}^2$) over c_1, so $\hat{y}$ is a Nash equilibrium. The opposite is true when $\hat{x}_r^1 > \hat{y}_r^1$. If $\hat{x}_r^1 = \hat{y}_r^1$, then at this point the column player is indifferent between c_1, c_2 and c_3, resulting in a continuum of Nash equilibria. This completes the claim.

(*Claim:* there exists a point z where $\alpha(z) = \hat{x}$ and $\omega(z) = \hat{y}$.) In Figure 7 we show a more detailed depiction of this game for the purposes of the proof. The game has two sources and two sinks, which we name r_1, r_2, s_1 and s_2 respectively. The remaining two profiles are named a and d. The sinks are PNEs, both of which are attractors of the replicator. Further, because every attractor must contain a sink equilibrium (Biggar and Shames, 2023b), these are the only attractors.

Let $\omega^{-1}(s_1)$ and $\omega^{-1}(s_2)$ be the sets of points z whose limit set $\omega(z)$ is $\{s_1\}$ or $\{s_2\}$ respectively. These are disjoint invariant sets, and they are open by the continuity of the replicator dynamic. Hence there exists a closed set W of points whose limit sets are neither $\{s_1\}$ nor $\{s_2\}$. Note that W is also an invariant set. Because each proper subgame of this game is two-dimensional, we can completely characterize the intersection of W and the boundary $\mathrm{bd}\, X$ of the strategy space by considering each subgame in turn. The result is the set depicted in Figure 7. It contains all fixed points except the sinks s_1 and s_2, as well as the one-dimensional trajectories joining them, and a

two-dimensional region in the subgame spanned by a, $\hat{x}$, r_1 and d, and the trajectories between them.

By Proposition 5.1 of Hofbauer et al. (2009), the limit set of the time average of an interior point z is a closed subset of X that is both invariant under *best-response dynamics*. For $z \in \operatorname{int} X \cap W$, $\omega(z)$ is on the boundary (Hofbauer and Sigmund, 1998), and so must additionally be a subset of $W \cap \operatorname{bd} X$. The only subset of this set that is invariant under best-response dynamics is the Nash equilibrium set $\{\hat{y}\}$. Hence for any $z \in W \cap \operatorname{int} X$, $\omega(z) = \{\hat{y}\}$.

Now we repeat the steps above, but with the role of forward and backward limit sets reversed. Let S be the set of points z in X whose *backward* limit set $\alpha(z)$ is neither $\{r_1\}$ nor $\{r_2\}$. Again we can characterize the set $S \cap \operatorname{bd} X$, which has an analogous structure to W. The flow of the replicator in reverse time is equal to the flow of the replicator on the negated game $-u$, so we can apply the same arguments about ω-limits to α-limits. We conclude, again, that for every $z \in S \cap \operatorname{int} X$, $\alpha(z) = \{\hat{x}\}$.

To complete the argument, we just need to show that $S \cap W \cap \operatorname{int} X$ is not empty. For this, observe that if $z \in \operatorname{bd} X \cap \omega^{-1}(s_1)$, then because $\omega^{-1}(s_1)$ is open there is an $\epsilon > 0$ such that all points within ϵ of z are in $\omega^{-1}(s_1)$. In particular, there are points near z in $\operatorname{int} X \cap \omega^{-1}(s_1)$.

Now consider a neighborhood of $\hat{x}$ in $S \cap \operatorname{int} X$. $\hat{x}$ lies on the boundary of $\omega^{-1}(s_1)$ and $\omega^{-1}(s_2)$, so every neighborhood in S contains points in both of these sets. Similarly, every neighborhood in $S \cap \operatorname{int} X$ must contain points in both of these sets, by the argument in the previous paragraph. Finally, every neighborhood of $\hat{x}$ in $S \cap \operatorname{int} X$ must therefore contain points in the boundary of $\omega^{-1}(s_1)$ and $\omega^{-1}(s_1)$, which is entirely contained in W. Hence $S \cap W \cap \operatorname{int} X$ is nonempty. ■

Proof of Lemma 3.4 Let $w = (\alpha, \beta)$ be a profile in a sink equilibrium H, and $x = (\gamma, \delta)$ be another profile outside H such that the subgame $\{\alpha, \gamma\} \times \{\beta, \delta\}$ is a cavity of H. Because $x \notin H$, there are arcs $(\gamma, \delta) \longrightarrow (\gamma, \beta)$ and $(\gamma, \delta) \longrightarrow (\alpha, \delta)$ and so $W_{(\gamma,\beta),(\gamma,\delta)} > 0$ and $W_{(\alpha,\delta),(\gamma,\delta)} > 0$. Lemma 4.7 of Biggar and Shames (2024) establishes that $W_{(\gamma,\beta),(\gamma,\delta)} + W_{(\alpha,\delta),(\gamma,\delta)} = -(W_{(\gamma,\beta),(\alpha,\beta)} + W_{(\alpha,\delta),(\alpha,\beta)})$, so $W_{(\gamma,\beta),(\alpha,\beta)} + W_{(\alpha,\delta),(\alpha,\beta)} < 0$, thus this cavity is pseudoconvex. ■

Proof of Lemma 3.5 By Lemma A.1 of Biggar and Shames (2023b), the replicator equation can be written

$$\begin{aligned}
\dot{x}_s^i &= x_s^i \sum_{r \in S_i} x_r^i \sum_{p_{-i} \in Z_{-i}} z_{p_{-i}} \left(u_i(s; p_{-i}) - u_i(r; p_{-i})\right) \\
&= x_s^i \sum_{p_{-i} \in Z_{-i}} \sum_{r \in S_i} x_r^i z_{p_{-i}} \left(u_i(s; p_{-i}) - u_i(r; p_{-i})\right) \\
&= x_s^i \sum_{p_{-i} \in Z_{-i}} \sum_{r \in S_i} z_{(r; p_{-i})} \left(u_i(s; p_{-i}) - u_i(r; p_{-i})\right) \\
&= x_s^i \sum_{p \in Z} z_p \left(u_i(s; p_{-i}) - u_i(p)\right)
\end{aligned}$$

Now we observe that for $q = (q_1, q_2, \ldots, q_N)$,

$$\begin{aligned}
\dot{z}_q &= \frac{\mathrm{d}}{\mathrm{d}t}\left(\prod_{i=1}^{N} x^i_{q_i}\right) = z_q \sum_{i=1}^{N} \frac{\dot{x}^i_{q_i}}{x^i_{q_i}} \quad \text{(product rule)} \\
&= z_q \sum_{i=1}^{N} \sum_{p \in Z} z_p \left(u_i(q_i; p_{-i}) - u_i(p)\right) \quad \text{(by above)} \\
&= z_q \sum_{p \in Z} z_p \sum_{i=1}^{N} \left(u_i(q_i; p_{-i}) - u_i(p)\right) = z_q \sum_{p \in Z} z_p M_{q,p} \\
&= z_q (Mz)_q
\end{aligned}$$

■

Proof of Theorem 3.6 Let H be a sink equilibrium, which we assume is pseudoconvex. We will show it is asymptotically stable by a Lyapunov argument, using a similar structure to Theorem 4.3 of Biggar and Shames (2024). First, we define $z_H := \sum_{h \in H} z_h$. That is, z_H is the cumulative total distributed over the profiles in H in the product distribution z. Note that $z_H = 1$ if and only if $x \in \mathrm{content}(H)$. z_H is uniformly continuous, and so to prove that $\mathrm{content}(H)$ is an attractor it is sufficient to show that $\dot{z}_H > 0$ in some neighborhood of $\mathrm{content}(H)$. Now fix some small $1 > \epsilon > 0$. We will assume that $z_H = 1 - \epsilon$, and we will show that for small enough ϵ, $\dot{z}_H > 0$.

From Lemma 3.5 we have that $\dot{z}_H = \sum_{h \in H} z_h (Mz)_h = \sum_{p \in Z} \sum_{h \in H} z_p z_h M_{h,p}$. Each term in this sum corresponds to a pair of profiles p and h with $h \in H$. First, we divide this sum into comparable and non-comparable pairs of profiles:

$$\dot{z}_H = \sum_{p,h \in H \text{ comparable}} z_p z_h M_{h,p} + \sum_{p,h \in H \text{ not comparable}} z_p z_h M_{h,p}$$

If p and h are comparable and $p \in H$, then the first sum contains the terms $z_p z_h M_{p,h}$ and $z_p z_h M_{h,p}$. Because they are comparable, $M_{p,h} = u_i(p) - u_i(h) = W_{p,h} = -W_{h,p} = -M_{h,p}$ (by equation (1)), so these terms cancel. The sum becomes

$$\dot{x}_H = \sum_{p \notin H, h \in H \text{ comparable}} z_p z_h M_{h,p} + \sum_{p,h \in H \text{ not comparable}} z_p z_h M_{h,p}$$

If p and h are comparable, and $p \notin H$ and $h \in H$, $M_{h,p} > 0$ because the arc $p \longrightarrow h$ is necessarily directed into h (H is a sink component). Hence all terms in the first sum are non-negative.

Now consider the case where p and h are non-comparable. Suppose $p = (\alpha, \beta)$ and $h = (\gamma, \delta)$ where none of these are equal, because p and h are not comparable. Let $a := (\alpha, \delta)$ and $b := (\gamma, \beta)$ be the other two profiles in this 2×2 subgame. First observe that $z_p z_h = x_\alpha y_\beta x_\gamma y_\delta = z_a z_b$. So, for instance, if $z_p z_h M_{h,p}$ and $z_a z_b M_{a,b}$ are both in the sum, we can combine them into a single term $z_p z_h (M_{h,p} + M_{a,b})$. We will now group all such terms with common z-coefficients. Which of these pairs actually appear in the sum depends on which of p, a, b, h are in H. We know $h \in H$, so the remaining cases are:

1. All of p, a, b, h are in H: The sum contains all of the terms $M_{p,h} + M_{h,p} + M_{a,b} + M_{b,a}$. Expanding this by Definition 1 gives $W_{a,p} + W_{b,p} + W_{a,h} + W_{b,h} + W_{p,b} + W_{h,b} + W_{h,a} + W_{p,a} = 0$, because each term $W_{i,j} = -W_{j,i}$. Hence if all are in H, these terms cancel in the sum.

2. Three are in H—we assume w.l.o.g. that $p \notin H$ and $a, b, h \in H$. This is a cavity of H. Then the terms in the sum are $M_{h,p} + M_{a,b} + M_{b,a}$. By the same argument as above, $M_{h,p} + M_{a,b} + M_{b,a} = -M_{p,h}$, and $M_{p,h} = W_{a,h} + W_{b,h}$. Hence, the sum of these terms is non-negative *if and only if* this cavity is pseudoconvex. By assumption, all cavities are pseudoconvex, so $M_{h,p} + M_{a,b} + M_{b,a} = -M_{p,h}$ is non-negative.

3. Two are in H—assume w.l.o.g. that $p, a \notin H$ and $b, h \in H$. The sum therefore contains the terms $M_{h,p}$ and $M_{b,a}$. By Definition 1, $M_{h,p} + M_{b,a} = W_{p,a} + W_{h,a} + W_{a,p} + W_{b,p} = W_{h,a} + W_{b,p}$. Since b and h are in H and q and b are not, the arcs $a \longrightarrow h$ and $p \longrightarrow b$ must be directed into the component, so $W_{h,a} > 0$ and $W_{b,p} > 0$. Hence this term is also non-negative.

4. Only h is in H. This last case is the most difficult, because here the terms can be negative. Each such pair contains only the term $z_p z_h M_{h,p}$. However, note that because $a, b \notin H$, $z_a, z_b < \epsilon$. Hence $z_p z_h = z_a z_b < \epsilon^2$. Thus all negative terms in $\dot{x}_H$ have coefficient at most ϵ^2.

We have now grouped the terms in this sum by their distinct z-coefficients, so each term has the form $x_\alpha y_\beta x_\gamma y_\delta (M_{i,j} + \dots)$. For simplicity, we write $K_{\alpha,\beta,\gamma,\delta} := \sum M_{i,j}$ where $z_i z_j = x_\alpha y_\beta x_\gamma y_\delta$. We now define $\mu := \min\{K_{\alpha,\beta,\gamma,\delta} \ : \ K_{\alpha,\beta,\gamma,\delta} > 0\}$ and similarly $m := \max\{|K_{\alpha,\beta,\gamma,\delta}| \ : \ K_{\alpha,\beta,\gamma,\delta} < 0\}$. These are the smallest and largest positive and negative terms respectively. There are at most N^2 terms in this sum, where N is the number of profiles. By the above, each negative term has coefficient at most ϵ^2, so the total sum of negative terms is at most $-mN^2\epsilon^2$. Now select an $h \in H$ where $z_h \geq (1-\epsilon)/|H|$. Since $z_H = 1 - \epsilon$, such a node must exist. Then:

$$\dot{x}_H \geq \sum_{p \notin H,\ z_p z_h = x_\alpha y_\beta x_\gamma y_\delta,\ K_{\alpha,\beta,\gamma,\delta} > 0} z_p z_h \mu - mN^2\epsilon^2 = \mu z_h \sum_{p \notin H,\ z_p z_h = x_\alpha y_\beta x_\gamma y_\delta,\ K_{\alpha,\beta,\gamma,\delta} > 0} z_p - mN^2\epsilon^2$$

This inequality holds because we have retained the contribution from all negative terms (in the $mN^2\epsilon^2$ term), and reduced the set of positive terms. Specifically we have included terms where the coefficient equals $z_p z_h$ for our fixed h and where $p \notin H$. Now we must determine the sum $\sum z_p$ over these p. The total sum over z_p with $p \notin H$ is ϵ, but some z_p are not included in this sum, if they correspond to a negative term $z_w z_h K_{\alpha,\beta,\gamma,\delta} < 0$. However, by the argument above, this occurs only in case (4). There, the remaining profiles a and b are not in H, and so contribute two positive terms $z_a z_h W_{h,a}$ and $z_b z_h W_{h,b}$ to this sum. By this argument, at least 2/3 of the terms in this sum must be positive. Also, each such z_p has $z_p < \epsilon^2$. We obtain

$$\begin{aligned} \dot{z}_H &\geq k((1-\epsilon)/|H|)(\epsilon - (1 - |H|)\epsilon^2/3) - mN^2\epsilon^2 \\ &\geq k\epsilon/|H| - o(\epsilon^2) \end{aligned}$$

Thus, for small enough $\epsilon > 0$, this term is strictly positive. ■