The Projection Dynamic, the Replicator Dynamic, and the Geometry of Population Games*†

William H. Sandholm, Emin Dokumacı, and Ratul Lahkar‡

December 4, 2006

Abstract

Every population game defines a vector field on the set of strategy distributions X. The projection dynamic maps each population game to a new vector field: namely, the one closest to the payoff vector field among those that never point outward from X. We investigate the geometric underpinnings of the projection dynamic, describe its basic game-theoretic properties, and establish a number of close connections between the projection dynamic and the replicator dynamic.

1 Introduction

Many strategic interactions that arise in applications involve large numbers of small anonymous agents, each of whom plays one of a finite number of roles.[1] Population games provide a simple and general model for capturing interactions of this kind. In a population game, agents in the same role choose from the same set of strategies, and each agent's payoff to a given strategy is the same function of the distribution of opponents' behavior. Since large numbers of agents make equilibrium assumptions especially suspect, population games are best analyzed using an explicitly dynamic approach, with agents adjusting their choices in response to their current strategic environment.

We begin this paper by considering population games from a geometric point of view. By seeing a population game as defined not by a collection of payoff functions, but by a vector field, we can describe complicated strategic interactions entirely in pictures, allowing a game's incentive

*This paper contains color figures.

†We thank seminar audiences at Concordia, Penn State, the Stockholm School of Economics, Wisconsin, and the 2005 SIAM Conference on Optimization for helpful comments. Financial support from NSF Grants SES-0092145 and SES-0617753 is gratefully acknowledged.

‡Department of Economics, University of Wisconsin, 1180 Observatory Drive, Madison, WI 53706, USA. e-mail: whs@ssc.wisc.edu, edokumaci@wisc.edu, rlahkar@wisc.edu; websites: http://www.ssc.wisc.edu/∼whs, http://www.ssc.wisc.edu/∼edokumac.

[1]These applications range from externality pricing and macroeconomic spillovers (Sandholm (2001,2005), Cooper (1999)), to natural selection and animal behavior (Maynard Smith (1982), Hofbauer and Sigmund (1988)), to route selection and mode choice in transportation networks (Beckmann et. al. (1956), Sheffi (1985)), to selfish routing and congestion in computer networks (Roughgarden (2005)). See Sandholm (2006b) for an overview and further references.

structure to be digested at a glance. Moreover, the geometric approach provides a new perspective on standard game-theoretic concepts. By noting a simple geometric characterization of Nash equilibrium, we can find a game's Nash equilibria using pictures alone. By the same token, an evolutionary dynamic can be viewed as a method of constructing "feasible distortions" of payoff vector fields—that is, of mapping each payoff vector field to a new vector field that specifies a feasible direction of motion from every strategy distribution.

Viewing evolutionary dynamics this way suggests the following question: can we find an evolutionary dynamic that always minimizes the distortion imposed on the payoff vector field? We call the dynamic described by this criterion the *projection dynamic*. This dynamic was first introduced in the transportation science literature in the work of Nagurney and Zhang (1997). In this paper, we study the projection dynamic's geometric underpinnings, its basic game-theoretic properties, and its many close connections with the replicator dynamic of Taylor and Jonker (1978), the original dynamic of evolutionary game theory.

s we noted earlier, any population game can be represented as a vector field F on the space $X \subset \mathbf{R}^n$. of strategy distributions. Here n is the total number of strategies in the game; $F : X \to \mathbf{R}^n$ assigns each strategy distribution $x \in \mathbf{R}^n$ a payoff vector $F(x) \in \mathbf{R}^n$, where component $F_i(x)$ represents the payoff to strategy i when the state is x. The projection dynamic for the game F is defined by a new vector field $V : X \to \mathbf{R}^n$; the vector $V(x) \in \mathbf{R}^n$ is the best approximation of the payoff vector $F(x)$ by a feasible direction of motion through X from x. In formal terms, $V(x)$ is the closest point projection of $F(x)$ onto $TX(x)$, the tangent cone of the set X at point x.

An obvious technical difficulty with the projection dynamic is that it is discontinuous: $V(x)$ can change abruptly when the state x reaches the boundary of the state space. Fortunately, it follows from general results of Henry (1973) and Aubin and Cellina (1984) (see also Dupuis and Ishii (1991) and Dupuis and Nagurney (1993)) that solutions to the projection dynamic exist, are unique, and are Lipschitz continuous in their initial conditions. Nevertheless, solutions to this dynamic have some properties quite different from those of standard dynamics: its solutions can merge in finite time, and can enter and exit the boundary of X repeatedly as time passes.

Since the projection dynamic respects payoffs to the greatest possible extent, it comes as no surprise that it exhibits appealing game-theoretic properties: the rest points of the projection dynamic are always identical to the Nash equilibria of the underlying game, and that the dynamic converges to Nash equilibrium from every initial conditions in all potential games and all stable games.

Still, while the projection dynamic is derived from a natural mathematical motivation, whether it admits a similarly natural economic foundation is not obvious. Fortunately, we are able to provide such a foundation by deriving the projection dynamic from an explicit model of individual choice. In this model, agents occasionally receive opportunities to choose new strategies, with both the timing of these opportunities and the probabilities of switches being described by a function of payoffs and the population state called a revision protocol.[2] We show that if agents follow certain

[2]This model of microfoundations for evolutionary dynamics is introduced in Sandholm (2006a), building on earlier

revision protocols under which unpopular strategies are more likely to be abandoned than popular ones, then their aggregate behavior is described by the projection dynamic.

In fact, these microfoundations provide the first connection between the projection dynamic and the replicator dynamic of Taylor and Jonker (1978). If we take any of the protocols noted above based on *abandonment* of *unpopular* strategies, and reformulate them to be driven by *adoption* of *popular* strategies—in other words, by imitation—then aggregate behavior under these new protocols is described by the replicator dynamic.

Unlike the revision protocols that underlie many other evolutionary dynamics,[3] those that generate the replicator and projection dynamics condition directly on the population state. The exact form of this conditioning leads these dynamics to enjoy a property we call *inflow-outflow symmetry*. This property is important because of its implications for dominated strategies: namely, that dominated strategies must always be losing ground to their dominating strategies at interior population states.[4]

It has been known since the work of Akin (1980) that the replicator dynamic eliminates strictly dominated strategies so long as play begins at an interior initial condition. This result follows easily from the invariance of the interior of X under the replicator dynamic, combined with inflow-outflow symmetry. Since the projection dynamic also satisfies inflow-outflow symmetry, it is natural to expect similar conclusions about dominated strategies to follow. This intuition turns out to be wrong. Using the fact that solutions of the projection dynamic can enter and leave the boundary of X, we construct an example in which a strictly dominated strategy appears and then disappears from the population in perpetuity. Thus, the fact that the dominated strategy is losing ground to the dominating strategy at almost all states is not enough to guarantee its elimination.

In the final section of the paper, we explore a striking connection between the replicator and projection dynamics in potential games. To start, we note that on the interior of X, the projection dynamic of a potential game is actually the gradient system generated by the game's potential function: in other words, the dynamic always follows the path that ascends potential in the most direct fashion. The connection with the replicator dynamic then arises by way of a result of Akin (1979): building on work of Kimura (1958) and Shahshahani (1979), Akin (1979) shows that in potential games, the replicator dynamic of a potential game is also a gradient system defined by the game's potential function, but only after the state space has been transformed by a nonlinear change of variable, one that attaches greater weight to changes in the use of rare strategies. We conclude the paper with a direct proof of Akins (1979) result: unlike Akins (1979) original proof, ours does not require the introduction of tools from differential geometry. This result and those described above demonstrate the deep connections between two seemingly unrelated models of the evolution of behavior in games.

---

work by Bjornerstedt and Weibull (1996), Benaim and Weibull (2003), and Sandholm (2003)

[3]For instance, the best response dynamic (Gilboa and Matsui (1991)), the logit dynamic (Fudenberg and Levine (1998)), the Brown-von Neumann-Nash dynamic (Brown and von Neumann (1950)), and the pairwise difference dynamic (Smith (1984)).

[4]This is in stark contrast with the work of Hofbauer and Sandholm (2006b), who construct a game with a strictly dominated strategy under which a large class of evolutionary dynamics admit an interior attractor.

2 Population Games

2.1 Definitions

Let $\mathcal{P} = \{1, \ldots, p\}$ be a *society* consisting of $p \geq 1$ *populations* of agents. Agents in population p form a continuum of *mass* $m^p > 0$. Agents in population p choose strategies from the set $S^p = \{1, \ldots, n^p\}$. The total number of strategies in all populations is $n = \sum_{p \in \mathcal{P}} n^p$.

During play, each agent in population p selects a (pure) strategy from S^p. The set of *population states* (or *strategy distributions*) for population p is $X^p = \{x^p \in \mathbf{R}_+^{n^p} : \sum_{i \in S^p} x_i^p = m^p\}$. The scalar $x_i^p \in \mathbf{R}_+$ represents the mass of players in population p choosing strategy $i \in S^p$. Elements of $X = \prod_{p \in \mathcal{P}} X^p = \{x = (x^1, \ldots, x^p) : x^p \in X^p\}$, the set of *social states*, describe behavior in all $\mathcal{P}$ populations at once. When there is only one population, we omit the superscript p from our notation and assume that the population's mass is 1.

We generally take the sets of populations and strategies as fixed and identify a game with its payoff function. A *payoff function* $F : X \to \mathbf{R}^n$ is a Lipschitz continuous map that assigns each social state a vector of payoffs, one for each strategy in each population. Observe that F defines a *vector field* on $\mathbf{R}^n$: it sends a "nice" subset of $\mathbf{R}^n$ into $\mathbf{R}^n$ itself. $F_i^p : X \to \mathbf{R}$ denotes the payoff function for strategy $i \in S^p$, while $F^p : X \to \mathbf{R}^{n^p}$ denotes the payoff functions for all strategies in S^p. Similar notational conventions are used throughout the paper.

Social state $x \in X$ is a *Nash equilibrium* of F if all agents in all populations play best responses. Formally, $x \in NE(F)$ if

$$x_i^p > 0 \implies i \in \operatorname*{argmax}_{j \in S^p} F_j^p(x) \text{ for all } i \in S^p \text{ and } p \in \mathcal{P}$$

.

2.2 Examples

Example 2.1 (Random matching in symmetric normal form games) *An n-strategy symmetric normal form game is defined by a payoff matrix $A \in \boldsymbol{R}^{n \times n}$. A_{ij} is the payoff a player obtains when he chooses strategy i and his opponent chooses strategy j; this payoff does not depend on whether the player in question is called player 1 or player 2. When a population of agents are randomly matched to play this game, the (expected) payoff to strategy i at population state is x is $F_i(x) = \sum_{j \in S} A_{ij} x_j$; hence, the population game associated with A is the linear game $F(x) = Ax$.*

Example 2.2 (Congestion games) *Congestion games provide a basic model of multilateral externalities. For concreteness, we describe these games using the context of highway network congestion. Consider a collection of towns is connected by a network of links L. For each ordered pair $p \in \mathcal{P}$ of towns, there is a population of agents, each of whom needs to commute from the first town in the pair (where he lives) to the second (where he works). To accomplish this, the agent must choose a path: each path $i \in S^p$ consists of a set of links $L_i^p \subseteq L$ connecting the towns in pair p. An agent's payoff from choosing path i is the negation of the delay on this path. The delay on a path*

is the sum of the delays on its links, and the delay on link is a function of the number of agents using that link. Formally,

$$F_i^p(x) = -\sum_{l \in L_i^p} c_l(u_l(x)), \quad \textit{where } u_l(x) = \sum_{p \in \mathcal{P}} \sum_{i \in S^p:\, l \in L_i^p} x_i^p.$$

When congestion games are used to model highway networks or other environments in which externalities are negative, the cost functions c_l are increasing in the utilization levels u_l. But one can also use congestion games to capture positive externalities by assuming that the cost functions c_l are decreasing.

3 The Geometry of Population Games

3.1 Tangent Spaces and Orthogonal Projections for Population Games

3.1.1 Definitions

The *tangent space* of X^p, denoted TX^p, is the smallest subspace of $\mathbf{R}^{n^p}$ that contains all vectors describing motions between points in X^p. In other words, if x^p, $y^p \in X^p$, then $y^p - x^p \in TX^p$, and TX^p is the span of all vectors of this form. Evidently, TX^p contains exactly those vectors in $\mathbf{R}^{n^p}$ whose components sum to zero: $TX^p = \{z^p \in \mathbf{R}^{n^p} : \sum_{i \in S^p} z_i^p = 0\}$. All directions of motion between points in the set of social states X are contained in its tangent space, the product set $TX = \prod_{p \in \mathcal{P}} TX^p$.

Projections—in particular, projections of payoff vectors onto sets of feasible directions of motion—play a central role throughout this paper. When the target space of a projection is a linear subspace, like the tangent spaces TX^p and TX, the appropriate notion of projection is *orthogonal projection*. The geometric definition of orthogonal projection is well-known; algebraically, orthogonal projections are the linear operations represented by symmetric idempotent matrices.

We represent the orthogonal projection onto the subspace $TX^p \subset \mathbf{R}^{n^p}$ by the matrix $\Phi \in \mathbf{R}^{n^p \times n^p}$, defined by $\Phi = I - \frac{1}{n^p}\mathbf{1}\mathbf{1}'$. Here $\mathbf{1} = (1, ..., 1)'$ is the vector of ones, so $\frac{1}{n^p}\mathbf{1}\mathbf{1}'$ is the matrix whose entries are all $\frac{1}{n^p}$.

The projection Φ has a simple interpretation. If v^p is a payoff vector in $\mathbf{R}^{n^p}$, the projection of v^p onto TX^p is

$$\Phi v^p = v^p - \frac{1}{n^p}\mathbf{1}\mathbf{1}'v^p = v^p - \mathbf{1}\left(\frac{1}{n^p}\sum_{k \in S^p} v_k^p\right).$$

Thus, the ith component of the vector Φv^p is the difference between the actual payoff to strategy i and the unweighted average payoff of all strategies in S^p. Put differently Φv^p discards information about the absolute level of payoffs under v^p while retaining information about relative payoffs. This is interesting from a game-theoretic point of view, since incentives, and hence Nash equilibria, only depend on payoff differences. Therefore, when incentives (as opposed to, e.g., efficiency) are our main concern, the projected payoff vectors Φv^p are sufficient statistics for the actual payoff vectors

v^p.

Since $TX = \prod_{p\in\mathcal{P}} TX^p$ is a product set, the orthogonal projection onto TX is represented by a block diagonal matrix, $\mathbf{\Phi} = \text{diag}(\Phi, \ldots, \Phi) \in \mathbf{R}^{n\times n}$. If we apply $\mathbf{\Phi}$ to the payoff vector $v = (v^1, \ldots, v^p)$, the resulting vector $\mathbf{\Phi}v = (\Phi v^1, \ldots, \Phi v^p)$ lists the normalized payoffs in each population.

3.1.2 Drawing Population Games

We noted earlier that the population game $F : X \to \mathbf{R}^n$ can be viewed as a vector field. This perspective enables us to present population games graphically. We begin by considering two single-population, two-strategy games. The first is a coordination game, the second a Hawk-Dove game:

$$F(x) = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ 2x_2 \end{pmatrix} \quad \text{and} \quad F(x) = \begin{pmatrix} -1 & 2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_H \\ x_D \end{pmatrix} = \begin{pmatrix} 2x_D - x_H \\ x_D \end{pmatrix}.$$

Figures 1(a) and 1(b) present these two games' payoff vector fields, along with their projections,

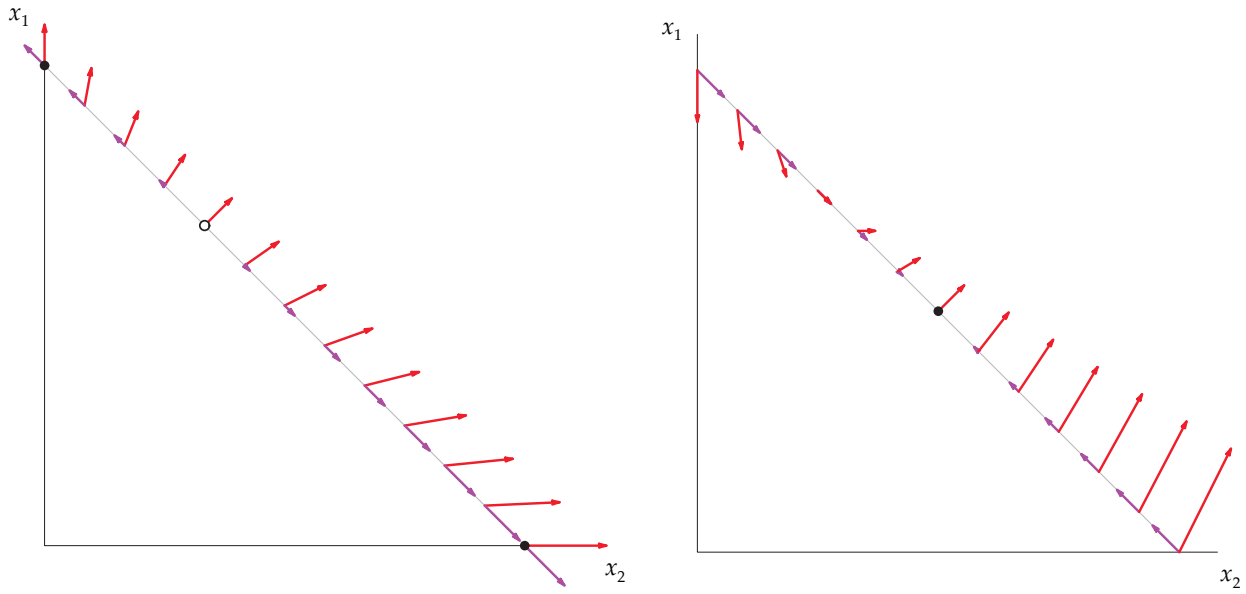
$$\Phi F(x) = \Phi \begin{pmatrix} x_1 \\ 2x_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{2}x_1 - x_2 \\ -\frac{1}{2}x_1 + x_2 \end{pmatrix} \quad \text{and} \quad \Phi F(x) = \Phi \begin{pmatrix} 2x_D - x_H \\ x_D \end{pmatrix} = \begin{pmatrix} \frac{1}{2}(x_D - x_H) \\ \frac{1}{2}(x_H - x_D) \end{pmatrix}.$$

The figures are synchronized with the payoff matrix by using the *vertical* coordinate to represent the mass on the *first* strategy and the *horizontal* coordinate to represent the mass on the *second* strategy. At various states x, we draw (scaled down) versions of the corresponding payoff vectors $F(x)$ and projected payoff vectors $\Phi F(x)$.

Let us focus first on Figure 1(a), representing the coordination game. At the pure state $e_1 = (1, 0)$, at which all agents play strategy 1, the payoffs to the two strategies are $F_1(e_1) = 1$ and $F_2(e_1) = 0$, so the payoff vector $F(e_1)$ points directly *upward*. At the interior Nash equilibrium $x^* = (x_1^*, x_2^*) = (\frac{2}{3}, \frac{1}{3})$, each strategy earns a payoff of $\frac{2}{3}$; the arrow representing payoff vector $F(x^*) = (\frac{2}{3}, \frac{2}{3})$ is drawn at a right angle to the simplex, implying that the projected payoff vector $\Phi F(x^*) = (0, 0)$ is null. Similar logic explains how the payoff vectors are drawn at other states, and how Figure 1(b) is constructed as well.

These diagrams help us visualize the incentives faced by agents playing these games. In the coordination game, the payoff vectors "push outward" toward the two axes, reflecting an incentive structure that drives the population toward the two pure Nash equilibria. In contrast, payoff vectors in the Hawk-Dove game "push inward", away from the axes, reflecting forces leading the population toward the interior Nash equilibrium $x^* = (\frac{1}{2}, \frac{1}{2})$.

Representing three-strategy games in two-dimensional pictures requires more care. Figure 2 presents a "three-dimensional" picture of the simplex X situated in its ambient plane $\text{aff}(X) = \{x \in \mathbf{R}^3 : \sum_{i\in S} x_i = 1\}$ in $\mathbf{R}^3$, known as the *affine hull* of X. When we draw the simplex on a sheet of paper as an equilateral triangle, the paper represents this plane. Each payoff vector $F(x)$



(a) Coordination (b) Hawk-Dove

Figure 1: The payoff vector field $F(\cdot)$ and its projection $\Phi F(\cdot)$ in two two-strategy games.

in a three-strategy game is an element of $\mathbf{R}^3$, but the projected payoff vector $\Phi F(x)$ lies in the two-dimensional tangent space TX. If the vector $\Phi F(x)$ is drawn as an arrow rooted at state x, then by construction this arrow will lie in the plane $\text{aff}(X)$ (see Figure 2).

In Figures 3(a) and 3(b), we use this approach to draw pictures of two single-population games with three strategies: a coordination game and a standard Rock-Paper-Scissors game:

$$F(x) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 \\ 2x_2 \\ 3x_3 \end{pmatrix} \quad \text{and} \quad F(x) = \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_R \\ x_P \\ x_S \end{pmatrix} = \begin{pmatrix} x_S - x_P \\ x_R - x_S \\ x_P - x_R \end{pmatrix}.$$

Of course, these drawings are actually of the projected payoffs

$$\Phi F(x) = \Phi \begin{pmatrix} x_1 \\ 2x_2 \\ 3x_3 \end{pmatrix} = \begin{pmatrix} \frac{1}{3}(2x_1 - 2x_2 - 3x_3) \\ \frac{1}{3}(-x_1 + 4x_2 - 3x_3) \\ \frac{1}{3}(-x_1 - 2x_2 + 6x_3) \end{pmatrix} \quad \text{and} \quad \Phi F(x) = \Phi \begin{pmatrix} x_S - x_P \\ x_R - x_S \\ x_P - x_R \end{pmatrix} = \begin{pmatrix} x_S - x_P \\ x_R - x_S \\ x_P - x_R \end{pmatrix}.$$

But since standard RPS is symmetric zero-sum (i.e., since its payoff matrix is skew-symmetric), the original and projected vector fields for this game are identical.

Much like Figure 1(a), Figure 3(a) shows that in the three-strategy coordination game, the projected payoff vectors push outward toward the extreme points of the simplex. Figure 3(b) exhibits a property that is only possible when there are three or more strategies: instead of heading toward Nash equilibria, the vectors in this figure describe cycle around state $x^* = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, the unique Nash equilibrium of standard RPS

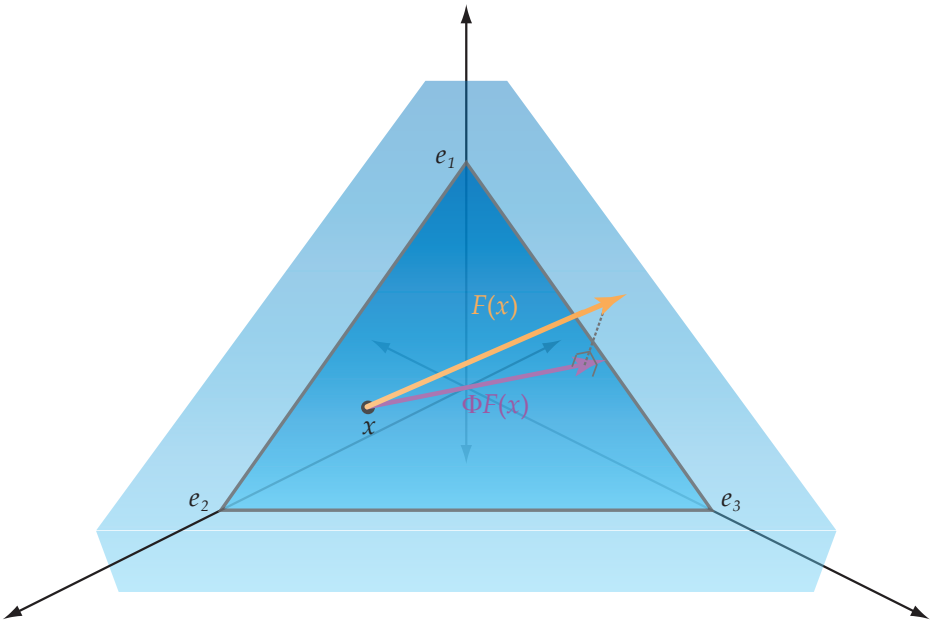
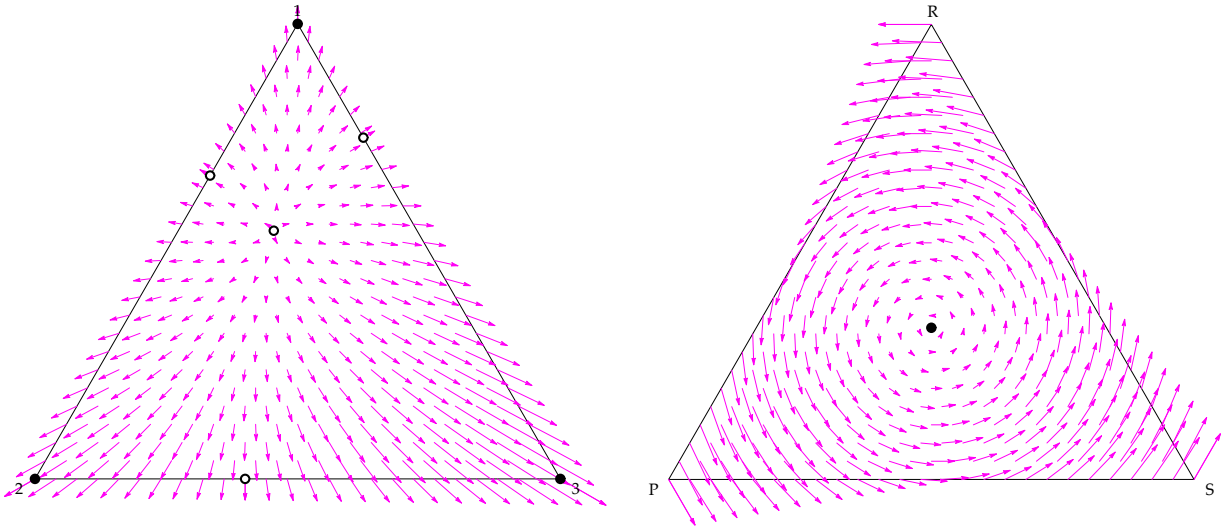


Figure 2: The simplex $X \subset \mathbf{R}^3$ and its affine hull.



(a) Coordination

(b) Standard Rock-Paper-Scissors

Figure 3: The projected payoff vector field $\Phi F(\cdot)$ in two three-strategy games.

3.2 Tangent Cones and Normal Cones for Convex Sets

All vectors $z \in TX$ represent feasible motions from social states in the (relative) interior of X. To represent a feasible motion from a boundary state, a vector cannot cause an unused strategy to lose mass. To describe sets of such motions, we introduce the notion of a tangent cone to a convex set; then, defining closest point projections onto these sets will enable us to define the projection dynamic. For further background on these topics, see Hiriart-Urruty and Lemaréchal (2001).

3.2.1 Definitions

The set $K \subseteq \mathbf{R}^n$ is a *cone* if whenever it contains the vector z, it contains each vector αz with $\alpha > 0$. If K is a closed convex cone, its *polar cone* K° is a new closed convex cone:

$$K^\circ = \{y \in \mathbf{R}^n : y'z \leq 0 \text{ for all } z \in K\}.$$

In words, K° contains all vectors that form a weakly obtuse angle with each vector in K.

If the closed convex cone K is symmetric, in the sense that $K = -K$, then K is actually a linear subspace of $\mathbf{R}^n$; in this case, $K^\circ = K^\perp$, the orthogonal complement of K. More generally, polarity defines an *involution* on the set of closed convex cones: that is, $(K^\circ)^\circ = K$ for any closed convex cone K.

If $C \subset \mathbf{R}^n$ is a closed convex set, then the *tangent cone* of C at state $x \in C$, denoted $TC(x)$, is the closed convex cone

$$TC(x) = \text{cl}\left(\{z \in \mathbf{R}^n : z = \alpha(y - x) \text{ for some } y \in C \text{ and some } \alpha \geq 0\}\right).$$

If $C \subset \mathbf{R}^n$ is a *polytope* (i.e., the convex hull of a finite number of points), then the closure operation is redundant. In this case, $TC(x)$ is the set of directions of motion from x that initially remain in C; more generally, $TC(x)$ also contains the limits of such directions. If x is in the (relative) interior of C, then $TC(x)$ is just TC, the tangent space of C; otherwise, $TC(x)$ is a strict subset of TC.

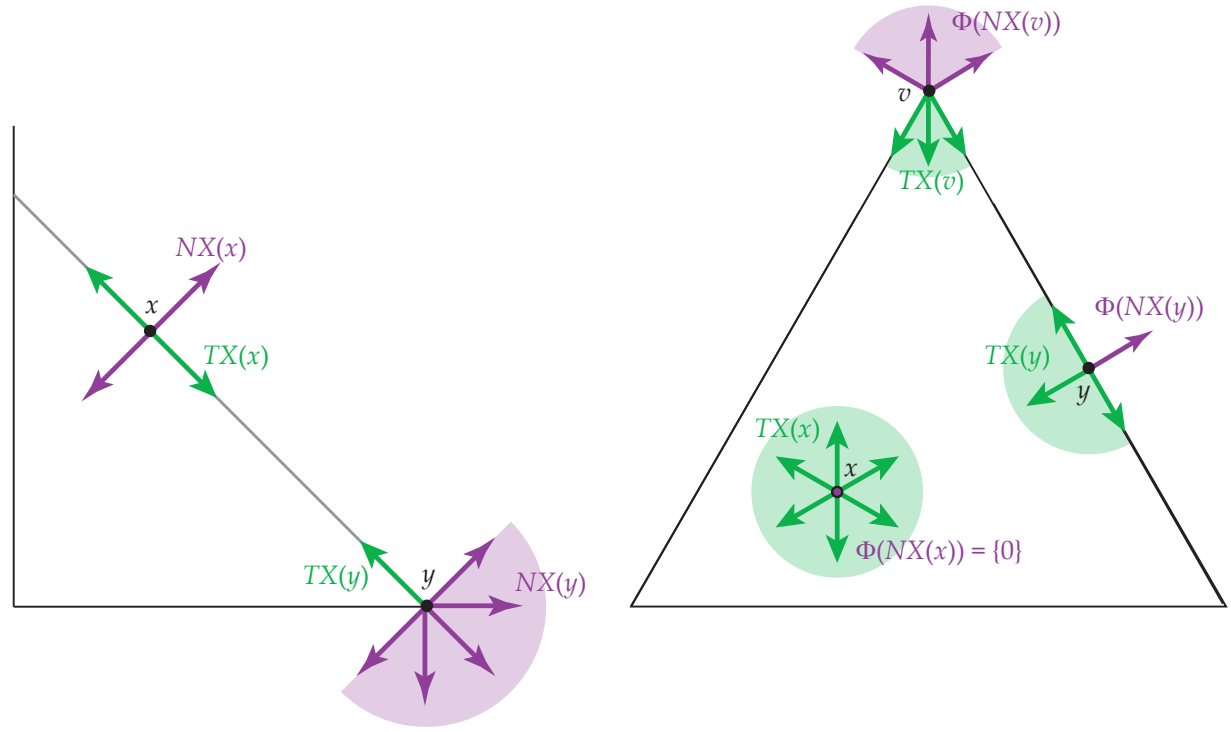
The *normal cone* of C at x is the polar of the tangent cone of C at x: that is, $NC(x) = TC(x)^\circ$. By definition, $NC(x)$ is a closed convex cone, and it contains every vector that forms a weakly obtuse angle with every feasible displacement vector at x.

3.2.2 Normal Cones and Nash Equilibria

When X is the set of social states of a population game, each tangent cone $TX(x)$ contains the feasible directions of motion from social state $x \in X$. In multipopulation cases $TX(x)$ can be decomposed population by population.[5]

In Figures 4(a) and 4(b), we sketch examples of tangent cones and normal cones when X is the state space of a two-strategy game and of a three-strategy game. Since Figure 4(b) is two-

[5]That is: since $X = \prod_{p \in \mathcal{P}} X^p$ is a product set, $TX(x) = \prod_{p \in \mathcal{P}} TX^p(x^p)$ is a product set as well. Similarly, we have that $NX(x) = \prod_{p \in \mathcal{P}} NX^p(x^p)$; this fact is used in the proof of Theorem 3.1 below.



(a) $n = 2$ (b) $n = 3$

Figure 4: Tangent cones and (projected) normal cones in two- and three-strategy games.

dimensional, with the sheet of paper representing the plane $\text{aff}(X)$, the figure actually displays the projected normal cones $\Phi NX(x)$.

At first glance, normal cones might appear to be less relevant to game theory than tangent cones. Theorem 3.1 shows that this impression is false, and provides us with a simple geometric description of Nash equilibria of population games. Versions of this result can be found in the literature on variational inequalities—see Harker and Pang (1990) and Nagurney and Zhang (1996).

Theorem 3.1 *Let F be a population game. Then $x \in NE(F)$ if and only if $F(x) \in NX(x)$.*

Proof.
$$\begin{aligned} x \in NE(F) &\Leftrightarrow [x_i^p > 0 \Rightarrow F_i^p(x) \geq F_j^p(x) \text{ for all } i, j \in S^p, p \in \mathcal{P}] \\ &\Leftrightarrow (x^p)' F^p(x) \geq (y^p)' F^p(x) \text{ for all } y^p \in X^p, p \in \mathcal{P} \\ &\Leftrightarrow (y^p - x^p)' F^p(x) \leq 0 \text{ for all } y^p \in X^p, p \in \mathcal{P} \\ &\Leftrightarrow (z^p)' F^p(x) \leq 0 \text{ for all } z^p \in TX^p(x), p \in \mathcal{P} \\ &\Leftrightarrow F^p(x) \in NX^p(x^p) \text{ for all } p \in \mathcal{P} \\ &\Leftrightarrow F(x) \in NX(x). \blacksquare \end{aligned}$$

In the figures above, Nash equilibria are marked with dots. In the two-strategy games (Figures 1(a) and 1(b)), the Nash equilibria are those states x at which the payoff vector $F(x)$ lies in the

normal cone $NX(x)$, as Theorem 3.1 shows. In the three-strategy games, the Nash equilibria are those states x at which the projected payoff vector $\Phi F(x)$ lies in the projected normal cone $\Phi NX(x)$; this is an easy corollary of Theorem 3.1. Even if the dots were left out of these figures, we could locate the Nash equilibria by examining the arrows alone.

3.3 Closest Point Projections onto Convex Cones

3.3.1 Definition and Characterization

To define the projection dynamic on the boundary of X, we need to introduce projections onto convex cones. If $K \subset \mathbf{R}^n$ is a closed convex cone, then the *closest point projection* $\Pi_K : \mathbf{R}^n \to K$ is defined by

$$\Pi_K(v) = \operatorname*{argmin}_{z \in K} |z - v|.$$

If K is a subspace of $\mathbf{R}^n$ (i.e., if $K = -K$), then the closest point projection onto K is simply the orthogonal projection onto K.

The fundamental result about projections onto closed convex cones is the *Moreau Decomposition Theorem*, which generalizes the notion of orthogonal decomposition to the "one-sided" world of convex cones. In words, this theorem tells us that the projections of the vector v onto K and K° are the unique vectors in K and K° that are orthogonal to one another and that sum to v. For a proof, see Hirriat-Uruty and Lemaréchal (2001).

Theorem 3.2 (The Moreau Decomposition Theorem) *Let $K \subseteq \mathbf{R}^n$ and $K^\circ \subseteq \mathbf{R}^n$ be a closed convex cone and its polar cone, and let $v \in \mathbf{R}^n$. Then the following are equivalent:*

(i) $v_K = \Pi_K(v)$ *and* $v_{K^\circ} = \Pi_{K^\circ}(v)$.
(ii) $v_K \in K, v_{K^\circ} \in K^\circ, (v_K)' v_{K^\circ} = 0$, *and* $v = v_K + v_{K^\circ}$.

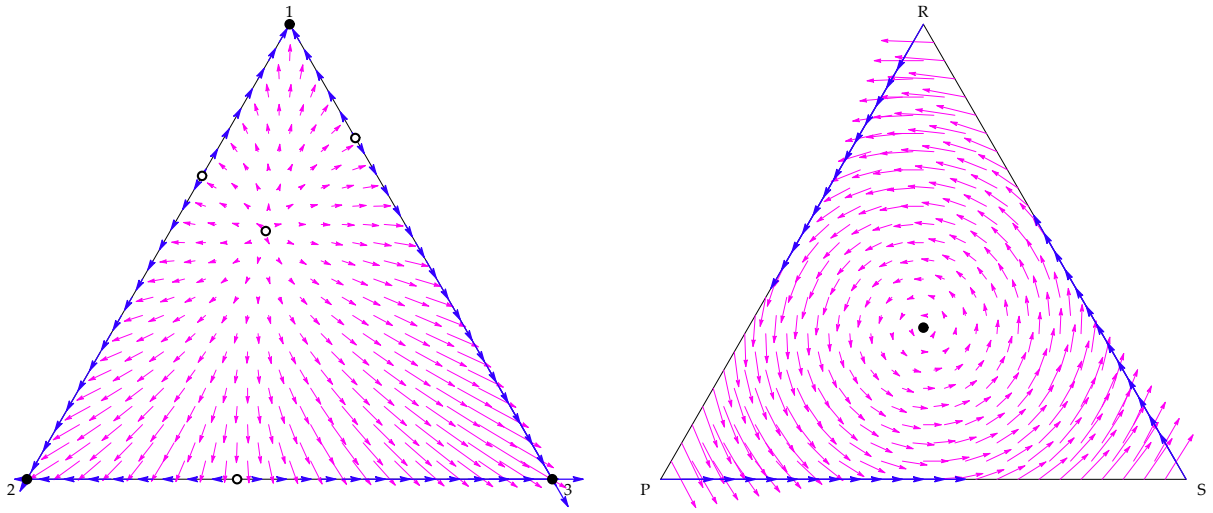
3.3.2 Projecting Payoff Vectors onto Tangent Cones of X

In Figures 5(a) and 5(b), we draw the projected vector fields $V(\cdot) = \Pi_{TX(\cdot)} F(\cdot)$ for our two three-strategy games. In these figures, each payoff vector $F(x)$ is represented by $\Pi_{TX(x)}(F(x))$, the best approximation by a feasible direction of motion from x. Evidently, the states x at which the projected payoff vector is null are precisely the Nash equilibria of the underlying game. That this is true in general is an immediate consequence of Theorems 3.1 and 3.2:

Corollary 3.3 *Let F be a population game. Then $x \in NE(F)$ if and only if $\Pi_{TX(x)}(F(x)) = \mathbf{0}$.*

3.3.3 An Explicit Formula for $\Pi_{TX(x)}(v)$

In general, explicitly computing a closest point projection onto a convex cone requires solving a quadratic program. The next result, Theorem 3.4, shows that the projection onto $TX^p(x^p)$ admits a simple explicit description. Since the projection onto $TX(x)$ can be decomposed population by



(a) Coordination (b) Standard Rock-Paper-Scissors

Figure 5: Vector fields $V(\cdot) = \Pi_{TX(\cdot)}F(\cdot)$ obtained by projecting payoffs onto tangent cones.

population, this formula is sufficient to determine $\Pi_{TX(x)}(v)$.[6] This formula underlies much of the analysis to follow: in addition to providing us with an explicit expression for the projection dynamic, it provides the key to establishing microfoundations for this dynamic, and it undergirds our proof that dominated strategies can survive.

Theorem 3.4 *The projection $\Pi_{TX^p(x^p)}(v^p)$ can be expressed as follows:*

$$(\Pi_{TX^p(x^p)}(v^p))_i = \begin{cases} v_i^p - \frac{1}{\#\mathcal{S}^p(v^p,x^p)} \sum_{j\in\mathcal{S}^p(v^p,x^p)} v_j^p & \text{if } i \in \mathcal{S}^p(v^p,x^p), \\ 0 & \text{otherwise.} \end{cases}$$

Here, the set $\mathcal{S}^p(v^p,x^p) \subseteq S^p$ contains all strategies in $\text{support}(x^p)$*, along with any subset of $S^p - \text{support}(x^p)$ that maximizes the average $\frac{1}{\#\mathcal{S}^p(v^p,x^p)} \sum_{j\in\mathcal{S}^p(v^p,x^p)} v_j^p$.*

The proof of Theorem 3.4 can be found in Appendix A.

To explain Theorem 3.4, let us avoid superscripts by focusing on the single population case. Imagine that $v \in \mathbf{R}^n$ is the vector of payoffs earned by strategies in $S = \{1,\dots,n\}$ at state $x \in X$. When x is in the interior of X, the tangent cone $TX(x)$ is the just the subspace TX; therefore, the closest point projection onto $TX(x)$ is the orthogonal projection $\Phi = I - \frac{1}{n}\mathbf{1}\mathbf{1}'$ from Section 3.1, which subtracts the average payoff under v from each component of v:

$$(\Pi_{TX(x)}(v))_i = (\Phi v)_i = v_i - \frac{1}{n}\sum_{j\in S} v_j. \tag{1}$$

[6]More explicitly: since $TX(x) = TX^1(x^1) \times \cdots \times TX^p(x^p)$ is a product set, we have that $\Pi_{TX(x)}(v) = \Pi_{TX^1(x^1)}(v^1) \times \cdots \times \Pi_{TX^p(x^p)}(v^p)$.

If instead there is exactly one unused strategy at state x, say strategy n, then the tangent cone $TX(x)$ consists of vectors in TX whose nth component is nonnegative. In this case, if strategy n earns an above average payoff, then $(\Phi v)_n \geq 0$, and Theorem 3.4 tells us that formula (1) still applies. But if strategy n earns a below average payoff, then $(\Phi v)_n < 0$, so $\Pi_{TX(x)}(v)$ cannot equal Φv. Instead, according to Theorem 3.4, the nth component of $\Pi_{TX(x)}(v)$ is set to 0, while the remaining components of $\Pi_{TX(x)}(v)$ are obtained from those of v by subtracting $\frac{1}{n-1}(v_1+\ldots+v_{n-1})$ from each. More generally, the components of $\Pi_{TX(x)}(v)$ corresponding to "bad" unused strategies are set to 0, while the remaining components are obtained from v by normalizing away the average of these components only.

4 The Projection Dynamic

4.1 Definition

An *evolutionary dynamic* is a map that assigns each population game F a dynamical system on the set of social states X. Typically, this dynamical system is described by a differential equation $\dot{x} = V(x)$. To define the projection dynamic, we suppose that the vector $V(x)$ is the best approximation of the payoff vector $F(x)$ by a feasible direction of motion from x. As we noted in the Introduction, this dynamic first appears in the transportation science literature in the work of Nagurney and Zhang (1997) (see also Nagurney and Zhang (1996, Ch. 8)).

Definition 4.1 *The* projection dynamic *assigns each population game F the differential equation*

$$\dot{x} = \Pi_{TX(x)}(F(x)). \tag{P}$$

When $x \in \text{int}(X)$, the tangent cone $TX(x)$ is just the subspace TX, so the explicit formula for (P) is simply $\dot{x} = \mathbf{\Phi}F(x)$. When $x \in \text{bd}(X)$, the explicit formula for (P) is given by Theorem 3.4.

When F is generated by random matching to play the normal form game A (i.e., when F takes the linear form $F(x) = Ax$), the dynamic (P) is especially simple. On $\text{int}(X)$, the dynamic is described by the linear equation $\dot{x} = \Phi Ax$. On $\text{bd}(X)$, Theorem 3.4 tells us that

$$(\Pi_{TX(x)}(Ax))_i = \begin{cases} (Ax)_i - \frac{1}{\#\mathcal{S}(Ax,x)} \sum_{j\in\mathcal{S}(Ax,x)} (Ax)_j & \text{if } i \in \mathcal{S}(Ax,x), \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Notice that once the set of strategies $\mathcal{S}(Ax,x)$ is fixed, this expression is a linear function of x. Thus, under single population random matching, the projection dynamic is piecewise linear. This observation is crucial to our proof of Theorem 5.2 on the survival of dominated strategies under (P).[7]

[7] A relative of the projection dynamic from the economics literature is the *linear dynamic* of Friedman (1991, p. 643 and 661). Like the projection dynamic, the linear dynamic is defined by $\dot{x} = \mathbf{\Phi}F(x)$ on the interior of X. But at boundary states, the linear dynamic posits that all unused strategies have growth rates of zero, making the boundary

4.2 Basic Properties

4.2.1 Existence, Uniqueness, and Continuity of Forward Solutions

Since the projection dynamic (P) is discontinuous at the boundary of X, standard results on the existence of solutions to differential equations do not apply to it. Indeed, the appropriate notion of solution for this equation must allow for kinks at boundary states: we call the trajectory $\{x_t\}_{t\geq 0} \subset X$ a (*Carathéodory*) *solution* to (P) if it is absolutely continuous and satisfies equation (P) at almost every $t \geq 0$. Theorem 4.2 shows that despite these inconveniences, forward-time solutions to (P) exist, are unique, and are Lipschitz continuous in their initial conditions.

Theorem 4.2 *Let a Lipschitz continuous population game F and an initial condition $\xi \in X$ be given. Then there exists a unique solution $\{x_t\}_{t\geq 0}$ to the projection dynamic* (P) *with $x_0 = \xi$. Solutions to* (P) *are Lipschitz continuous in their initial conditions: if $\{x_t\}_{t\geq 0}$ and $\{y_t\}_{t\geq 0}$ are solutions to* (P)*, then $|y_t - x_t| \leq |y_0 - x_0|\, e^{Kt}$ for all $t \geq 0$, where K is the Lipschitz coefficient for F.*

The existence result in Theorem 4.2 follows from more general existence results proved by Henry (1973) and Aubin and Cellina (1984, Sec. 5.6), and later rediscovered by Dupuis and Ishii (1991) and Dupuis and Nagurney (1993). In Appendix B, we briefly present a proof of Theorem 4.2 based on the Viability Theorem for differential inclusions.

While the projection dynamic admits a unique solution from every initial condition, these solutions differ from solutions to standard Lipschitz differential equations in a number of important respects, as the following examples illustrate.

Example 4.3 (A three-strategy coordination game) *In Figure 6, we present the phase diagram for the projection dynamic in the three-strategy coordination game*

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

This phase diagram provides the solution trajectories generated by the vector field pictured in Figure 5(a). In both Figure 6 and Figure 7, the background color represents the speed of motion: regions where motion is fastest are red, while regions where motion is slowest are blue. (Since the dynamic changes discontinuously at the boundary of X, the colors are only guaranteed to represent speeds at interior states.)

As one travels away from the completely mixed equilibrium, the speed of motion increases until the boundary of the state space is reached; thus, bd(X) is reached in finite time. At this point, the

forward invariant, while the growth rate of each strategy in use is the difference between its payoff and the average payoff of the strategies in use. By Theorem 3.4, the linear dynamic is identical to the projection dynamic if and only if at each state $x \in \text{bd}(X)$ and in all populations $p \in \mathcal{P}$, every unused strategy earns a payoff no greater than the average payoff of the strategies in use.

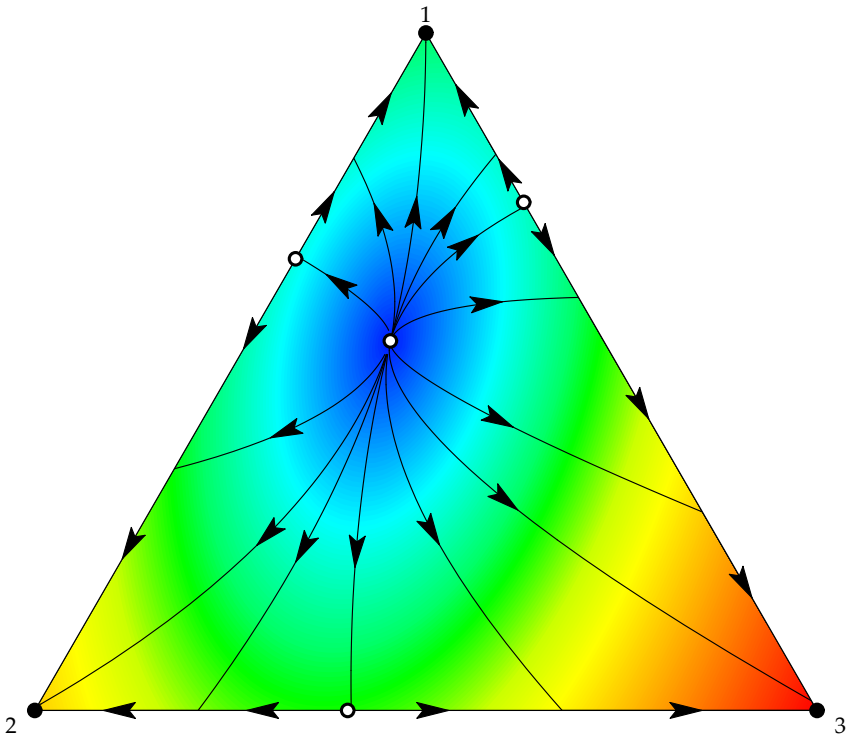


Figure 6: Phase diagram of (P) for a coordination game. Colors represent speeds.

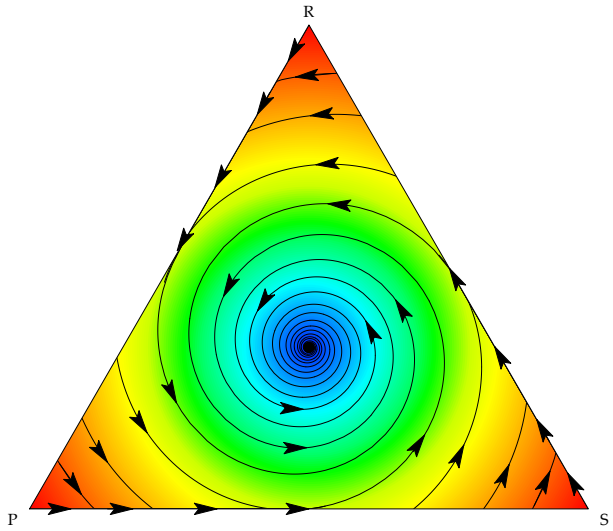
solution changes direction, merging with a solution that travels along the boundary of the simplex, implying that backward-time solutions from boundary states are not unique. All solutions reach one of the seven symmetric Nash equilibria of A in finite time, with solutions from almost all initial conditions leading to one of the three strict equilibria.

Example 4.4 (Rock-Paper-Scissors games) *Consider the payoff matrix*

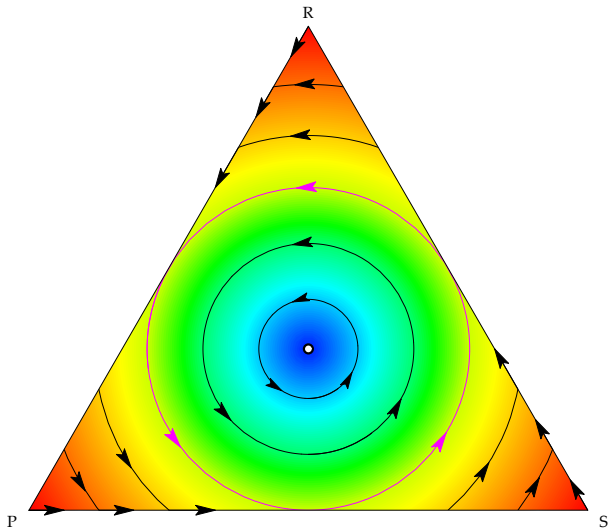
$$A = \begin{pmatrix} 0 & -l & w \\ w & 0 & -l \\ -l & w & 0 \end{pmatrix}.$$

with $w, l > 0$. A is a good RPS game *if $w > l$ (that is, if the winner's profit is higher than the loser's loss). A is a* standard RPS game *if $w = l$, and A is a* bad RPS game *if $w < l$. In all cases, the unique symmetric Nash equilibrium of A is $x^* = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$.*

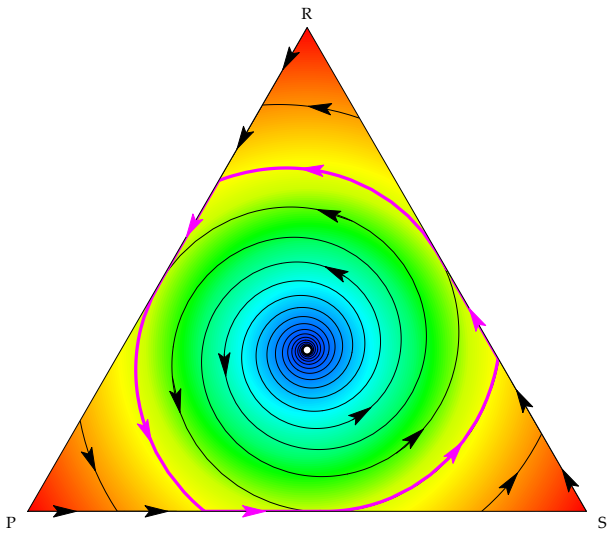
Figure 7 presents phase diagrams for the projection dynamic in good RPS ($w = 3, l = 2$), standard RPS ($w = l = 1$), and bad RPS ($w = 2, l = 3$). In all three games, solutions spiral around the Nash equilibrium in a counterclockwise direction. In good RPS (Figure 7(a)), all solutions converge to the Nash equilibrium. Solutions that start at an interior state close to a vertex first hit and then travel along the boundary of X; they then reenter $\operatorname{int}(X)$ *and spiral inward toward x^*. In standard RPS (Figure 7(b)), all solutions enter closed orbits at a fixed distance from x^*. Solutions starting at distance $\frac{1}{\sqrt{6}}$ or greater from x^* (i.e., all solutions at least as far from x^* as the state*



(a) Good RPS ($w = 3, l = 2$).



(b) Standard RPS ($w = l = 1$).



(c) Bad RPS ($w = 2, l = 3$).

Figure 7: Phase diagrams of (P) for three Rock-Paper-Scissors games. Colors represent speeds.

(0, $\frac{1}{2}$, $\frac{1}{2}$)) merge with the closed orbit at distance $\frac{1}{\sqrt{6}}$ from x^*; other solutions maintain their initial distance from x^* forever. In bad RPS (Figure 7(c)), all solutions other than the one starting at x^* enter the same closed orbit. This orbit alternates between segments through the interior of X and segments that traverse the boundaries of X. This game is the starting point for our analysis of the survival of dominated strategies in Section 5.2.

In all three versions of RPS, there are solution trajectories starting in $\operatorname{int}(X)$ *that reach* $\operatorname{bd}(X)$ *in finite time. By Theorem 3.4, solutions leave* $\operatorname{bd}(X)$ *at the point where the unused strategy's payoff exceeds the average payoff of the two strategies in use.*

4.2.2 Nash Stationarity and Positive Correlation

Next, we first establish two basic game-theoretic properties of the projection dynamic. To state these properties, suppose that F is a population game and $\dot{x} = V(x)$ an evolutionary dynamic for this game. Define $RP(V) = \{x \in X : V(x) = \mathbf{0}\}$ to be the set of rest points of V.

(NS)	*Nash stationarity*	$RP(V) = NE(F)$.
(PC)	*Positive correlation*	$[V^p(x) \neq \mathbf{0} \implies V^p(x)'F^p(x) > 0]$ for all $p \in \mathcal{P}$.

Nash stationarity (NS) requires that the Nash equilibria of the game F and the rest points of the dynamic V coincide. Dynamics satisfying this condition provide the strongest support for the fundamental solution concept of noncooperative game theory. Positive correlation (PC) imposes restrictions on disequilibrium dynamics. It requires that whenever population $p \in \mathcal{P}$ is not at rest, there is a positive correlation between the growth rates and payoffs of strategies in S^p. In geometric terms, (PC) demands that the direction of motion $V^p(x)$ and the payoff vector $F^p(x)$ form acute angles with one another whenever $V^p(x)$ is not null. This property, versions of which have been studied by Friedman (1991), Swinkels (1993), and Sandholm (2001), is an important ingredient in establishing global convergence results—see Section 4.2.3 below.

Both (NS) and (PC) are simple consequences of the developments in Section 3.

Proposition 4.5 *The projection dynamic satisfies Nash stationarity* (NS) *and positive correlation* (PC).

Proof. Property (NS) is a restatement of Corollary 3.3. To prove property (PC), we take the Moreau decomposition of the payoff vector $F^p(x)$:

$$\begin{aligned} V^p(x)'F^p(x) &= \Pi_{TX^p(x^p)}(F^p(x))' \left(\Pi_{TX^p(x^p)}(F^p(x)) + \Pi_{NX^p(x^p)}(F^p(x))\right) \\ &= \left|\Pi_{TX^p(x^p)}(F^p(x))\right|^2 \\ &\geq 0. \end{aligned}$$

The inequality is strict if and only if $\Pi_{TX^p(x^p)}(F^p(x)) = V^p(x) \neq \mathbf{0}$. ■

4.2.3 Global Convergence in Potential Games

To conclude this section, we show that the projection dynamic converges to Nash equilibrium from all initial conditions in two important classes of games: potential games (Monderer and Shapley (1996), Sandholm (2001, 2006c)) and stable games (Hofbauer and Sandholm (2006a)).

In a potential game, information about all strategies' payoffs is encoded in a single scalar-valued function on the state space X. Following Sandholm (2006c), we call the population game $F : X \to \mathbf{R}^n$ a *potential game* if it admits a *potential function* $f : X \to \mathbf{R}$ satisfying

$$\nabla f(x) = \mathbf{\Phi} F(x) \text{ for all } x \in X. \tag{PG}$$

Since the domain of f is X, the gradient vector $\nabla f(x)$ is the unique vector in TX that represents the derivative of f at X, in the sense that $f(y) = f(x) + \nabla f(x)'(y - x) + o(|y - x|)$ for all $y \in X$. Definition (PG) requires that this gradient vector always equal the projected payoff vector $\mathbf{\Phi} F(x)$. Common interest games, congestion games, and games defined by variable externality pricing schemes are all potential games.

In potential games, all limit points of the projection dynamic are Nash equilibria.

Theorem 4.6 *Let F be a potential game with potential function f. Then f is a strict Lyapunov function for the projection dynamic* (P) *on X. Therefore, each solution to* (P) *converges to a connected set of Nash equilibria of F.*

Proof. Property (PC) and the fact that $V(x) \in TX$ imply that

$$\dot{f}(x) = \nabla f(x)'\dot{x} = (\mathbf{\Phi} F(x))'V(x) = F(x)'V(x) = \sum_{p \in \mathcal{P}} F^p(x)'V^p(x) \geq 0,$$

and that $\dot{f}(x) = 0$ if and only if $V(x) = 0$. Therefore, standard results (e.g., Theorem 7.6 of Hofbauer and Sigmund (1988)) imply that every solution of (P) converges to a connected set of rest points of V. By Nash stationarity, these rest points are all Nash equilibria. ■

The three strategy coordination game from Example 4.3 is a potential game with potential function $f(x) = \frac{1}{2}((x_1)^2 + 2(x_2)^2 + 3(x_3)^2)$. In Figure 8, we present another phase diagram of the projection dynamic for this game, this time over a contour plot of the potential function f. Theorem 4.6 tells us that all solutions of (P) ascend this function.

In fact, we can see in the figure that on the interior of X, the solutions cross the level sets of f orthogonally. This reflects the fact that the projection dynamic is actually the *gradient system* for f on $\text{int}(X)$: in other words,

$$\dot{x} = \nabla f(x) \text{ on } \text{int}(X), \tag{3}$$

as follows immediately from equations (1), (P), and (PG). This observation is the source of one of the many connections between the projection and replicator dynamics that we explore in Section 5.

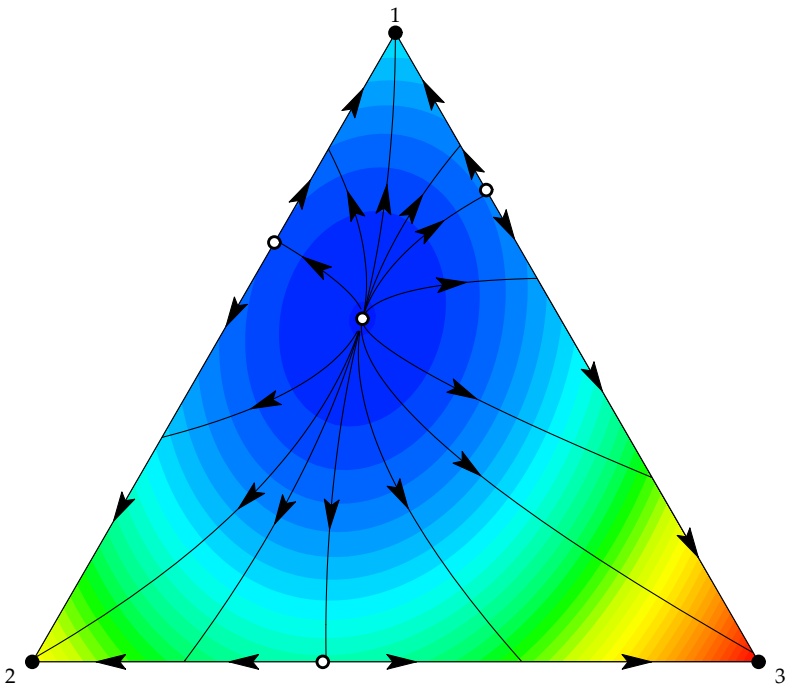


Figure 8: Phase diagram of (P) for a coordination game. Colors represent the value of potential.

4.2.4 Global Convergence and Cycling in Stable Games

The population game F is a *stable game* (Hofbauer and Sandholm (2006a)) if

$$(y - x)'(F(y) - F(x)) \leq 0 \text{ for all } x, y \in X. \tag{SG}$$

If inequality (SG) is strict whenever $y \neq x$, then F is a *strictly stable game*; if (SG) is always satisfied with equality, then F is a *null stable game*. The set of Nash equilibria of any stable game is convex; if F is strictly stable, then $NE(F)$ is a singleton. Games with an interior ESS, wars of attrition, and congestion games in which congestion is a bad are all stable games, while zero-sum games are null stable games.

Let

$$E_{x^*}(x) = |x - x^*|^2,$$

the squared Euclidean distance from the Nash equilibrium x^*. Nagurney and Zhang (1997) show that E_{x^*} is a Lyapunov function in any stable game. The analysis to follow includes a streamlined proof of their result.

Theorem 4.7 *Let x^* be a Nash equilibrium of F.*

(*i*) *If F is a stable game, then E_{x^*} is a Lyapunov function for* (P)*, so x^* is Lyapunov stable under* (P)*.*

(*ii*) *If F is a strictly stable game, then E_{x^*} is a strict Lyapunov function for* (P)*, so $NE(F) = x^*$ is globally asymptotically stable under* (P)*.*

(*iii*) *If F is a null stable game and $x^* \in \text{int}(X)$, then E_{x^*} defines a constant of motion for* (P) *on* $\text{int}(X)$.

Proof. The proof of Theorem 3.1 shows that x^* is a Nash equilibrium if and only if

$$(x - x^*)'F(x^*) \leq 0 \text{ for all } x \in X. \tag{4}$$

By adding this inequality to inequality (SG), Hofbauer and Sandholm (2006a) show that if F is a stable game, then $x^* \in NE(F)$ if and only if x^* is a *globally neutrally stable state* of F:

$$(x - x^*)'F(x) \leq 0 \text{ for all } x \in X; \tag{5}$$

while if F is a strictly stable game, then its unique Nash equilibrium x^* is also its unique *globally evolutionarily stable state*:

$$(x - x^*)'F(x) < 0 \text{ for all } x \in X - x^*. \tag{6}$$

Now suppose that F is stable. Using the Moreau decomposition, equation (5), and the fact that $x^* - x \in TX(x)$, we compute the time derivative of E_{x^*} over a solution to (P):

$$\begin{aligned} \dot{E}_{x^*}(x) &= \nabla E_{x^*}(x)'\dot{x} \\ &= 2(x - x^*)'\Pi_{TX(x)}(F(x)) \\ &= 2(x - x^*)'F(x) + 2(x^* - x)'\Pi_{NX(x)}(F(x)) \\ &\leq 2(x^* - x)'\Pi_{NX(x)}(F(x)) \\ &\leq 0. \end{aligned} \tag{7}$$

Thus, E_{x^*} is a Lyapunov function for (P), which implies that x^* is Lyapunov stable. If F is strictly stable, then equation (6) implies that the first inequality in (7) is strict; thus, E_{x^*} is a strict Lyapunov function for (P), and so x^* is globally asymptotically stable.

Finally, suppose that F is null stable and that x^* is an interior Nash equilibrium. The first of these assumptions tells us that equation (SG) always holds with equality, while the second implies that all pure strategies are optimal, and hence that equation (4) always holds with equality. Adding these two equalities shows that equation (5), and hence the first inequality in (7), always holds with equality. If $x \in \text{int}(X)$, then $NX(x)$ and $TX(x)$ are orthogonal, so the the second inequality in (7) holds with equality as well. Therefore, $\dot{E}_{x^*}(x) = 0$ on $\text{int}(x)$, and so E_{x^*} defines a constant of motion for (P) on this set. ■

To conclude this section, we show that at interior states, the squared speed of motion under (P) also serves as a Lyapunov function for (P). Unlike that of the distance function E_{x^*}, the definition of this function does not directly incorporate the Nash equilibrium x^*.

Theorem 4.8 *Let F be continuously differentiable. If F is a stable game, then $L(x) = |\mathbf{\Phi}F(x)^2|$ is a Lyapunov function for* (P) *on* $\text{int}(X)$*. If F is a null stable, then L defines a constant of motion for* (P) *on* $\text{int}(X)$*.*

Proof. Since F is C^1, the Fundamental Theorem of Calculus implies that F is stable if and only if

$$z'DF(x)z \leq 0 \text{ for all } z \in TX \text{ and } x \in X, \tag{8}$$

and that F is null stable game if and only if inequality (8) always binds.

When $x \in \text{int}(x)$, the projection dynamic is given by $\dot{x} = \mathbf{\Phi}F(x)$. Therefore, since $D\mathbf{\Phi}F(x) = \mathbf{\Phi}DF(x)$ and $\mathbf{\Phi}^2 = \mathbf{\Phi}$, we find that

$$\dot{L}(x) = \nabla L(x)'\dot{x} = 2(\mathbf{\Phi}F(x))'D\mathbf{\Phi}F(x)(\mathbf{\Phi}F(x)) = 2(\mathbf{\Phi}F(x))'DF(x)(\mathbf{\Phi}F(x)).$$

Of course, $\mathbf{\Phi}F(x) \in TX$. Therefore, if F is stable, then $\dot{L}(x) \leq 0$ on $\text{int}(X)$, and if F is null stable, then $\dot{L}(x) = 0$ on $\text{int}(X)$. ■

The results in this section are illustrated by our phase diagrams for the projection dynamic in RPS games (Example 4.4). Good RPS defines a strictly stable population game. In its phase diagram (Figure 7(a)), we see that distance from the equilibrium $x^* = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ falls over time, and that the speed of motion falls over time on $\text{int}(X)$. Standard RPS is a zero-sum game, and so defines a null stable population game. In this game's phase diagram (Figure 7(b)), both distance from equilibrium and the speed of motion remain fixed over time on $\text{int}(X)$.

5 The Projection Dynamic and the Replicator Dynamic

In this final section of the paper, we explore the many close connections between the projection dynamic and the replicator dynamic. To simplify notation, we restrict attention to the single population setting; however, all of the results to follow continue to hold when there are multiple populations. The replicator dynamic was introduced in the mathematical biology literature by Taylor and Jonker (1978), who use it to model natural selection driven by differences in birth and death rates. In economic contexts, the replicator dynamics describes the changes in the strategy distribution as agents imitates successful opponents (Björnerstedt and Weibull (1996), Schlag (1998)).

In the single population setting, the *replicator dynamic* for the population game F is defined by

$$\dot{x}_i = x_i\big(F_i(x) - \textstyle\sum_{k \in S} x_k F_k(x)\big). \tag{R}$$

In words: the *percentage* growth rate of strategy i equals the difference between the payoff to strategy i and the *weighted* average payoff under F at x (that is, the average payoff obtained by members of the population).

For the sake of comparison, recall that the projection dynamic takes the form

$$\dot{x}_i = \begin{cases} F_i(x) - \frac{1}{\#\mathcal{S}(F(x),x)} \sum_{j \in \mathcal{S}(F(x),x)} F_j(x) & \text{if } i \in \mathcal{S}(F(x),x), \\ 0 & \text{otherwise,} \end{cases} \tag{P}$$

Since $\mathcal{S}(F(x),x) = S$ on $\text{int}(X)$, we see that at interior states, (P) specifies that the *absolute* growth rate in population (P) equals the difference between the payoff to strategy i and the *unweighted* average payoff. At boundary states, some poorly performing unused strategies (namely, those not in $\mathcal{S}(F(x),x)$) are ignored, while the absolute growth rates of the remaining strategies are defined as before.

Thus, we see that both the replicator and projection dynamics convert payoff vector fields in differential equations in similar fashions, the key difference being that the replicator dynamic uses relative definitions, while the projection dynamic employs the corresponding absolute definitions. In the remainder of the paper, we explore the game-theoretic ramifications of this link.

5.1 Microfoundations

To provide an explicit link between individual choices and aggregate behavior, we need to derive deterministic dynamics like (R) from explicit models of individual choice. We derive deterministic dynamics on X from a model of individual choice, using *revision protocols* $\rho : \mathbf{R}^n \times X \to \mathbf{R}^{n \times n}_+$, which describes the process through which individual agents make decisions. As time passes, agents are chosen at random from the population and granted opportunities to switch strategies. When an i player receives such an opportunity, he switches to strategy j with probability proportional to the conditional switch rate $\rho_{ij}(F(x),x)$. Aggregate behavior in the game F is then described by the *mean dynamic*

$$\dot{x}_i = V_i(x) = \sum_{j \in S} x_j \rho_{ji}(F(x),x) - x_i \sum_{j \in S} \rho_{ij}(F(x),x). \tag{M}$$

The first term in equation (M) captures the inflow of agents into strategy i from other strategies, while the second term captures the outflow of agents from strategy i to other strategies.

As examples, consider the following three pairs of revision protocols:

$$\begin{aligned} \rho_{ij} &= x_j[F_j(x) - F_i(x)]_+, & \rho_{ij} &= \frac{1}{nx_i}[F_j(x) - F_i(x)]_+, \\ \rho_{ij} &= x_j(K - F_i(x)), & \rho_{ij} &= \frac{1}{nx_i}(K - F_i(x)), \\ \rho_{ij} &= x_j(F_j(x) + K), & \rho_{ij} &= \frac{1}{nx_i}(F_j(x) + K). \end{aligned}$$

The protocols in the first column are all based on imitation, as reflected in the initial x_j term. For instance, to implement the first protocol, an agent whose clock rings picks an opponent from his population at random; he then imitates this opponent only if the opponents' payoff is higher, doing

so with probability proportional to the payoff difference.[8] Each of these protocols generates the replicator dynamic as its mean dynamic. For the first protocol, we have that

$$\begin{aligned}\dot{x}_i &= \sum\nolimits_j x_j\rho_{ji} - x_i\sum\nolimits_j \rho_{ij} = \sum\nolimits_j x_j x_i[F_i(x) - F_j(x)]_+ - x_i\sum\nolimits_j x_j[F_j(x) - F_i(x)]_+ \\ &= x_i\sum\nolimits_j x_j(F_i(x) - F_j(x)) = x_i\left(F_i(x) - \sum\nolimits_j x_j F_j(x)\right).\end{aligned}$$

The derivations for the other two protocols are similar.

The protocols in the second column are obtained from those in the first by replacing x_j with $\frac{1}{nx_i}$. Thus, while in each of the imitative protocols, ρ_{ij} is *proportional* to the mass of agents playing the *candidate strategy* j, in the abandonment protocols, ρ_{ij} *inversely proportional* to the mass of agents playing the *current strategy* i: in other words, agents are more likely to abandon strategies that are currently unpopular.[9] It is easy to check that each of the latter protocols generates the projection dynamic as its mean dynamic on the interior of the state space X: for the first protocol, we have that

$$\begin{aligned}\dot{x}_i &= \sum\nolimits_j x_j\rho_{ji} - x_i\sum\nolimits_j \rho_{ij} = \sum\nolimits_j x_j\frac{[F_i(x)-F_j(x)]_+}{nx_j} - x_i\sum\nolimits_j \frac{[F_j(x)-F_i(x)]_+}{nx_i} \\ &= \tfrac{1}{n}\sum\nolimits_j (F_i(x) - F_j(x)) = F_i(x) - \tfrac{1}{n}\sum\nolimits_j F_j(x)\end{aligned}$$

However, it is easy to check that these protocols do not generate (P) on the boundary of X, where some strategies are not in use.

We now present a new pair of protocols which aggregate to (R) and (P) on the entire state space X. For each state $x \in X$, let

$$\hat{F}_i(x) = F_i(x) - \sum\nolimits_{k\in S} x_k F_k(x)$$

denote the difference between strategy i's payoff and the weighted average payoff, and let

$$\tilde{F}_i^{\mathcal{S}}(x) = F_i(x) - \tfrac{1}{\#\mathcal{S}(F(x),x)}\sum\nolimits_{k\in\mathcal{S}(F(x),x)} F_k(x)$$

denote the difference between strategy i's payoff and the unweighted average payoff of strategies in $\mathcal{S}(F(x), x)$. Consider the following two revision protocols:

$$\rho_{ij} = \begin{cases} [\hat{F}_i(x)]_- \cdot \dfrac{x_j\,[\hat{F}_j(x)]_+}{\sum_{k\in S} x_k\,[\hat{F}_k(x)]_+} & \text{if } \sum\limits_{k\in S} x_k\,[\hat{F}_k(x)]_+ > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

[8] This protocol is the *proportional imitation* protocol of [**?**]. The second protocol in the first column is the *imitation driven by dissatisfaction* protocol of [**?**].

[9] This protocol can be implemented in the following way: Suppose that at every moment in time, each agent chooses members of his population at random until he draws someone playing his own strategy, and then considers switching strategies at a rate proportional to the number of draws. As this number of draws has a *geometric*(x_i) distribution, the expected number of draws is $\frac{1}{x_i}$.

$$\rho_{ij} = \begin{cases} \dfrac{[\tilde{F}_i^{\mathcal{S}}(x)]_-}{x_i} \cdot \dfrac{[\tilde{F}_j^{\mathcal{S}}(x)]_+}{\sum_{k \in \mathcal{S}(F(x),x)} [\tilde{F}_k^{\mathcal{S}}(x)]_+} & \text{if } \sum\limits_{k \in \mathcal{S}(F(x),x)} x_i \, [\tilde{F}_k^{\mathcal{S}}(x)]_+ > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

In both protocols, the first term describes the rate at which an i player who receives a revision opportunity considers revising strategies. In both cases, the player only revises when i's payoff is below average (where the meaning of "average" is different in (9) and (10)), and the rate depends linearly on how far below average; in (10), the rate is also inversely proportional to x_i. The second term describes the probability with which an agent who opts to revise chooses strategy j. In both cases, only "above average" strategies are chosen, with strategies further above average being chosen more often; in (9), the choice probability is also weighted by the popularity of strategy j.

Theorem 5.1 verifies that these two protocols have the desired mean dynamics throughout the state space X. Its proof is presented in Appendix C.

Theorem 5.1

(*i*) *Revision protocol* (9) *generates the replicator dynamic* (R) *as its mean dynamic.*

(*ii*) *Revision protocol* (10) *generates the projection dynamic* (P) *as its mean dynamic.*

5.2 Inflow-Outflow Symmetry and Dominated Strategies

It is natural to expect evolutionary dynamics to eliminate dominated strategies. The first positive result on this question was proved by Akin (1980), who showed that the replicator dynamic eliminates strictly dominated strategies so long as the initial state is interior. This result was extended to broader classes of imitative dynamics by Samuelson and Zhang (1992) and Hofbauer and Weibull (1996). But while these results seem encouraging, they are actually quite special: Hofbauer and Sandholm (2006b) show continuous evolutionary dynamics that are not based exclusively on imitation do not eliminate strictly dominated strategies in all games.

In this section, we show that the projection dynamic shares with the replicator dynamic a property called inflow-outflow symmetry, and we show why this property leads to selection against dominated strategies under both of these dynamics on the interior of X. Despite this common property, the long run prospects for dominated strategies under the two dynamics are quite different: we show that because solutions to the projection dynamic can enter and exit the boundary of X, inflow-ouflow symmetry is not enough to ensure that dominated strategies are eliminated.

In our revision protocol model from the previous section, agents are occasionally selected at random and provided with revision opportunities; the recipients of these opportunities switch strategies at frequencies determined by the conditional switch rates $\rho_{ij} = \rho_{ij}(F(x), x)$. The evolution of the population state is then described by the mean dynamic (M).

$$\dot{x}_i = \sum_{j \in S} x_j \rho_{ji} - x_i \sum_{j \in S} \rho_{ij}. \tag{M}$$

Notice that x_i, representing the mass of players choosing strategy i, appears in this expression

for $\dot{x}_i$ asymmetrically. Since in order for an agent to switch away from strategy i, he must first be selected at random, x_i appears in the (negative) outflow term. But since agents switching to strategy i were previously playing other strategies, x_i does not appear in the inflow term.

We say that an evolutionary dynamic satisfies *inflow-outflow symmetry* if this asymmetry in equation (M) is eliminated by the dependence of the revision protocol ρ on the population state x. Under the replicator dynamic and other imitative dynamics, ρ_{ji} is proportional to x_i, making both the inflow and outflow terms in (M) proportional to x_i; thus, these dynamics exhibit inflow-outflow symmetry. Similarly, under the projection dynamic, which is based on abandonment, ρ_{ij} is inversely proportional to x_i whenever x_i is positive. As a result, neither the inflow nor the outflow term in equation (M) depends directly on x_i, yielding inflow-ouflow symmetry on $\text{int}(X)$.

Importantly, inflow-outflow symmetry can make it possible to show that a dominated strategy i will lose ground to the strategy j that dominates it. In the case of the replicator dynamic (R), the ratio x_i/x_j falls over time throughout $\text{int}(X)$:

$$\begin{aligned}\frac{d}{dt}\left(\frac{x_i}{x_j}\right) &= \frac{\dot{x}_i x_j - \dot{x}_j x_i}{x_j^2} = \frac{x_i\left(F_i(x) - \sum_{k\in S} x_k F_k(x)\right)\cdot x_j - x_j\left(F_j(x) - \sum_{k\in S} x_k F_k(x)\right)\cdot x_i}{x_j^2} \\ &= \frac{x_i}{x_j}(F_i(x) - F_j(x)) < 0 \text{ on } \text{int}(X).\end{aligned} \tag{11}$$

Under the the projection dynamic (P), the difference $x_i - x_j$ falls instead:

$$\begin{aligned}\frac{d}{dt}(x_i - x_j) &= \left(F_i(x) - \tfrac{1}{n}\sum_{k\in S} F_k(x)\right) - \left(F_j(x) - \tfrac{1}{n}\sum_{k\in S} F_k(x)\right) \\ &= F_i(x) - F_j(x) < 0 \text{ on } \text{int}(X).\end{aligned} \tag{12}$$

By combining equation (11) with the fact that $\text{int}(X)$ is invariant under (R), it is easy to prove that the replicator dynamic eliminates strictly dominated strategies along solutions in $\text{int}(X)$. But because solutions of the projection dynamic can enter and leave $\text{int}(X)$, the analogous argument based on equation (12) does not go through: while i will lose ground to j in $\text{int}(X)$, it might gain ground back on $\text{bd}(X)$, leaving open the possibility of survival.

To pursue this idea, we consider the following game due to Berger and Hofbauer (2006):

$$F(x) = Ax = \begin{pmatrix} 0 & -3 & 2 & 2 \\ 2 & 0 & -3 & -3 \\ -3 & 2 & 0 & 0 \\ -3-c & 2-c & -c & -c \end{pmatrix}\begin{pmatrix} x_R \\ x_P \\ x_S \\ x_T \end{pmatrix}. \tag{13}$$

The game defined by the first three strategies is bad RPS with $w = 2$ and $l = 3$. Figure 7(c), which presents the phase diagram for (P) in this game, shows that solutions other than the one at the Nash equilibrium $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ enter a closed orbit, an orbit that enters and exits the three edges

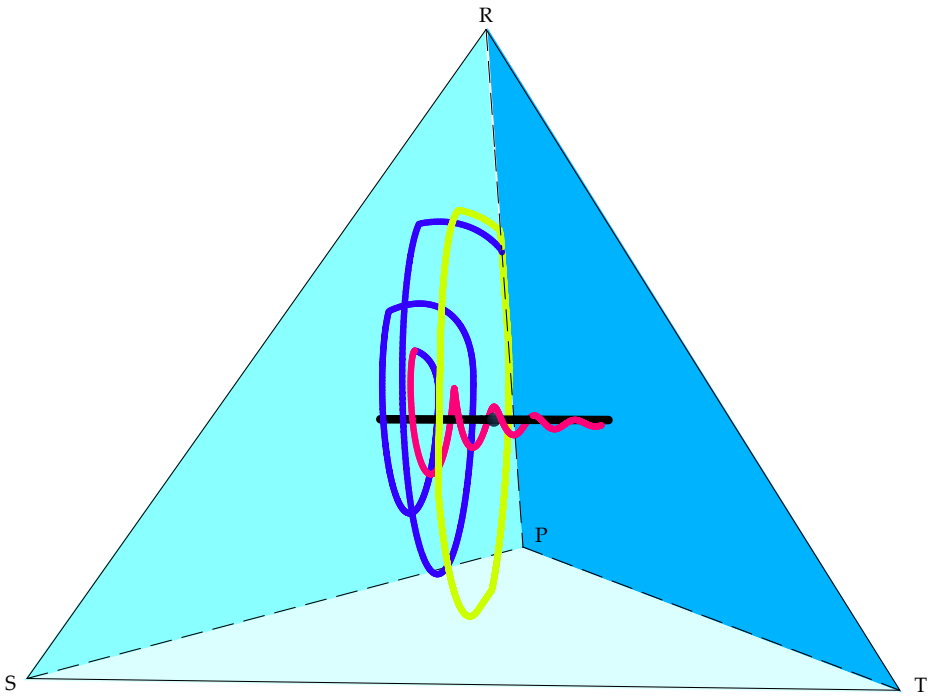


Figure 9: A solution to (P) in Rock-Paper-Scissors-Twin.

of the simplex. The fourth strategy of game (13), Twin, is a duplicate of Scissors, except that its payoff is always $c \geq 0$ lower than that of Scissors. When $c = 0$, the set of Nash equilibria of game (13) is the line segment L between $x^* = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0)$ and $(\frac{1}{3}, \frac{1}{3}, 0, \frac{1}{3})$. If $c > 0$, then Twin is strictly dominated, and the game's unique Nash equilibrium is x^*.

Figure 9 presents the solution to (P) from initial condition $(\frac{97}{300}, \frac{103}{300}, \frac{1}{100}, \frac{97}{300})$ when $c = \frac{1}{10}$. At first, the trajectory spirals down line segment L, as agents switch from Rock to Paper to Scissors/Twin to Rock, with Scissors replacing Twin as time passes (since $\dot{x}_T - \dot{x}_S = -\frac{1}{10}$ on $\text{int}(X)$). When Twin is eliminated, both it and Scissors earn less than the average of the payoffs to Rock, Paper, and Scissors; therefore, x_S falls while x_T stays fixed at 0, and so $x_T - x_S$ *rises*. Soon the solution enters $\text{int}(X)$, and so $\dot{x}_T - \dot{x}_S = -\frac{1}{10}$ once again. When the solution reenters face RPS, it does so at a state further away from the Nash equilibrium x^* than the initial point of contact. Eventually, the trajectory appears to enter a closed orbit on which the mass on Twin varies between 0 and roughly .36.

The existence and stability of this closed orbit is established rigorously in Theorem 5.2, whose proof can be found in Appendix D.

Theorem 5.2 *In game* (13) *with* $c = \frac{1}{10}$*, the projection dynamic* (P) *has an asymptotically stable closed orbit* γ *that absorbs all solutions from nearby states in finite time. This orbit, pictured in Figure 10, combines eight segments of solutions to linear differential equations as described in equation* (2)*; the approximate endpoints and exit times of these segments are presented in Table 1. Along the orbit, the value of* x_T *varies between 0 and approximately .359116.*

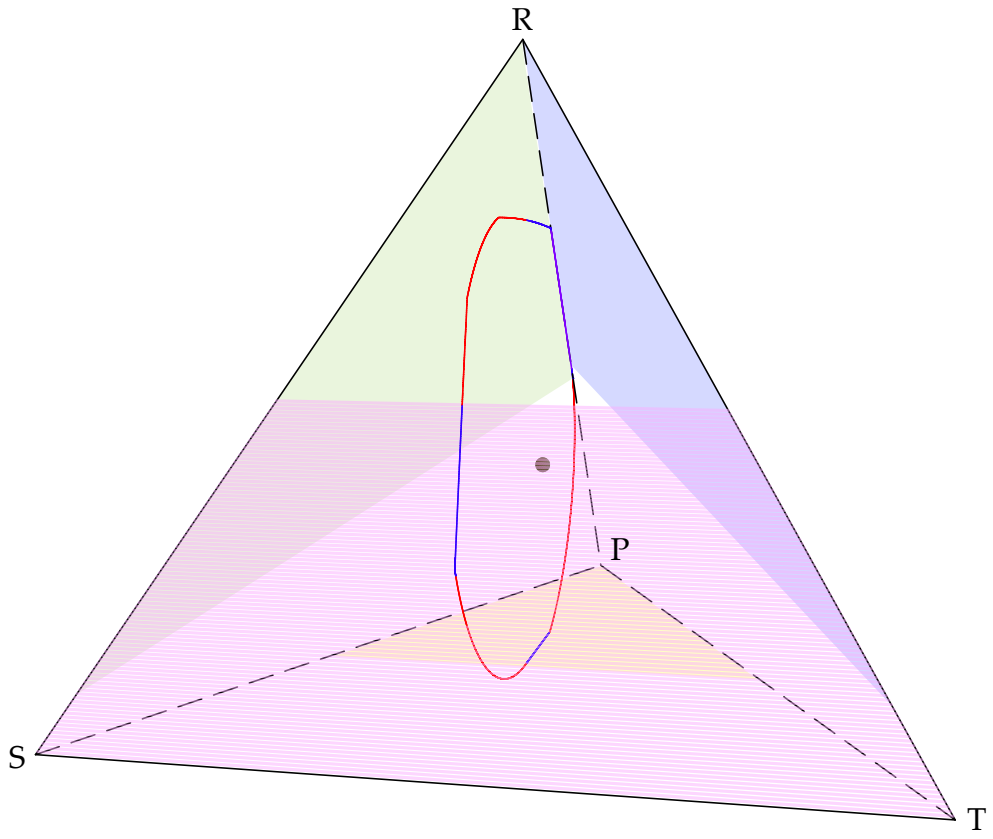


Figure 10: The closed orbit of (P) in Rock-Paper-Scissors-Twin.

Segment	Support $= \mathcal{S}(Ax, x)$	Exit point	Exit time
(initial state)	RP	(.466667, .533333, 0, 0)	0
1	RPS	(.446354, .552864, .000782, 0)	.015678
2	$RPST$	(0, .564668, .227024, .208308)	.324883
3	PST	(0, .413636, .307144, .279219)	.416973
4	$RPST$	(.256155 ,0, .395751, .348094)	.656509
5	RST	(.473913, 0, .288747, .237340)	.793914
6	$RPST$	(.709788, .244655, .045576, 0)	1.028310
7	RPS	(.693072, .306928, 0, 0)	1.065574
8	RP	(.466667, .533333, 0, 0)	1.252812

Table 1: Approximate transition points and transition times of the closed orbit γ.

Recall that in piecewise linear games like game (13), the projection dynamic is piecewise linear. On the interior of X, the dynamic is described by $\dot{x} = \Phi Ax$. On the boundary of X, it is described by equation (2). Thus, when the only unused strategy is strategy i, there are two possibilities for $\dot{x}$: if $F_i(x)$ does not exceed the average payoff of the other three strategies, then $\dot{x}_i = 0$, so the solution travels along the face of X where strategy i is absent; if instead $F_i(x)$ is greater than this average payoff, then the solution from x immediately enters $\text{int}(X)$. We illustrate these regions in Figure 10, where we shade the portions of the faces of X to which solutions "stick".

Similar considerations determine the behavior of (P) on the edges of the simplex. For instance, solutions starting at vertex R travel along edge RP until reaching state $\xi = (\frac{7}{15}, \frac{8}{15}, 0, 0)$, at which point they enter face RPS.[10]

The proof of Theorem 5.2 takes advantage of the piecewise linearity of the dynamic, the Lipschitz continuity of its solutions in their initial conditions, and the fact that solutions to the dynamic can merge in finite time. Because of piecewise linearity, we can obtain analytic solutions to the (P) within each region where (P) is linear. The point where a solution leaves one of these regions generally cannot be expressed analytically, but it can be approximated numerically to an arbitrary degree of precision. This approximation introduces a small error; however, the Lipschitz continuity of solutions places a tight bound on how quickly this error can propagate. Ultimately, our approximate solution starting from state ξ returns to edge RP. Since solutions cycle outward (cf Figure 7(c)), edge RP is reached between ξ and vertex R. While the point of contact we compute is only approximate, solutions from all states between vertex R and state ξ pass through state ξ. Therefore, since our total approximation error is very small, our calculations prove that the true solution must return to state ξ.

5.3 "Distance" from Equilibrium in Stable Games

Theorem 4.7 showed that in any stable game, distance from equilibrium serves as a Lyapunov function for the projection dynamic. Almost exact analogues of all of the statements in Theorem 4.7 can be proved for the replicator dynamic by replacing the distance function E_{x^*} with the distance-like function[11]

$$H_{x^*}(x) = \sum_{i:\, x_i^* > 0} x_i^* \log \frac{x_i^*}{x_i},$$

This Lyapunov function for (R) for stable games was introduced by Hofbauer et al. (1979) and Zeeman (1980); see Hofbauer and Sigmund (1988) and Hofbauer and Sandholm (2006a) for further details.

Figure 11 presents a phase diagram of the replicator dynamic in standard RPS, drawn on top of a contour plot of the function H_{x^*}. Evidently, interior solutions to (R) lie on the level sets of H_{x^*}. Compare this picture to Figure 7(c), which performs the same illustration for the projection

[10] State ξ lies between the vertices on edge RP of the "sticky" regions in faces RPT and RPS. These vertices lie at states $(\frac{71}{150}, \frac{79}{150}, 0, 0)$ and $(\frac{67}{150}, \frac{83}{150}, 0, 0)$, respectively.

[11] $H_{x^*}(x)$ is the relative entropy of x^* with respect to x. While $H_{(\cdot)}(\cdot)$ is not a true distance, $H_{x^*}(\cdot)$ is strictly concave, nonegative, and equal to 0 only when its argument x equals x^*.

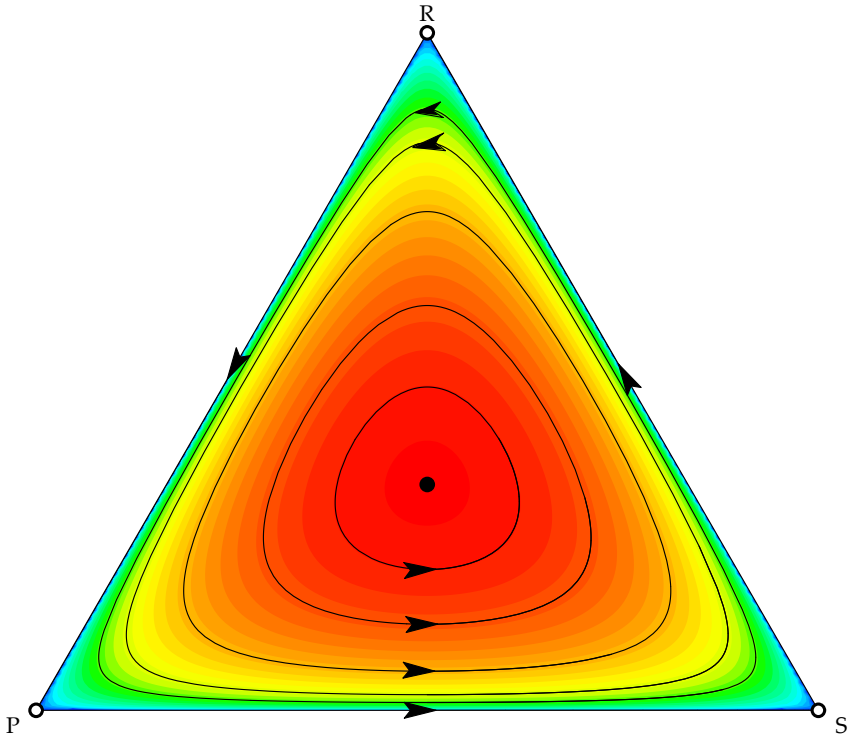


Figure 11: Phase diagram of the replicator dynamic in standard RPS. Colors represent the value of the Lyapunov function H_{x^*}.

dynamic and the function E_{x^*}.

5.4 Gradient Systems for Potential Games

In Section 4.2.3, we saw that the if $F : X \to \mathbf{R}^n$ is a potential game, then on the interior of X, the projection dynamic for F is actually the gradient system defined by for F's potential function f. Interestingly, it is possible to view the replicator dynamic for F as a gradient system for f as well.

Shahshahani (1979), building on the early work of Kimura (1958), showed that the replicator dynamic for a potential game is a gradient dynamic after a "change in geometry"—in particular, after the introduction of an appropriate Riemannian metric on $\text{int}(X)$. Subsequently, Akin (1979) (see also Akin (1990)) showed that Shahshahanis (1979) Riemannian manifold is isometric to the set $\mathcal{X} = \{\chi \in \mathbf{R}^n_+ : \sum_{i \in S} x_i^2 = 4\}$, the portion of the raidus 2 sphere lying in the positive orthant, endowed with the usual Euclidean metric. It follows that if we use Akin's (1979) isometry to transport the replicator dynamic for the potential game F to the set $\mathcal{X}$, this transported dynamic is a gradient system in the usual Euclidean sense. To conclude the paper, we provide a direct proof of this striking fact, a proof that does not require intermediate steps through differential geometry.

Akin's (1979) transformation, which we denote by $H : \text{int}(\mathbf{R}^n_+) \to \text{int}(\mathbf{R}^n_+)$, is defined by $H_i(x) = 2\sqrt{x_i}$. As we noted earlier, H is a diffeomorphism that maps the simplex X onto the set $\mathcal{X}$. We wish to prove

Theorem 5.3 *Let $F : X \to \mathbf{R}^n$ be a potential game with potential function $f : X \to \mathbf{R}$. Suppose we transport the replicator dynamic for F on $\operatorname{int}(X)$ to the set $\operatorname{int}(\mathcal{X})$ using the transformation H. Then the resulting dynamic is the (Euclidean) gradient dynamic for the transported potential function $\phi = f \circ H^{-1}$.*

Since $H_i(x) = 2\sqrt{x_i}$, the transformation H makes changes in component x_i look large when x_i itself is small. Therefore, Theorem 5.3 tells us that the replicator dynamic is a gradient dynamic on $\operatorname{int}(X)$ after a change of variable that makes changes in the use of rare strategies look important relative to changes in the use of common ones. Intuitively, this reweighting accounts for the fact that under imitative dynamics, both increases and decreases in the use of rare strategies are necessarily slow.

Proof. We prove Theorem 5.3 in two steps: first, we derive the transported version of the replicator dynamic; then we derive the gradient system for the transported version of the potential function, and show that it is the same dynamic on $\mathcal{X}$. The following notation will simplify our calculations: when $y \in \mathbf{R}^n$ and $a \in \mathbf{R}$, we let $[y^a] \in \mathbf{R}^n$ be the vector whose ith component is $(y_i)^a$.

We can express the replicator dynamic on X as

$$\dot{x} = R(x) = \operatorname{diag}(x)\left(F(x) - \mathbf{1}x'F(x)\right) = \left(\operatorname{diag}(x) - xx'\right)F(x).$$

The transported version of this dynamic can be computed as

$$\dot{\chi} = \mathcal{R}(\chi) = DH(H^{-1}(\chi))R(H^{-1}(\chi)).$$

In words: given a state $\chi \in \mathcal{X}$, we first find the corresponding state $x = H^{-1}(\chi) \in X$ and direction of motion $R(x)$. Since $R(x)$ represents a displacement from state x, we transport it to $\mathcal{X}$ by premultiplying it by $DH(x)$, the derivative of H evaluated at x.

Since $\chi = H(x) = 2\,[x^{1/2}]$, the derivative of H at x is given by $DH(x) = \operatorname{diag}([x^{-1/2}])$ Using this fact, we derive a primitive expression for $\mathcal{R}(\chi)$ in terms of $x = H^{-1}(\chi) = \frac{1}{4}[\chi^2]$:

$$\begin{aligned}
\dot{\chi} &= \mathcal{R}(\chi) \\
&= DH(x)R(x) \\
&= \operatorname{diag}([x^{-1/2}])(\operatorname{diag}(x) - xx')F(x) \\
&= \left(\operatorname{diag}([x^{1/2}]) - [x^{1/2}]x'\right)F(x).
\end{aligned} \tag{14}$$

Now, we derive the gradient system on $\mathcal{X}$ generated by $\phi = f \circ H^{-1}$. To compute $\nabla\phi(\chi)$, we need to define an extension of ϕ to all of $\mathbf{R}^n_+$, compute its gradient, and then project the result onto the tangent space of $\mathcal{X}$ at χ. The easiest way to proceed is to let $\tilde{f} : \operatorname{int}(\mathbf{R}^n_+) \to \mathbf{R}$ be an arbitrary C^1 extension of f, and to define the extension $\tilde{\phi} : \operatorname{int}(\mathbf{R}^n_+) \to \mathbf{R}$ by $\tilde{\phi} = \tilde{f} \circ H^{-1}$.

Since $\mathcal{X}$ is a portion of a sphere centered at the origin, the tangent space of $\mathcal{X}$ at χ is the subspace $T_\chi\mathcal{X} = \{z \in \mathbf{R}^n : \chi'z = 0\}$. The orthogonal projection onto this set is represented by the

$n \times n$ matrix

$$P_{T_\chi \mathcal{X}} = I - \frac{1}{\chi'\chi}\chi\chi' = I - \frac{1}{4}\chi\chi' = I - [x^{1/2}][x^{1/2}]'.$$

Also, since $\nabla f(x) = \Phi \nabla \tilde{f}(x)$ by construction, it follows that $\nabla \tilde{f}(x) = \nabla f(x) + c(x)\mathbf{1}$ for some scalar-valued function $c : X \to \mathbf{R}$. Therefore, the gradient system on $\mathcal{X}$ generated by ϕ is

$$\begin{aligned}
\dot{\chi} &= \nabla \phi(\chi) \\
&= P_{T_\chi \mathcal{X}} \nabla \tilde{\phi}(\chi) \\
&= P_{T_\chi \mathcal{X}} DH^{-1}(\chi)' \nabla \tilde{f}(x) \\
&= P_{T_\chi \mathcal{X}} \left(DH(x)^{-1}\right)' \left(\nabla f(x) + c(x)\mathbf{1}\right) \\
&= \left(I - [x^{1/2}][x^{1/2}]'\right) \operatorname{diag}([x^{1/2}]) \left(F(x) + c(x)\mathbf{1}\right) \\
&= \left(\operatorname{diag}([x^{1/2}]) - [x^{1/2}]x'\right) \left(F(x) + c(x)\mathbf{1}\right) \\
&= \left(\operatorname{diag}([x^{1/2}]) - [x^{1/2}]x'\right) F(x).
\end{aligned}$$

This agrees with equation (14), completing the proof of the theorem. ■

In Figure 12, we illustrate Theorem 5.3 with phase diagrams of the transported replicator dynamic $\dot{\chi} = \mathcal{R}(\chi)$ for the three-strategy coordination game from Example 4.3. These phase diagrams on $\mathcal{X}$ are drawn atop contour plots of the transported potential function $\phi(\chi) = (f \circ H^{-1})(\chi) = \frac{1}{32}((\chi_1)^4 + 2(\chi_2)^4 + 3(\chi_3)^4)$. According to Theorem 5.3, the solution trajectories of $\mathcal{R}$ should cross the level sets of ϕ orthogonally.

Looking at Figure 12(a), we find that the crossings look orthogonal at the center of the figure, but not by the boundaries. This is an artifact of our drawing a portion of the sphere in $\mathbf{R}^3$ by projecting it orthogonally onto a sheet of paper.[12] To check whether the crossings near a given state $\chi \in \mathcal{X}$ are truly orthogonal, we must minimize the distortion of angles near χ by making χ the origin of the projection.[13] In the phase diagrams in Figure 12, we mark the projection origins with pink dots. Evidently, the crossings are orthogonal near these points.

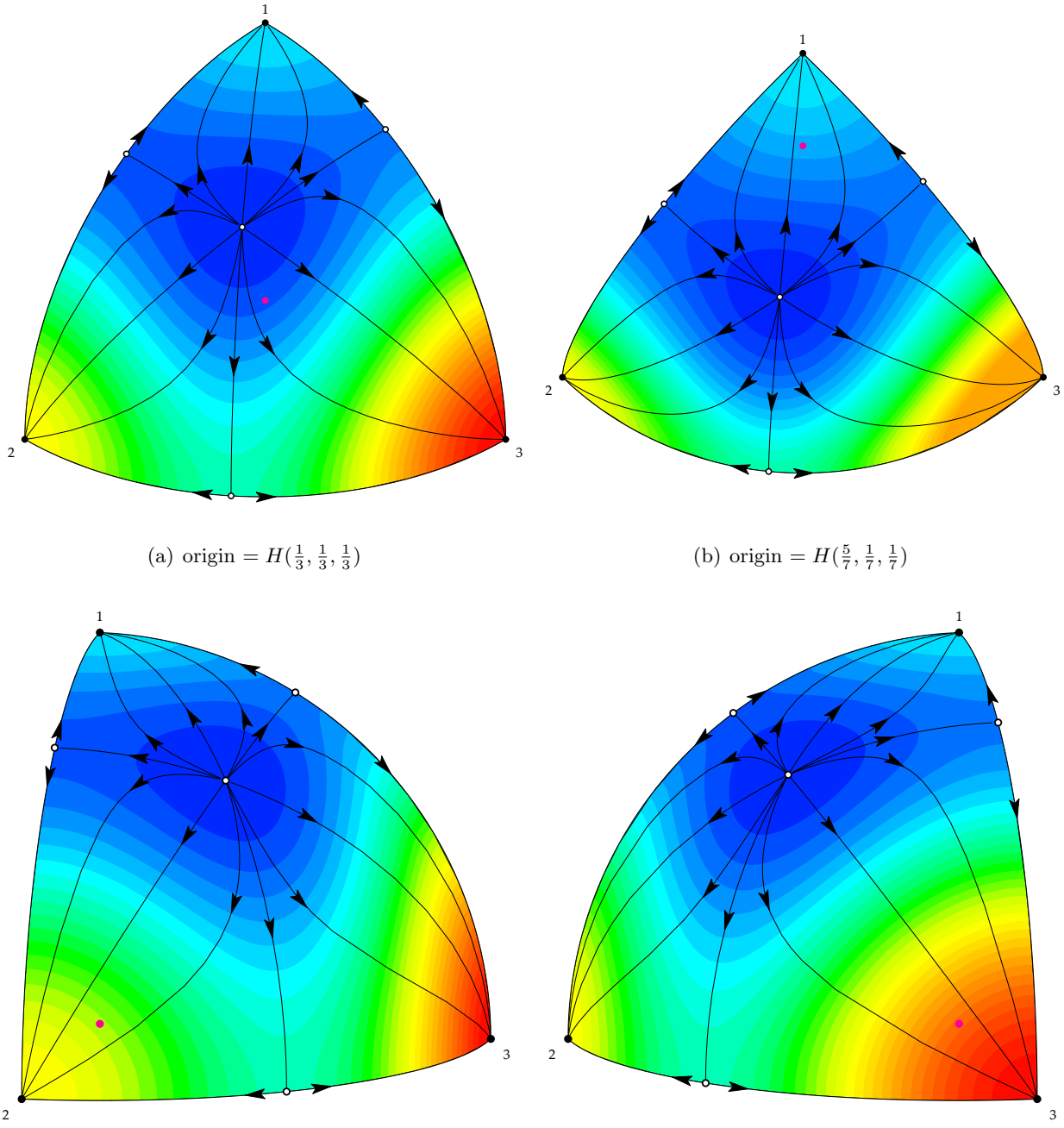
Appendix

A The Proof of Theorem 3.4

In this section, we derive the formula for the projection map $\Pi_{TX^p(x^p)}(v^p)$ stated in Theorem 3.4. To eliminate superscripts, we suppose that $p = 1$. Then Theorem 3.4 is an immediate consequence of the two propositions to follow.

[12] For exactly the same reason, latitude and longitude lines in an orthographic projection of the Earth only appear to cross at right angles in the center of the projection, not on the left and right sides.

[13] The origin of the projection, $o \in \mathcal{X}$, is the point where the sphere touches the sheet of paper. If we view the projection from any point on the ray that exits the sheet of paper orthogonally from o, then the center of the sphere is directly behind o.



(a) origin $= H(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$

(b) origin $= H(\frac{5}{7}, \frac{1}{7}, \frac{1}{7})$

(c) origin $= H(\frac{1}{7}, \frac{5}{7}, \frac{1}{7})$

(d) origin $= H(\frac{1}{7}, \frac{1}{7}, \frac{5}{7})$

Figure 12: The phase diagram of the transported replicator dynamic $\dot{\chi} = \mathcal{R}(\chi)$ for a coordination game. The pink dots represent the positions of the projection origins.

Proposition A.1 *The vector $z^* = \Pi_{TX(x)}(v)$ is defined by*

$$z_i^* = \begin{cases} v_i - \mu & \text{if } x_i > 0, \text{ or if } x_i = 0 \text{ and } v_i > \mu, \\ 0 & \text{if } x_i = 0 \text{ and } v_i \leq \mu \end{cases} \tag{15}$$

for some $\mu = \mu(v, x)$.

Proof. Let $Y = \{i \in S : x_i > 0\}$ and $N = \{i \in S : x_i = 0\}$ denote the sets of strategies that are used and unused at x. Then z^* is the solution to the following quadratic program:

$$\min_{z \in \mathbf{R}^n} \tfrac{1}{2} \sum_{i \in S} (z_i - v_i)^2 \quad \text{s.t.} \quad \sum_{i \in S} z_i = 0$$
$$\text{and } z_i \geq 0 \text{ for all } i \in N.$$

The Lagrangian for this program is

$$L(z, \mu, \gamma) = -\tfrac{1}{2} \sum_{i \in S} (z_i - v_i)^2 - \mu \sum_{i \in S} z_i + \sum_{j \in N} \lambda_j z_j,$$

so the Kuhn-Tucker first order conditions are

$$z_i^* = v_i - \mu \qquad \text{for all } i \in Y, \tag{16}$$
$$z_j^* = v_j - \mu + \lambda_j \qquad \text{for all } j \in N, \tag{17}$$
$$\lambda_j \geq 0 \qquad \text{for all } j \in N, \text{ and} \tag{18}$$
$$\lambda_j z_j^* = 0 \qquad \text{for all } j \in N. \tag{19}$$

If we let $N^+ = \{i \in N : z_i^* > 0\}$ and $N^0 = \{j \in N : z_j^* = 0\}$ we can rewrite the Kuhn-Tucker conditions (16-19) as

$$z_i^* = v_i - \mu > 0 \qquad \text{for all } i \in Y \cup N^+, \tag{20}$$
$$v_j^* = z_j - \mu + \lambda_j = 0 \qquad \text{for all } j \in N^0, \text{ and} \tag{21}$$
$$\lambda_j \geq 0 \qquad \text{for all } j \in N^0. \tag{22}$$

It follows immediately from equations (20-22) that

$$N^+ = \{i \in N : v_i > \mu\} \text{ and } N^0 = \{i \in N : v_i \leq \mu\}. \tag{23}$$

Equations (20-22) and (23) together yield equation (15). ■

To complete our description of $z^* = \Pi_{TX(x)}(v)$, and hence of the projection dynamic (P), we need to determine the value of $\mu = \mu(x, v)$. To accomplish this, we permute the names of the strategies in S so that strategies 1 through $n_Y = \#Y$ are in Y and strategies $n_Y + 1$ through n are

in N, with the latter strategies ordered so that $v_{n_Y+1} \geq v_{n_Y+2} \geq \cdots \geq v_n$. For $k \in \{1, \ldots, n\}$, let

$$\mu_k = \frac{1}{k} \sum_{i=1}^{k} v_i$$

be the average of the first k components of v. Throughout the analysis, we make use of the recursion

$$\mu_{k+1} = \frac{1}{k+1} v_{k+1} + \frac{k}{k+1} \mu_k,$$

which tells us that μ_{k+1} and v_{k+1} deviate from μ_k in the same direction. Finally, to simplify the statement of the result to follow, we set $\mu_{n+1} = -\infty$.

Proposition A.2 *The Lagrange multiplier μ is equal to μ_{k^*}, where*

$$k^* = \min\{k \in \{n_Y, \ldots, n\} : \mu_{k+1} \leq \mu_k\}.$$

Therefore, the set of unused strategies with positive components in z^* is $N^+ = \{n_Y + 1, \ldots, k^*\}$.

Proof. To prove the proposition, it is enough to show that μ_{k^*} satisfies

$$v_i > \mu_{k^*} \text{ for all } i \in N^+, \text{ and}$$
$$\mu_{k^*} \geq v_j \text{ for all } j \in N^0.$$

Therefore, by our ordering assumption, we need only show that

$$\text{if } k^* > n_Y, \text{ then } v_{k^*} > \mu_{k^*}, \text{ and} \tag{24}$$
$$\text{if } k^* < n, \text{ then } \mu_{k^*} \geq v_{k^*+1}. \tag{25}$$

We divide the analysis into four cases. If $n_Y = n$, then $k^* = n$ as well, and conditions (24) and (25) are vacuous. (In this case, we obtain $\mu = \mu_n = \frac{1}{n} \sum_{i=1}^{n} v_i$ and $z^* = \Phi v$, as expected.) Therefore, for the remainder of the proof we can suppose that $n_Y < n$.

If $k^* = n_Y$, then condition (24) is vacuous. Since $\mu_{k^*+1} \leq \mu_{k^*}$ by the definition of k^*, it follows from the recursion that $v_{k^*+1} \leq \mu_{k^*}$, which is condition (25).

If $k^* \in \{n_Y, \ldots, n-1\}$, then the definition of k^* tells us that $\mu_{k^*} > \mu_{k^*-1}$. It then follows from the recursion that $v_{k^*} > \mu_{k^*-1}$, and hence that

$$v_{k^*} > \frac{1}{k^*} v_{k^*} + \frac{k^* - 1}{k^*} \mu_{k^*-1} = \mu_{k^*},$$

which is condition (24). The definition of k^* also tells us that $\mu_{k^*+1} \leq \mu_{k^*}$; the recursion then reveals that $v_{k^*+1} \leq \mu_{k^*}$, which is condition (25).

Finally, if $n_Y < n$ and $k^* = n$, then condition (25) is vacuous, and repeating the corresponding proof from the previous case gives us condition (24). ■

B The Proof of Theorem 4.2

We first sketch a proof of existence of solutions to (P) due to Aubin and Cellina (1984). Define the multivalued map $V : X \to \mathbf{R}^n$ by

$$V(x) = \bigcap_{\varepsilon>0} \operatorname{cl}\left(\operatorname{conv}\left(\bigcup_{y\in X:|y-x|\leq\varepsilon} \Pi_{TX(y)}(F(y))\right)\right).$$

In words, $V(x)$ is the closed convex hull of all values of $\Pi_{TX(y)}(F(y))$ that obtain at points y arbitrarily close to x. It is easy to check that V is upper hemicontinuous with closed convex values. Moreover, $V(x) \cap TX(x)$, the set of feasible directions of motion from x contained in $V(x)$, is always equal to $\{\Pi_{TX(x)}(F(x))\}$, and so in particular is nonempty. Because $V(x) \cap TX(x) \neq \emptyset$, the Viability Theorem for differential inclusions ((Aubin and Cellina (1984, Theorem 4.2.1) implies that for each $\xi \in X$, a Carathéodory solution $\{x_t\}_{t\geq 0}$ to $\dot{x} \in V(x)$ exists. But since $V(x) \cap TX(x) = \{\Pi_{TX(x)}(F(x))\}$, this solution must also solve the original equation (P).

Our proof of uniqueness and Lipschitz continuity of solutions to (P) follows Cojocaru and Jonker (2004). Let $\{x_t\}$ and $\{y_t\}$ be solutions to (P). Using the chain rule, Theorem 3.2, and the Lipschitz continuity of F, we see that

$$\begin{aligned}
\tfrac{d}{dt}\left|y_t - x_t\right|^2 &= 2(y_t - x_t)'(\Pi_{TX(y_t)}(F(y_t)) - \Pi_{TX(x_t)}(F(x_t))) \\
&= 2(y_t - x_t)'(F(y_t) - F(x_t)) - 2(y_t - x_t)'(\Pi_{NX(y_t)}(F(y_t)) - \Pi_{NX(x_t)}(F(x_t))) \\
&= 2(y_t - x_t)'(F(y_t) - F(x_t)) + 2(x_t - y_t)'\Pi_{NX(y_t)}(F(y_t)) \\
&\quad + 2(y_t - x_t)'\Pi_{NX(x_t)}(F(x_t)) \\
&\leq 2(y_t - x_t)'(F(y_t) - F(x_t)) \\
&\leq 2K\left|y_t - x_t\right|^2,
\end{aligned}$$

and hence that

$$\left|y_t - x_t\right|^2 \leq \left|y_0 - x_0\right|^2 + \int_0^t 2K\left|y_s - x_s\right| ds.$$

Gronwall's inequality then implies that

$$\left|y_t - x_t\right|^2 \leq \left|y_0 - x_0\right|^2 e^{2Kt}.$$

Taking square roots yields the inequality stated in the theorem. ■

C The Proof of Theorem 5.1

We begin with the proof of part (i). Observe that the weighted average of the components of the excess payoff vector $\hat{F}(x)$ is 0:

$$\sum_{k\in S} x_k \hat{F}_k(x) = \sum_{k\in S} x_k \left(F_k(x) - \sum_{l\in S} x_l F_l(x) \right) = \sum_{k\in S} x_k F_k(x) - \sum_{l\in S} x_l F_l(x) = 0.$$

It follows directly that

$$\sum_{k\in S} x_k [\hat{F}_k(x)]_+ = \sum_{k\in S} x_k [\hat{F}_k(x)]_-. \tag{26}$$

Now fix a state $x \in X$. If both sides of equation (26) equal 0 at x, then the mean dynamic generated by protocol (9) has a rest point at x; but since in this case $x_k \hat{F}_k(x) = 0$ for all $k \in S$, the dynamic (R) has a rest point at x as well. On the other hand, if both sides of equality (26) are positive at state x, we can use this equality to compute as follows:

$$\begin{aligned}
\dot{x}_i &= \sum_{j\in S} x_j \rho_{ji} - x_i \sum_{j\in S} \rho_{ij} \\
&= \sum_{j\in S} x_j \left([\hat{F}_j(x)]_- \frac{x_i [\hat{F}_i(x)]_+}{\sum_{k\in S} x_k [\hat{F}_k(x)]_+} \right) - x_i \sum_{j\in S} \left([\hat{F}_i(x)]_- \frac{x_j [\hat{F}_j(x)]_+}{\sum_{k\in S} x_k [\hat{F}_k(x)]_+} \right) \\
&= x_i [\hat{F}_i(x)]_+ - x_i [\hat{F}_i(x)]_- \\
&= x_i \hat{F}(x),
\end{aligned}$$

This too agrees with equation (R), completing the proof of part (i).

We now prove part (ii). Write $\mathcal{S}$ for $\mathcal{S}(F(x), x)$. Since

$$\sum_{k\in\mathcal{S}} \tilde{F}_k^{\mathcal{S}}(x) = \sum_{k\in\mathcal{S}} \left(F_k^{\mathcal{S}}(x) - \frac{1}{\#\mathcal{S}} \sum_{l\in\mathcal{S}} F_l^{\mathcal{S}}(x) \right) = 0,$$

we see that

$$\sum_{k\in\mathcal{S}} [\tilde{F}_k^{\mathcal{S}}(x)]_+ = \sum_{k\in\mathcal{S}} [\tilde{F}_k^{\mathcal{S}}(x)]_-. \tag{27}$$

Also, we note these two implications of Theorem 3.4:

$$\text{if } j \in \mathcal{S} \text{ and } x_j = 0, \text{ then } [\tilde{F}_j^{\mathcal{S}}(x)]_- = 0; \tag{28}$$

$$\text{if } j \notin \mathcal{S}, \text{ then } [\tilde{F}_j^{\mathcal{S}}(x)]_+ = 0. \tag{29}$$

Fix a state $x \in X$. If both sides of equation (27) equal 0 at x, then the mean dynamic generated by protocol (10) has a rest point at x; but since in this case $\tilde{F}_k^{\mathcal{S}}(x) = 0$ for all $k \in \mathcal{S}$, the dynamic (P) has a rest point at x as well.

Suppose instead that both sides of equation (27) are positive at state x. If $x_i > 0$, we can use

equations (28), (29), and (27) to compute as follows:

$$\begin{aligned}
\dot{x}_i &= \sum_{j\in S} x_j\rho_{ji} - x_i\sum_{j\in S}\rho_{ij} \\
&= \sum_{j:\,x_j>0} x_j\left(\frac{[\tilde{F}_j^{\mathcal{S}}(x)]_-}{x_j}\cdot\frac{[\tilde{F}_i^{\mathcal{S}}(x)]_+}{\sum_{k\in\mathcal{S}}[\tilde{F}_k^{\mathcal{S}}(x)]_+}\right) - x_i\sum_{j\in S}\left(\frac{[\tilde{F}_i^{\mathcal{S}}(x)]_-}{x_i}\cdot\frac{[\tilde{F}_j^{\mathcal{S}}(x)]_+}{\sum_{k\in S}[\tilde{F}_k^{\mathcal{S}}(x)]_+}\right) \\
&= \left(\frac{\sum\limits_{j:\,x_j>0}[\tilde{F}_j^{\mathcal{S}}(x)]_-}{\sum\limits_{k\in\mathcal{S}}[\tilde{F}_k^{\mathcal{S}}(x)]_+}\right)[\tilde{F}_i^{\mathcal{S}}(x)]_+ - \left(\frac{\sum\limits_{j\in S}[\tilde{F}_j^{\mathcal{S}}(x)]_+}{\sum\limits_{k\in\mathcal{S}}[\tilde{F}_k^{\mathcal{S}}(x)]_+}\right)[\tilde{F}_i^{\mathcal{S}}(x)]_- \\
&= \left(\frac{\sum\limits_{j\in\mathcal{S}}[\tilde{F}_j^{\mathcal{S}}(x)]_-}{\sum\limits_{k\in\mathcal{S}}[\tilde{F}_k^{\mathcal{S}}(x)]_+}\right)[\tilde{F}_i^{\mathcal{S}}(x)]_+ - \left(\frac{\sum\limits_{j\in\mathcal{S}}[\tilde{F}_j^{\mathcal{S}}(x)]_+}{\sum\limits_{k\in\mathcal{S}}[\tilde{F}_k^{\mathcal{S}}(x)]_+}\right)[\tilde{F}_i^{\mathcal{S}}(x)]_- \\
&= [\tilde{F}_i^{\mathcal{S}}(x)]_+ - [\tilde{F}_i^{\mathcal{S}}(x)]_- \\
&= \tilde{F}^{\mathcal{S}}(x).
\end{aligned}$$

This agrees with equation (P). If $x_i = 0$, then the second term in the previous calculation drops out immediately, and the calculation of the first term shows that $\dot{x}_i = [\tilde{F}_i^{\mathcal{S}}(x)]_+$, again in agreement with equation (P). This completes the proof of the theorem. ■

D The Proof of Theorem 5.2

The method used to construct the approximate closed orbit of (P) is described in the text after the statement of the theorem. Here, we verify that this approximation implies the existence of an exact closed orbit of (P). A minor modification of our argument shows that this orbit absorbs all nearby trajectories in finite time.

Let us review the construction the approximate closed orbit. We begin by choosing the initial state $\xi^0 = \xi = (\frac{7}{15}, \frac{8}{15}, 0, 0)$. The (exact) solution to (P) from ξ initially travels through face RPS in a fashion described by the linear differential equation (2), and so be computed analytically. The solution exits face RPS into the interior of X when it hits the line on which the payoff to T equals the average of the payoffs to R, P, and S. The exit point cannot be determined analytically, but it can be approximated to any desired degree of accuracy. We call this approximate exit point, which we compute to 16 decimal places, $\xi^1 \approx (.446354, .552864, .000782, 0)$, and we call the time that the solution to (P) reaches this point $t^1 \approx .015678$.

Next, we consider the (exact) solution to (P) from starting from state ξ^1. This solution travels through $\text{int}(X)$ until it reaches face PST. We again compute an approximate exit point ξ^2, and we let t^2 be the total time expended during the first two solution segments. Continuing in this fashion, we compute the approximate exit points $\xi^3, \ldots, \xi^7$, and the transition times $t^3, \ldots, t^7$.

Now for each state ξ^k, let $\{x_t^k\}_{t\geq t^k}$ be the solution to (P) that starts from state ξ^k at time t^k.

Because solutions to (P) are Lipschitz continuous in their initial conditions (Theorem 4.2), we can bound the distance between state $x^0_{t^7}$, which is the location of the solution to (P) from state $\xi^0 = \xi$ at time t^7, and state $x^7_{t^7} = \xi^7$, as follows:

$$|x^0_{t^7} - x^7_{t^7}| \leq \sum_{k=1}^{7} |x^{k-1}_{t^7} - x^k_{t^7}| \leq \sum_{k=1}^{7} e^{K(t_7 - t_k)}.$$

Here, K is the Lipschitz coefficient for the payoff vector field F, and ε is an upper bound on the roundoff error introduced when we compute the approximate exit point ξ^k for the solution to (P) from state ξ^{k-1}.

Since $F(x) = Ax$ is linear, its Lipschitz coefficient is the spectral norm of the payoff matrix A: that is, the square root of the largest eigenvalue of $A'A$ (see Horn and Johnson (1985)). A computation reveals that the spectral norm of A is approximately 5.718145. Since we compute our approximate exit points to 16 decimal places, our roundoff errors are no greater than 5×10^{-17}. Thus, since $t^7 - t^1 = 1.049900$, we obtain the following bound on the distance between $x^0_{t^7}$ and $x^7_{t^7}$:

$$|x^0_{t^7} - x^7_{t^7}| \leq 7\left(e^{K(t_7 - t_k)}\varepsilon\right) \approx 7\left(e^{(5.718145)(1.049900)} \times (5 \times 10^{-17})\right) \approx 1.416920 \times 10^{-13}.$$

It is easy to verify that any solution to (P) that starts within this distance of state $\xi^7 \approx (.693072, .306928, 0, 0)$ will hit edge RP between vertex R and state ξ, and so continue on to ξ. We therefore conclude that $\{x^0_t\}_{t \geq 0}$, the exact solution to (P) starting from state ξ, must return to state ξ. This completes the proof of the theorem. ■