Chapter 1

Payoff-Independent Action Update for Continuous Action Social Dilemmas: a Preliminary Investigation

Ath. Kehagias
Department of Electrical and Computer Engineering,
Faculty of Engineering, Aristotle University
Thessaloniki, Greece
kehagiat@ece.auth.gr

In this paper we introduce an *action update* model for two-player *continuous actions social dilemma* games. The model applies to continuous action versions of social dilemma games such as *Prisoner's Dilemma*, *Hawk and Dove* , *Stag Hunt* etc. In our formulation, action updates depend on the current *cooperation levels* of the two players, *independently of specific payoffs*. This means that the *same* update equations can be applied to each of the abovementioned games. We present a preliminary investigation, limited to a particular action update model. As we will explain in the sequel, this model admits a large number of modifications and extensions and in fact it belongs to a more general family which will be further explored in future publications.

1. Introduction

In this paper we introduce an *action update* model for two-player *continuous actions social dilemma* games as defined by Dawes and Lange.[6,7,14] The model applies to continuous action versions of social dilemma games such as *Prisoner's Dilemma*,[21] *Hawk and Dove*,[28] *Stag Hunt*[27] etc. In our formulation, action updates depend on the current *cooperation levels* of the two players, *independently of specific payoffs*. This means that the *same* update equations can be applied to each of the abovementioned games. We present a preliminary investigation, limited to a particular action update model. As we will explain in the sequel, this model admits a large number of modifications and extensions and in fact it belongs to a more general family which will be further explored in future publications.

The seminal papers on social dilemmas have been authored by Dawes[6] and Liebrand,[16] which have been followed by an extensive literature; for example, see Dawes[7] and, for a recent list of references, the book by Lange.[15] The study of continuous action games forms a basic branch of game theory[3,18,20] and provides the tools to study continuous action social dilemmas. For instance, several authors[8,9,31] have used differential equations to model the evolution of actions over time. These equations are actually quite similar to the ones used to model the evolution of *probabilities of discrete actions*;[4,21,26] in addition differential equations (the *replicator* equations) are also used in *evolutionary* game theory[10,11,25,30] to model the evolution of action *frequencies* in a population.

The paper is organized as follows. In Section 2 we present mathematical preliminaries regarding differential equations and game theory. In Section 3 we discuss social dilemma games with both discrete and continuous actions. Section 4 is devoted to the introduction and analysis of our basic action update model for a continuous action social dilemma game. In Section 5 we briefly present various modifications and generalizations of the basic model. In Section 6 we study an inverse problem related to the selection of update coefficients such that the update equations have a prescribed equilibrium. Finally, in Section 7 we summarize and present future research directions.

2. Preliminaries

2.1. *Differential Equations*

Our model involves continuous actions evolving in continuous time. Consequently, our main tool will be vector differential equations of the form

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}\left(\mathbf{x}\right).$$

Here $\mathbf{x}, \mathbf{f}\left(\mathbf{x}\right) \in \mathbb{R}^N$ (for some $N \in \mathbb{N}$, usually $N = 2$); in other words

$$\mathbf{x} = \left(x_1, ..., x_N\right) \text{ and } \mathbf{f}\left(\mathbf{x}\right) = \left(f_1\left(\mathbf{x}\right), ..., f_N\left(\mathbf{x}\right)\right).$$

Notation 1. We will denote the norm of any $\mathbf{z} \in \mathbb{R}^N$ by $|\mathbf{z}| = \left(\sum_{n=1}^{N} z_n^2\right)^{1/2}$.

We will always assume that $\mathbf{f}$ satisfies appropriate conditions to ensure the existence of a unique solution of the problem

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}\left(x\right), \quad \mathbf{x}\left(0\right) = \mathbf{x}_0. \tag{1}$$

Hence the following notation is meaningful.

Notation 2. We denote the unique solution of (1) by $\mathbf{x}\left(t|\mathbf{x}_0\right)$.

Definition 1. We say that the set $S \subseteq \mathbb{R}^n$ is an *invariant set* of the DE $\frac{d\mathbf{x}}{dt} = \mathbf{f}\left(\mathbf{x}\right)$ (with $\mathbf{x} \in \mathbb{R}^N$) iff we have

$$\mathbf{x}_0 \in S \Rightarrow \left(\forall t \geq 0 : \mathbf{x}\left(t|\mathbf{x}_0\right) \in S\right).$$

Definition 2. We say that $\overline{\mathbf{x}} \in \mathbb{R}^n$ is an *invariant point* or an *equilibrium* of the DE $\frac{d\mathbf{x}}{dt} = \mathbf{f}\left(\mathbf{x}\right)$ (with $\mathbf{x} \in \mathbb{R}^N$) iff we have

$$\mathbf{x}_0 = \overline{\mathbf{x}} \Rightarrow \left(\forall t \geq 0 : \mathbf{x}\left(t|\mathbf{x}_0\right) = \overline{\mathbf{x}}\right).$$

Remark 1. Clearly $\overline{\mathbf{x}}$ is an equilibrium of $\frac{d\mathbf{x}}{dt} = f(\mathbf{x})$ iff $\{\overline{\mathbf{x}}\}$ is an invariant set of $\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x})$.

Proposition 1. The point $\overline{\mathbf{x}}$ is an equilibrium of $\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x})$ iff $\overline{\mathbf{x}}$ is a solution of the algebraic system $\mathbf{f}(\mathbf{x}) = 0$.

Definition 3. We say that $\overline{\mathbf{x}} \in \mathbb{R}^n$ is a *stable equilibrium* of the DE $\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x})$ iff

$$\forall \varepsilon > 0 : \exists \delta_\varepsilon > 0 : |x_0 - \overline{x}| < \delta_\varepsilon \Rightarrow (\forall t \geq 0 : |\mathbf{x}(t|\mathbf{x}_0) - \overline{\mathbf{x}}| < \varepsilon).$$

Otherwise we say that $\overline{\mathbf{x}} \in \mathbb{R}^n$ is an *unstable equilibrium.*

Definition 4. Given the DE system $\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x})$ where $\mathbf{x} \in \mathbb{R}^N$, the *Jacobian* matrix $J(\mathbf{x})$ is defined by $J_{ij}(\mathbf{x}) = \frac{\partial f_i}{\partial x_j}(\mathbf{x})$ for $i, j \in \{1, ..., N\}$.

Proposition 2. Given the DE system $\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x})$ where $\mathbf{x} \in \mathbb{R}^N$, assume it has an equilibrium $\overline{\mathbf{x}}$ and $\mathbf{f}$ has continuous first derivatives with respect to $x_1, ..., x_N$ in a neighborhood of $\overline{\mathbf{x}}$. Let $\lambda_1, ..., \lambda_N$ be the eigenvalues of $J(\overline{\mathbf{x}})$. If $\mathrm{Re}(\lambda_n) < 0$ for all $n \in \{1, ..., N\}$ then $\overline{\mathbf{x}}$ is a stable equilbrium; otherwise (i.e., if there exists some n such that $\mathrm{Re}(\lambda_n) \geq 0$) then $\overline{\mathbf{x}}$ is an unstable equilibrium.

Definition 5. Given the DE system $\frac{d\mathbf{x}}{dt} = f(\mathbf{x})$, assume it has an equilibrium $\overline{\mathbf{x}}$. We define the *basin of attraction of* $\overline{\mathbf{x}}$ by

$$U(\overline{\mathbf{x}}) = \left\{ \mathbf{z} : \lim_{t \to \infty} \mathbf{x}(t|\mathbf{z}) = \overline{\mathbf{x}} \right\}.$$

Theorem 1 (Petrovitsch[19]). If the scalar functions $y(t)$ and $z(t)$ satisfy

$$\frac{dy}{dt} = f(t, y(t)), \qquad \frac{dz}{dt} > f(t, z(t))$$

then the following inequality holds

$$\forall c, \forall t \in (0, t_1] : z(t|c) > y(t|c).$$

2.2. *Bimatrix Games*

A *bimatrix game*[3,18] is a two-player game in which each player has a finite number of possible actions and the players choose their action simultaneously. The name comes from the fact that the "*normal form*" of such a game can be described by two matrices: matrix A^1 (resp. A^2) describes the *payoffs* of the first player P_1 (resp. the second player P_2). If P_1 has M_1 possible actions and P_2 has M_2 possible actions (in which case we will speak of a "*finite actions game*") then A^1 and A^2 are $M_1 \times M_2$ matrices. When P_1 selects his m-th action and P_2 selects his n-th action, the payoff to P_1 (resp. P_2) is A^1_{mn} (resp. A^2_{mn}). These payoffs can also be combined into a *bimatrix*

$$A = \left(A^1, A^2\right) = \begin{bmatrix} \left(A^1_{11}, A^2_{11}\right) & \left(A^1_{12}, A^2_{21}\right) \\ \left(A^1_{21}, A^2_{21}\right) & \left(A^1_{22}, A^2_{22}\right) \end{bmatrix}.$$

Definition 6. A (*mixed*) *strategy* for P_1 (resp. for P_2) is a non-negative vector $\mathbf{p}^1 = \left(p_1^1, ..., p_{M_1}^1\right)$ (resp. $\mathbf{p}^2 = \left(p_1^2, ..., p_{M_2}^2\right)$) such that $\sum_{m=1}^{M_1} p_m^1 = 1$ (resp. $\sum_{n=1}^{M_2} p_n^2 = 1$) where

$$\begin{aligned}\forall m : p_m^1 &= \Pr\left(P_1 \text{ plays his } m\text{-th action}\right),\\ \forall n : p_n^2 &= \Pr\left(P_2 \text{ plays his } n\text{-th action}\right).\end{aligned}$$

If $\mathbf{p}^i$ $(i \in \{1,2\})$ is such that there exists some $\overline{m}$ satisfying $p_{\overline{m}}^i = 1$ (and hence, for all $m \neq \overline{m}$, $p_m^i = 0$) then we say that $\mathbf{p}^i$ is a *pure strategy*.

Notation 3. When the players play mixed strategies $\mathbf{p}^1$ and $\mathbf{p}^2$, the expected payoffs for P_1 and P_2 are

$$Q_n\left(\mathbf{p}^1, \mathbf{p}^2\right) = \sum_{m=1}^{M_1} \sum_{n=1}^{M_2} p_m^1 A_{mn}^n p_n^2. \tag{2}$$

When P_1 (resp. P_2) plays a pure strategy $\mathbf{p}^1$ (resp. $\mathbf{p}^2$) concentrating all probability on action m_1 (resp. m_2), we will also write $Q\left(m_1, m_2\right)$ instead of $Q_n\left(\mathbf{p}^1, \mathbf{p}^2\right)$.

Definition 7. A *Nash equilibrium* (NE) of the bimatrix game $\left(A^1, A^2\right)$ is a pair of mixed strategies $\left(\widehat{\mathbf{p}}^1, \widehat{\mathbf{p}}^2\right)$ such that

$$\forall \mathbf{p}^1 \in \Sigma_1 : Q_1\left(\widehat{\mathbf{p}}^1, \widehat{\mathbf{p}}^2\right) \geq Q_1\left(\mathbf{p}^1, \widehat{\mathbf{p}}^2\right) \quad \text{and} \quad \forall \mathbf{p}^2 \in \Sigma_2 : Q_2\left(\widehat{\mathbf{p}}^1, \widehat{\mathbf{p}}^2\right) \geq Q_2\left(\widehat{\mathbf{p}}^1, \mathbf{p}^2\right).$$

The interpretation is that when both players have committed to a NE $\left(\widehat{\mathbf{p}}^1, \widehat{\mathbf{p}}^2\right)$, neither player gains an advantage by *unilaterally* changing his strategy.

Proposition 3 (Nash[18]). Every bimatrix game has a Nash equilibrium in (possibly) mixed strategies.

It is worth noting that in a bimatrix game the interests if the two players are not, in general, diametrically opposed. In other words, one player's gain is not necessarily the other player's loss. The special case of completely opposed interests occurs in so-called *zero-sum* games, i.e., when $A^2 = -A^1$. But in a general (i.e., nonzero-sum) bimatrix game there is scope for *cooperation* between the players.

3. Social Dilemmas

In this paper we follow Liebrand:[16] a social dilemma is "*defined as a [game] in which (1) there is a strategy that yields the person the best payoff in at least one configuration of strategy choices and that has a negative impact on the interests of the other persons involved, and (2) the choice of that particular strategy by all persons results in a deficient outcome*". We now introduce some very simple social dilemma games which will be the focus of our investigation in the current paper.

3.1. *Two Players, Two Actions*

The simplest form of social dilemma games emerges when we consider two-action bimatrix games. Since there exists an infinite number of bimatrices, there also exists an infinite number of bimatrix games. However the essential characteristics of such games have been elucidated by Rapoport[22,23] and Liebrand,[16] by concentrating on the *ordering* rather than the *magnitude* of the payoffs. In particular, a family of 78 "prototypical" games is defined by Rapoport[22] and, as explained by Liebrand,[16] exactly three of these games are *symmetric*[a] *social dilemmas.* These are the games determined by the following bimatrices[b]:

Prisoner's Dilemma (PD)	**Hawk and Dove(HD)**	**Stag Hunt(SH)**
$A = \begin{bmatrix} (3,3) & (1,4) \\ (4,1) & (2,2) \end{bmatrix}$	$A = \begin{bmatrix} (3,3) & (2,4) \\ (4,2) & (1,1) \end{bmatrix}$	$A = \begin{bmatrix} (4,4) & (1,3) \\ (3,1) & (2,2) \end{bmatrix}$

The names of the above games are obtained from accompanying "back-stories" giving illustrative examples of situations which can be modeled by the respective games. We omit these stories (the interested reader can consult Rapoport[21] for Prisoner's Dilemma, Smith[28] for Hawk and Dove, Skyrms[27] for Stag Hunt) and proceed to a discussion of the similarities and differences between the three games.

For reasons which will soon become apparent, we will denote the first action of each player by C (for cooperation) and the second by D (for defection). For example, supposing that P_1 chooses C and P_2 chooses D, the payoff pair is: $(1,4)$ in PD, $(2,4)$ in HD and $(1,3)$ in SH. Now let us briefly discuss the dilemma involved in each game.

(1) **Prisoner's Dilemma**: In this game each player should always play D irrespective of the other player's action. To see this, suppose P_2 plays C: if P_1 plays C his payoff is 3 and if he plays D his payoff is $4 > 3$; similarly, suppose P_2 plays D: if P_1 plays C his payoff is 1 and if he plays D his payoff is $2 > 1$. Hence, in every case C is better than D for P_1; by symmetry the same holds for P_2. More generally, it is easy to prove that neither player gains anything by playing C with positive probability. In short, $\left(\widehat{\mathbf{p}}^1, \widehat{\mathbf{p}}^2\right) = ((0,1),(0,1))$ is the *unique* NE of PD. The "dilemma" or paradox is that both players would be better off if they played C (cooperation); in this case they would get

$$Q_1((1,0),(1,0)) = Q_2((1,0),(1,0)) = 3$$
$$Q_1((0,1),(0,1)) = Q_2((0,1),(0,1)) = 2$$

[a] I.e., A^2 is the transpose of A^1, resulting in symmetric payoffs for the two players.

[b] These are the simplest versions of the above games. We can obtain variants by changing the payoff values, as long as the ordering of the payoffs is preserved. For example the following bimatrix also describes a Prisoner's dilemma

$$\widehat{A} = \begin{bmatrix} (7,7) & (0,8) \\ (8,0) & (4,4) \end{bmatrix}$$

because the inequalities $A^1_{1,2} < A^1_{2,2} < A^1_{1,1} < A^1_{2,1}$ are preserved as $\widehat{A}^1_{1,2} < \widehat{A}^1_{2,2} < \widehat{A}^1_{1,1} < \widehat{A}^1_{2,1}$.

and $3 > 2$; however, $\left(\widetilde{\mathbf{p}}^1, \widetilde{\mathbf{p}}^2\right) = \left((1,0), (1,0)\right)$ is *not a* NE.

(2) **Hawk and Dove**: In this game each player's best payoff is 4, obtained when he plays D and the other player plays C; however, if both players play D each ends up with the worst possible payoff, namely 1. It turns out that the game has three NE: $((1,0),(0,1))$ with payoffs $(2,4)$, $((0,1),(1,0))$ with payoffs $(4,2)$, and the mixed strategies pair $\left(\left(\frac{1}{2},\frac{1}{2}\right),\left(\frac{1}{2},\frac{1}{2}\right)\right)$ with payoffs $\left(\frac{5}{2},\frac{5}{2}\right)$.

(3) **Stag Hunt**: In this game each player's best payoff is 4, when he plays C and the other player also plays C; this is a NE. However, if both players play D this is also a NE, in which each player plays D and receives a payoff of 2. So we have two pure NE: $((1,0),(1,0))$ with payoffs $(4,4)$ and $((0,1),(0,1))$ with payoffs $(2,2)$. It turns out there exists also a NE in mixed strategies: $\left(\left(\frac{1}{2},\frac{1}{2}\right),\left(\frac{1}{2},\frac{1}{2}\right)\right)$ with payoffs $\left(\frac{5}{2},\frac{5}{2}\right)$.

In each of the above games both players would be better off if they both played C (i.e., if they *cooperated*). The "social dilemma" is that this fact is not supported by the observed NE: in both PD and HD, (C,C) is not a NE at all; in SH (C,C) is a NE but there exist additional "deficient" NE[c]. In simpler terms, while both players would be better off under mutual cooperation, each is "tempted" to defect. This raises serious questions regarding the foundations of game theoretic analysis, in particular the concept of Nash equilibrium. Consequently a vast literature exists on social dilemmas (especially the Prisoner's Dilemma) and various approaches have been introduced to justify the "*emergence of cooperation*".[1,2,17]

An important observation, which is also supported by experimental study of actual games played between humans[12,32,33] is that, in many cases, what determines the outcome of a game is not so much the exact payoff values but the "attitude" of the players. For example, mutual cooperation tends to reinforce itself: in a game of PD, when both players play C once, they are more likely to play C again in future replays of the game (despite the fact that equilibrium analysis suggests the D is better). The same has been observed for mutual defection: players who repeatedly and mutually defect may in time recognize that defection generates further *mutual* defection resulting in mutual disadvantage and hence may switch to mutual cooperation. We will return to these considerations a little later.

3.2. *Two Players, Continuous Actions*

A *continuous action* game is one in which each player can choose from a continuum of actions. In the context of social dilemmas, this approach is quite appropriate to model situations which are characterized by "degrees of cooperation" (e.g., how much private income to contribute to a common cause).[9,13,24,31] For example, P_n's action can be a number $x_n \in [0,1]$; a larger value indicates higher cooperation (hence 0 indicates full defection and 1 indicates full cooperation).

[c] By "deficient" we mean that both players are worse off (in terms of their payoffs) than if they played (C,C).

"Continuous social dilemmas" can be obtained from the corresponding bimatrix games by interpolation. Recall that in a bimatrix game we have

$$Q_n(1,1) = A^n_{11}, \quad Q_n(1,2) = A^n_{12}, \quad Q_n(2,1) = A^n_{21}, \quad Q_n(2,2) = A^n_{22}.$$

Now consider the functions (for $n \in \{1,2\}$) $\overline{Q}_n = [0,1] \times [0,1] \to \mathbb{R}$ defined as follows:

$$\forall n : \overline{Q}_n(x_1, x_2) = A^n_{11} x_1 x_2 + A^n_{12} x_1 (1 - x_2) + A^n_{21} (1 - x_1) x_2 \\ + A^n_{22} (1 - x_1)(1 - x_2). \quad (3)$$

Obviously

$$Q_n(1,1) = \overline{Q}_n(1,1), \quad Q_n(1,2) = \overline{Q}_n(1,0), \\ Q_n(2,1) = \overline{Q}_n(0,1), \quad Q_n(2,2) = \overline{Q}_n(0,0).$$

What we have done is first to map C to 1 (i.e., full cooperation) and D to 0 (i.e., null cooperation) and then to interpolate the "interior" $\overline{Q}_n$ values from the "corner" Q_n values. In this way a bimatrix game has been extended to a "*game on the unit square*"[20] with polynomial payoff functions. We have the following.

Proposition 4 (Raghavan[20]). Every two-person game on the unit square, with the payoffs $\overline{Q}_1(x_1, x_2)$ and $\overline{Q}_2(x_1, x_2)$ being polynomials in x_1 and x_2, has at least one NE.

The previously discussed questions ("dilemmas") regarding the emergence of cooperation in bimatrix social dilemmas also appear in connection to continuous action social dilemmas. In fact, one reason for the introduction of continuous action social dilemmas has been the hope that these questions can be answered satisfactorily in the continuous action context. However, even in continuous social dilemmas, emergence of cooperation is not guaranteed.[9,31]

As a final remark, it is worth noting that the discrete action payoff function (2) and the continuous action payoff function (3) are remarkably similar. To see this, let us rewrite (2) as follows: for $n \in \{1,2\}$, let $p^1_n = p_n$, $p^2_n = 1 - p^1_n = 1 - p_n$. Then (2) becomes

$$\forall n : Q_n(p_1, p_2) = A^n_{11} p_1 p_2 + A^n_{12} p_1 (1 - p_2) + A^n_{21} (1 - p_1) p_2 \\ + A^n_{22} (1 - p_1)(1 - p_2) \quad (4)$$

and the similarity of (3) and (4) becomes obvious from the correspondence $x_1 \leftrightarrow p_1$ and $x_2 \leftrightarrow p_2$. However an important distinction must be made: under the (usual) assumption of *fully observable actions*, in the continuous action game both P_1 and P_2 will know the values of x_1, x_2, while in the finite action game P_1 will not know p_2 and P_2 will not know p_1.

4. The Basic Action Update Model

In this section we present and study a model for the evolution of actions in a two-player game on the unit square. The model is intended to capture the drift towards or away from cooperation in repeated plays of the game. The *update equations* will have the form

$$\frac{dx_1}{dt} = f_1(x_1, x_2), \qquad \frac{dx_2}{dt} = f_2(x_1, x_2), \tag{5}$$

where $x_n(t)$ is P_n's action at time t. The above equations imply that $x_1(t)$, $x_2(t)$ are functions of a continuous time variable t; in other words we assume that the replays of the game occur in *continuous* time. This is a common approach and can be understood as an approximation of replays at *discrete* times $t_i = i \cdot \Delta t$, with $i \in \mathbb{N}_0$ and $\Delta t \to 0$. Furthermore, (5) implies that, at each time t, each player knows not only his own but also the opponent's action (this is the "fully observable actions" hypothesis; in game theoretic terms we assume a game of *perfect information*).

4.1. *The Update Equations*

Our first task is to choose a concrete form for the equations (5). To make our arguments specific, let us initially assume that the game being played is a continuous action Prisoner's Dilemma.[d]

We assume that each player is continuously modifying his cooperation level $x_n(t)$ in response to the other player's cooperation level. We also assume that both players use the same (more accurately: symmetric) update equations. Specifically, we assume that the n-th player's action x_n (for $n \in \{1,2\}$) is updated according to

$$\frac{dx_1}{dt} = (a_{11}x_1x_2 + a_{10}x_1(1-x_2) + a_{01}(1-x_1)x_2 + a_{00}(1-x_1)(1-x_2))\,x_1(1-x_1) \tag{6}$$

$$\frac{dx_2}{dt} = (a_{11}x_1x_2 + a_{10}(1-x_1)x_2 + a_{01}x_1(1-x_2) + a_{00}(1-x_1)(1-x_2))\,x_2(1-x_2) \tag{7}$$

To justify the specific form of the above equations, let us look at (6) and consider separately each term in the right hand of the equation; a similar interpretation can be given for (7). It is easier to understand the following interpretations by first considering the "extremal" cases, i.e., when $(x_1, x_2) \in \{0,1\} \times \{0,1\}$.

(1) $a_{11}x_1x_2$ is P_1's tendency to change x_1 in the face of reciprocated cooperation.
(2) $a_{10}x_1(1-x_2)$ is P_1's tendency to change x_1 in the face of unreciprocated cooperation.
(3) $a_{01}(1-x_1)x_2$ is P_1's tendency to change x_1 in the face of unreciprocated defection.

[d] We will later argue that the exact same equations can be used for updates in a continuous action HD or SH game and in fact for any social dilemma in the unit square.

(4) $a_{00}x_1x_2$ is P_1's tendency to change x_1 in the face of reciprocated defection.

It remains to choose the values of the parameters a_{mn} $(m, n \in \{1, 2\})$. For the sake of simplicity, we assume $|a_{mn}| = 1$ and it remains to specify the *sign* of each a_{mn} coefficient. We make the following choices.

(1) $a_{11} = 1$ because reciprocated cooperation reinforces itself.
(2) $a_{10} = -1$ because unreciprocated cooperation makes one spiteful and uncooperative.
(3) $a_{01} = -1$ because unreciprocated defection is rewarded and hence discourages cooperation.
(4) $a_{00} = 1$ because reciprocated defection makes a player understand that cooperation is better.

Hence equations (6)-(7) reduce to

$$\frac{dx_1}{dt} = \left(x_1x_2 - x_1\left(1-x_2\right) - \left(1-x_1\right)x_2 + \left(1-x_1\right)\left(1-x_2\right)\right)x_1\left(1-x_1\right)$$
$$\frac{dx_2}{dt} = \left(x_1x_2 - x_1\left(1-x_2\right) - \left(1-x_1\right)x_2 + \left(1-x_1\right)\left(1-x_2\right)\right)x_2\left(1-x_2\right).$$

Simplifying the above equations we obtain

$$\frac{dx_1}{dt} = 4\left(\frac{1}{2}-x_1\right)\left(\frac{1}{2}-x_2\right)x_1\left(1-x_1\right) \tag{8}$$
$$\frac{dx_2}{dt} = 4\left(\frac{1}{2}-x_1\right)\left(\frac{1}{2}-x_2\right)x_2\left(1-x_2\right) \tag{9}$$

which will be the form we will use in our subsequent study. Before proceeding any further we note that, from standard existence and uniqueness theorems,[5,29] we have the following.

Proposition 5. For any initial conditions $\mathbf{x}(0) = \mathbf{x}_0 \in [0,1]\times[0,1]$, (8)-(9) have a unique solution $\mathbf{x}(t|\mathbf{x}_0)$.

4.2. *Equilibria*

Let us find and characterize the equilibria of (8)-(9). By definition, these are the solutions of the system of algebraic equations

$$0 = 4\left(\frac{1}{2}-x_1\right)\left(\frac{1}{2}-x_2\right)x_1\left(1-x_1\right) \tag{10}$$
$$0 = 4\left(\frac{1}{2}-x_1\right)\left(\frac{1}{2}-x_2\right)x_2\left(1-x_2\right) \tag{11}$$

We easily find that (10)-(11) has exactly the following solutions in $[0,1]\times[0,1]$.

$$\left(x_1,x_2\right)=\left(0,0\right), \quad \left(x_1,x_2\right)=\left(1,1\right), \quad \left(x_1,x_2\right)=\left(0,1\right), \quad \left(x_1,x_2\right)=\left(1,0\right),$$

$$(x_1,x_2)\in\left\{\left(\frac{1}{2},z\right):z\in[0,1]\right\},\quad (x_1,x_2)\in\left\{\left(z,\frac{1}{2}\right):z\in[0,1]\right\}.$$

Furthermore, after simplification, the Jacobian matrix is

$$J\left(x_1,x_2\right)=\begin{bmatrix}-\left(6x_1^2-6x_1+1\right)\left(2x_2-1\right) & -2\left(2x_1-1\right)x_1\left(x_1-1\right)\\ \\ -2\left(2x_2-1\right)x_2\left(x_2-1\right) & -\left(6x_2^2-6x_2+1\right)\left(2x_1-1\right)\end{bmatrix}$$

and, from Proposition 2, we have the following cases.

(1) $(0,0)$ is unstable because $J\left(0,0\right)=\begin{bmatrix}1\,0\\0\,1\end{bmatrix}$, eigenvalues are $\lambda_1=1$, $\lambda_2=1$.

(2) $(1,1)$ is stable because $J\left(1,1\right)=\begin{bmatrix}-1 & 0\\ 0 & -1\end{bmatrix}$, eigenvalues are $\lambda_1=-1$, $\lambda_2=-1$.

(3) $(1,0)$ is unstable because $J\left(1,0\right)=\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}$, eigenvalues are $\lambda_1=1$, $\lambda_2=-1$.

(4) $(0,1)$ is unstable because $J\left(0,1\right)=\begin{bmatrix}-1 & 0\\ & 1\end{bmatrix}$, eigenvalues are $\lambda_1=-1$, $\lambda_2=1$.

(5) $\left(\frac{1}{2},z\right)$ is unstable because $J\left(\frac{1}{2},z\right)=\begin{bmatrix}z-\frac{1}{2} & 0\\ 0 & 0\end{bmatrix}$, eigenvalues are $\lambda_1=z-\frac{1}{2}$, $\lambda_2=0$.

(6) $\left(z,\frac{1}{2}\right)$ is unstable because $J\left(z,\frac{1}{2}\right)=\begin{bmatrix}0 & 0\\ 0 & z-\frac{1}{2}\end{bmatrix}$, eigenvalues are $\lambda_1=0$, $\lambda_2=z-\frac{1}{2}$.

However the above stability analysis does not tell the full story. We can get a better idea of the behavior of (8)-(9) by studying its *phase plot*, illustrated in Fig. 1.

We see in Fig. 1 that $[0,1]\times[0,1]$ can be partitioned into several regions; furthermore the value of $\lim_{t\rightarrow\infty}\left(x_1\left(t\right),x_2\left(t\right)\right)$ is determined by the region to which the initial condition $\left(x_1\left(0\right),x_2\left(0\right)\right)$ belongs. For example, it appears from Fig. 1 that

$$\forall\left(x_1\left(0\right),x_2\left(0\right)\right)\in\left(\frac{1}{2},1\right)\times\left(\frac{1}{2},1\right):\lim_{t\rightarrow\infty}\left(x_1\left(t\right),x_2\left(t\right)\right)=\left(1,1\right).$$

Similar remarks can be made when $\left(x_1\left(0\right),x_2\left(0\right)\right)$ to some other region. Hence we will next embark upon a study of the *attraction basins* of (8)-(9).

We see in Fig. 1 that $[0,1]\times[0,1]$ can be partitioned into several regions; furthermore the value of $\lim_{t\rightarrow\infty}\left(x_1\left(t\right),x_2\left(t\right)\right)$ is determined by the region to which the initial condition $\left(x_1\left(0\right),x_2\left(0\right)\right)$ belongs. For example, it appears from Fig. 1 that

$$\forall\left(x_1\left(0\right),x_2\left(0\right)\right)\in\left(\frac{1}{2},1\right)\times\left(\frac{1}{2},1\right):\lim_{t\rightarrow\infty}\left(x_1\left(t\right),x_2\left(t\right)\right)=\left(1,1\right).$$

Similar remarks can be made when $\left(x_1\left(0\right),x_2\left(0\right)\right)$ to some other region. Hence we will next embark upon a study of the *attraction basins* of (8)-(9).

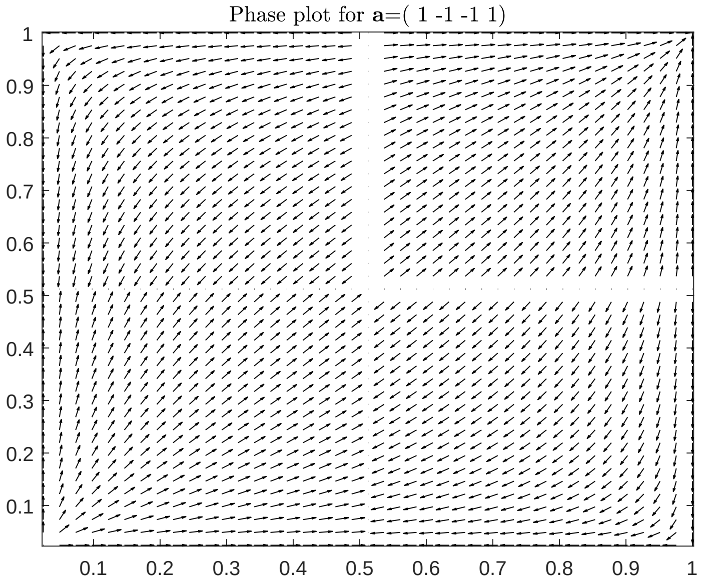


Fig. 1.: Phase plot of the system (8)-(9).

4.3. *Attraction Basins*

Looking at (8)-(9) we see that, for all $m \in \{1,2\}$, $x_m(1-x_m) > 0$. Hence the sign of $\frac{dx_m}{dt}$ is determined by $4\left(\frac{1}{2}-x_1\right)\left(\frac{1}{2}-x_2\right)$, which is positive on $\left(0,\frac{1}{2}\right)\times\left(0,\frac{1}{2}\right)$ and $\left(\frac{1}{2},1\right)\times\left(\frac{1}{2},1\right)$, and negative on $\left(0,\frac{1}{2}\right)\times\left(\frac{1}{2},1\right)$ and $\left(\frac{1}{2},1\right)\times\left(0,\frac{1}{2}\right)$. Accordingly, in our study of the basins of attraction we start by considering these four "subsquares" of $[0,1]\times[0,1]$.

Before proceeding with a detailed analysis of each subsquare (as well as of their boundaries), we remind the reader that, by virtue of Proposition 5, the system (8)-(9) has a unique solution for any initial condition $(x_1(0), x_2(0)) \in [0,1]\times[0,1]$.

4.3.1. *Initial Conditions* $(x_1(0), x_2(0)) \in \left(\frac{1}{2},1\right)\times\left(\frac{1}{2},1\right)$

Proposition 6. The system

$$\frac{dx_1}{dt} = 4\left(\frac{1}{2}-x_1\right)\left(\frac{1}{2}-x_2\right)x_1(1-x_1), \qquad x_1(0) = c_1 \in \left(\frac{1}{2},1\right)$$
$$\frac{dx_2}{dt} = 4\left(\frac{1}{2}-x_1\right)\left(\frac{1}{2}-x_2\right)x_2(1-x_2), \qquad x_2(0) = c_2 \in \left(\frac{1}{2},1\right)$$

has a unique solution $(x_1(t), x_2(t))$ which satisfies $\lim_{t\to\infty} x_1(t) = \lim_{t\to\infty} x_2(t) = 1$.

Proof. Existence and uniqueness follow from Proposition 5. For any $k \geq 3$ consider

$$\frac{dx_1}{dt} = 4\left(x_2 - \frac{1}{2}\right) x_1 \left(x_1 - \frac{1}{2}\right)(1 - x_1), \qquad x_1(0) = c_1 \in \left(\frac{1}{2} + \frac{1}{k}, 1\right) \tag{12}$$

$$\frac{dx_2}{dt} = 4\left(x_2 - \frac{1}{2}\right) x_2 \left(x_1 - \frac{1}{2}\right)(1 - x_2), \qquad x_2(0) = c_2 \in \left(\frac{1}{2} + \frac{1}{k}, 1\right) \tag{13}$$

The solution is $(x_1(t), x_2(t))$ where both $x_1(t)$ and $x_2(t)$ are strictly increasing functions. Hence

$$\forall m \in \{1, 2\}, \forall t : x_m(t) \in \left(\frac{1}{2} + \frac{1}{k}, 1\right] \tag{14}$$

Consequently we have

$$4\left(x_2 - \frac{1}{2}\right) x_1 \geq 4 \cdot \frac{1}{k} \cdot \frac{1}{2} = \frac{2}{k}$$

and

$$\frac{dx_1}{dt} \geq \frac{2}{k}\left(x_1 - \frac{1}{2}\right)(1 - x_1). \tag{15}$$

Now consider the DE

$$\frac{dy_k}{dt} = \frac{2}{k}\left(y_k - \frac{1}{2}\right)(1 - y_k), \qquad y_k(0) = c_1 \in \left(\frac{1}{2} + \frac{1}{k}, 1\right) \tag{16}$$

This has the solution

$$y_k(t) = \frac{\frac{2c_1 - 1}{c_1 - 1} e^{t/k} - 1}{\frac{2c_1 - 1}{c_1 - 1} e^{t/k} - 2} = 1 + \frac{1}{\frac{2c_1 - 1}{c_1 - 1} e^{t/k} - 2}$$

We see that $y_k(t)$ is increasing and $\lim_{t\to\infty} y_k(t) = 1$. On the other hand, from (12)-(16) we have

$$\forall t : \frac{dx_1}{dt} \geq \frac{dy_k}{dt}.$$

Hence, from Petrovich's theorem we have: $\forall t : x_1(t) \geq y_k(t)$. Hence

$$1 = \lim_{t\to\infty} y_k(t) \leq \lim_{t\to\infty} x_1(t) \leq 1 \Rightarrow \lim_{t\to\infty} x_1(t) = 1.$$

In the same manner we prove $\lim_{t\to\infty} x_2(t) = 1$. Since these hold when $(x_1(0), x_2(0)) \in \left(\frac{1}{2} + \frac{1}{k}, 1\right) \times \left(\frac{1}{2} + \frac{1}{k}, 1\right)$ and for any $k \geq 3$, we conclude that

$$\forall (x_1(0), x_2(0)) \in \left(\frac{1}{2}, 1\right) \times \left(\frac{1}{2}, 1\right) : \lim_{t\to\infty} x_1(t) = \lim_{t\to\infty} x_2(t) = 1$$

and we have proved the proposition. □

4.3.2. *Initial Conditions* $(x_1(0), x_2(0)) \in \left(\frac{1}{2}, 1\right) \times \left(0, \frac{1}{2}\right)$

Proposition 7. The system

$$\frac{dx_1}{dt} = 4\left(\frac{1}{2} - x_1\right)\left(\frac{1}{2} - x_2\right)x_1(1 - x_1), \qquad x_1(0) = c_1 \in \left(\frac{1}{2}, 1\right)$$

$$\frac{dx_2}{dt} = 4\left(\frac{1}{2} - x_1\right)\left(\frac{1}{2} - x_2\right)x_2(1 - x_2), \qquad x_2(0) = c_2 \in \left(0, \frac{1}{2}\right)$$

has a unique solution $(x_1(t), x_2(t))$ which satisfies $\lim_{t\to\infty} x_1(t) = \frac{1}{2}$ and $\lim_{t\to\infty} x_2(t) \in \left[0, \frac{1}{2}\right)$.

Proof. Existence and uniqueness follow from Proposition 5. Now fix some $k \geq 3$ and consider the system

$$\frac{dx_1}{dt} = -4\left(\frac{1}{2} - x_2\right)x_1\left(x_1 - \frac{1}{2}\right)(1 - x_1), \qquad x_1(0) = c_1 \in \left(\frac{1}{2}, 1 - \frac{1}{k}\right)$$

$$\frac{dx_2}{dt} = -4\left(\frac{1}{2} - x_2\right)\left(x_1 - \frac{1}{2}\right)x_2(1 - x_2), \qquad x_2(0) = c_2 \in \left(0, \frac{1}{2} - \frac{1}{k}\right)$$

(notice that we have rewriten the DE's (8)-(9) in slightly different but equivalent form). The problem has as unique solution $(x_1(t), x_2(t))$ where both $x_1(t)$ and $x_2(t)$ are strictly decreasing functions, with $x_1(t) \in \left[\frac{1}{2}, 1 - \frac{1}{k}\right)$ and $x_2(t) \in \left[0, \frac{1}{2} - \frac{1}{k}\right)$. The following limits exist:

$$\lim_{t\to\infty} x_1(t) = \bar{c}_1 \in \left[\frac{1}{2}, 1\right), \qquad \lim_{t\to\infty} x_2(t) = \bar{c}_2 \in \left[0, \frac{1}{2}\right).$$

Furthermore, we have

$$4\left(\frac{1}{2} - x_2\right)x_1 \geq \frac{2}{k}.$$

Hence

$$-4\left(\frac{1}{2} - x_2\right)x_1\left(x_1 - \frac{1}{2}\right)(1 - x_1) \leq -\frac{2}{k}\left(x_1 - \frac{1}{2}\right)(1 - x_1)$$

and

$$\frac{dx_1}{dt} \leq -\frac{2}{k}\left(x_1 - \frac{1}{2}\right)(1 - x_1), \qquad x_1(0) = c_1 \in \left(\frac{1}{2}, 1\right).$$

Now consider the DE

$$\frac{dy_k}{dt} = -\frac{2}{k}\left(y_k - \frac{1}{2}\right)(1 - y_k), \qquad y(0) = c_1 \in \left(\frac{1}{2}, 1 - \frac{1}{k}\right).$$

This has a unique solution $y_k(t)$ which is strictly decreasing and satisfies $\lim_{t\to\infty} y_k(t) = \frac{1}{2}$. Hence $\forall t : y_k(t) \geq x_1(t) \geq \frac{1}{2}$ and so we have

$$\frac{1}{2} = \lim_{t\to\infty} y_k(t) \geq \lim_{t\to\infty} x_1(t) \geq \frac{1}{2} \Rightarrow \lim_{t\to\infty} x_1(t) = \frac{1}{2}.$$

In short $\lim_{t\to\infty} x_1(t) = \frac{1}{2}$ and $\lim_{t\to\infty} x_2(t) \in \left[0, \frac{1}{2}\right)$. Since these hold when $(x_1(0), x_2(0)) \in \left(\frac{1}{2} + \frac{1}{k}, 1\right) \times \left(0, \frac{1}{2} - \frac{1}{k}\right)$ and for any $k \geq 3$, we conclude that

$$\forall (x_1(0), x_2(0)) \in \left(\frac{1}{2}, 1\right) \times \left(0, \frac{1}{2}\right) : \lim_{t\to\infty} x_1(t) = \frac{1}{2} \text{ and } \lim_{t\to\infty} x_2(t) \in \left[0, \frac{1}{2}\right)$$

and we have proved the proposition. □

4.3.3. *Initial Conditions* $(x_1(0), x_2(0)) \in \left(0, \frac{1}{2}\right) \times \left(\frac{1}{2}, 1\right)$

In the same manner as in the previous case, but with the roles of $x_1(t)$ and $x_2(t)$ interchanged, we prove the following.

Proposition 8. The system

$$\begin{aligned} \frac{dx_1}{dt} &= 4\left(\frac{1}{2} - x_1\right)\left(\frac{1}{2} - x_2\right) x_1(1 - x_1), \qquad x_1(0) = c_1 \in \left(0, \frac{1}{2}\right) \\ \frac{dx_2}{dt} &= 4\left(\frac{1}{2} - x_1\right)\left(\frac{1}{2} - x_2\right) x_2(1 - x_2), \qquad x_2(0) = c_2 \in \left(\frac{1}{2}, 1\right) \end{aligned}$$

has a unique solution $(x_1(t), x_2(t))$ which satisfies $\lim_{t\to\infty} x_1(t) \in \left[0, \frac{1}{2}\right)$ and $\lim_{t\to\infty} x_2(t) = \frac{1}{2}$.

4.3.4. *Initial Conditions* $(x_1(0), x_2(0)) \in \left(0, \frac{1}{2}\right) \times \left(0, \frac{1}{2}\right)$

The proof of the following proposition is a little more involved than the previous ones.

Proposition 9. The system

$$\begin{aligned} \frac{dx_1}{dt} &= 4\left(\frac{1}{2} - x_1\right)\left(\frac{1}{2} - x_2\right) x_1(1 - x_1), \qquad x_1(0) = c_1 \in \left(0, \frac{1}{2}\right) \\ \frac{dx_2}{dt} &= 4\left(\frac{1}{2} - x_1\right)\left(\frac{1}{2} - x_2\right) x_2(1 - x_2), \qquad x_2(0) = c_2 \in \left(0, \frac{1}{2}\right) \end{aligned}$$

has a unique solution $(x_1(t), x_2(t))$ which satisfies $\lim_{t\to\infty} x_1(t) = \overline{c}_1$ and $\lim_{t\to\infty} x_2(t) = \overline{c}_2$ and either $\overline{c}_1 = \frac{1}{2}$ or $\overline{c}_2 = \frac{1}{2}$ or both.

Proof. Fix some $k_0 \geq 3$ and consider the system

$$\begin{aligned} \frac{dx_1}{dt} &= 4\left(\frac{1}{2} - x_1\right)\left(\frac{1}{2} - x_2\right) x_1(1 - x_1), \qquad x_1(0) = c_1 \in \left(0, \frac{1}{2} - \frac{1}{k_0}\right) \\ \frac{dx_2}{dt} &= 4\left(\frac{1}{2} - x_1\right)\left(\frac{1}{2} - x_2\right) x_2(1 - x_2), \qquad x_2(0) = c_2 \in \left(0, \frac{1}{2} - \frac{1}{k_0}\right) \end{aligned}$$

This system has a unique solution $(x_1(t), x_2(t))$ where both $x_1(t)$ and $x_2(t)$ are strictly increasing functions which take values inside $(0, \frac{1}{2}]$. Since $x_1(t)$, $x_2(t)$ are bounded and increasing, the following limits exist: $\lim_{t\to\infty} x_1(t) = \overline{c}_1 \in (0, \frac{1}{2}]$ and $\lim_{t\to\infty} x_2(t) = \overline{c}_2 \in (0, \frac{1}{2}]$. Furthermore we have

$$\forall k \geq 3 : \forall (x_1(t), x_2(t)) \in \left(0, \frac{1}{2} - \frac{1}{k}\right) \times \left(0, \frac{1}{2} - \frac{1}{k}\right) : 4\left(\frac{1}{2} - x_2(t)\right)(1 - x_1(t)) \geq \frac{2}{k} \tag{17}$$

and

$$\frac{dx_1}{dt} \geq \frac{2}{k}\left(\frac{1}{2} - x_1\right) x_1, \qquad x_1(0) = c_1 \in \left(0, \frac{1}{2} - \frac{1}{k_0}\right)$$

Now take any $k \geq k_0$ and consider the DE

$$\frac{dy_k}{dt} = \frac{2}{k}\left(\frac{1}{2} - y_k\right) y_k, \qquad y_k\left(0\right) = c_1 \in \left(0, \frac{1}{2} - \frac{1}{k_0}\right).$$

This has the unique solution:

$$y_k\left(t\right) = \frac{1}{2 + \frac{1-2c_1}{c_1} e^{-\frac{1}{k}t}}$$

which is strictly increasing. From $y_k\left(0\right) = c_1 \in \left(0, \frac{1}{2} - \frac{1}{k_0}\right)$, $k_0 \leq k$ and $\lim_{t\to\infty} y_k\left(t\right) = \frac{1}{2}$, we conclude that there exists a t_k at which we will have $y_k\left(t_k\right) = \frac{1}{2} - \frac{1}{k}$. In fact, solving

$$\frac{1}{2 + \frac{1-2c_1}{c_1} e^{-\frac{1}{k}t}} = \frac{1}{2} - \frac{1}{k}$$

we get

$$t_k = k \ln \frac{\left(1 - 2c_1\right) k - 2 + 4c_1}{4c_1}$$

Hence the sequence t_k, t_{k+1}, ... is strictly increasing. Now we have two possibilities.

(1) The inequality (17) holds for every $t \in [0, t_k]$. Then, by Petrovitsch's theorem, at every t_k we have $x_1\left(t_k\right) \geq y_k\left(t_k\right) = \frac{1}{2} - \frac{1}{k} > \frac{1}{2} - \sqrt{\frac{2}{k}}$.

(2) The inequality (17) does not hold for some $\widetilde{t_k} \leq t_k$, i.e., $\left(\frac{1}{2} - x_2\left(\widetilde{t_k}\right)\right)\left(1 - x_1\left(\widetilde{t_k}\right)\right) < \frac{2}{k}$. In this case we must have:

   (a) either $1 - x_1\left(\widetilde{t_k}\right) < \sqrt{\frac{2}{k}}$ which implies $x_1\left(t_k\right) \geq x_1\left(\widetilde{t_k}\right) > 1 - \sqrt{\frac{2}{k}} > \frac{1}{2} - \sqrt{\frac{2}{k}}$;

   (b) or $\frac{1}{2} - x_2\left(\widetilde{t_k}\right) < \sqrt{\frac{2}{k}}$ which implies $x_2\left(t_k\right) \geq x_2\left(\widetilde{t_k}\right) \geq \frac{1}{2} - \sqrt{\frac{2}{k}}$.

In any case, for every large enough k, we have that either $x_1\left(t_k\right) > \frac{1}{2} - \sqrt{\frac{2}{k}}$ or $x_2\left(t_k\right) > \frac{1}{2} - \sqrt{\frac{2}{k}}$; hence one of the two inequalities must hold infinitely often. Consequently, there exist strictly increasing sequences k_1, k_2, ... and t_{k_1}, t_{k_2}, ... such that at least one of the following holds for all n and all $t \geq t_{k_n}$.

(1) Either $x_1\left(t\right) \geq x_1\left(t_{k_n}\right) > \frac{1}{2} - \sqrt{\frac{2}{k_n}}$; then, since also $\frac{1}{2} \geq x_1\left(t\right)$, we have $\lim_{t\to\infty} x_1\left(t\right) = \frac{1}{2}$.

(2) Or $x_2\left(t\right) \geq x_2\left(t_{k_n}\right) \geq \frac{1}{2} - \sqrt{\frac{2}{k_n}}$; then, since also $\frac{1}{2} \geq x_2\left(t\right)$, we have that $\lim_{t\to\infty} x_2\left(t\right) = \frac{1}{2}$.

Finally, the above holds for any $\left(x_1\left(0\right), x_2\left(0\right)\right) \in \left(0, \frac{1}{2} - \frac{1}{k_0}\right) \times \left(0, \frac{1}{2} - \frac{1}{k_0}\right)$ and for any $k_0 \geq 3$; hence it also holds for any $\left(x_1\left(0\right), x_2\left(0\right)\right) \in \left(0, \frac{1}{2}\right) \times \left(0, \frac{1}{2}\right)$ and the proof is complete. □

Remark 2. With a little more effort we can prove the following stronger result.

Proposition 10. The system

$$\begin{aligned}\frac{dx_1}{dt} &= 4\left(\frac{1}{2}-x_1\right)\left(\frac{1}{2}-x_2\right)x_1\left(1-x_1\right), \qquad x_1\left(0\right)=c_1\in\left(0,\frac{1}{2}\right)\\ \frac{dx_2}{dt} &= 4\left(\frac{1}{2}-x_1\right)\left(\frac{1}{2}-x_2\right)x_2\left(1-x_2\right), \qquad x_2\left(0\right)=c_2\in\left(0,\frac{1}{2}\right)\end{aligned}$$

has a unique solution $(x_1(t), x_2(t))$ which satisfies $\lim_{t\to\infty} x_1(t) = \overline{c}_1$ and $\lim_{t\to\infty} x_2(t) = \overline{c}_2$. Now define the sets

$$\begin{aligned}A &= \left\{(z_1,z_2)\in\left(0,\frac{1}{2}\right)\times\left(0,\frac{1}{2}\right) \text{ and } z_1<z_2\right\},\\ B &= \left\{(z_1,z_2)\in\left(0,\frac{1}{2}\right)\times\left(0,\frac{1}{2}\right) \text{ and } z_1>z_2\right\},\\ C &= \left\{(z_1,z_2)\in\left(0,\frac{1}{2}\right)\times\left(0,\frac{1}{2}\right) \text{ and } z_1=z_2\right\}.\end{aligned}$$

Then we have the following

$$\begin{aligned}(x_1(0),x_2(0))\in A &\Rightarrow \lim_{t\to\infty}x_1(t)=\frac{1}{2} \text{ and } \lim_{t\to\infty}x_2(t)<\frac{1}{2},\\ (x_1(0),x_2(0))\in B &\Rightarrow \lim_{t\to\infty}x_1(t)<\frac{1}{2} \text{ and } \lim_{t\to\infty}x_2(t)=\frac{1}{2},\\ (x_1(0),x_2(0))\in C &\Rightarrow \lim_{t\to\infty}x_1(t)=\frac{1}{2} \text{ and } \lim_{t\to\infty}x_2(t)=\frac{1}{2}.\end{aligned}$$

4.3.5. *Some Initial Conditions belonging to the set* $\left\{0,\frac{1}{2},1\right\}$

Now we present a proposition which applies to cases in which $x_1(0)\in\left\{0,\frac{1}{2},1\right\}$.

Proposition 11. For the problem

$$\begin{aligned}\frac{dx_1}{dt} &= 4\left(\frac{1}{2}-x_1\right)\left(\frac{1}{2}-x_2\right)x_1\left(1-x_1\right), \qquad x_1\left(0\right)=c_1\\ \frac{dx_2}{dt} &= 4\left(\frac{1}{2}-x_1\right)\left(\frac{1}{2}-x_2\right)x_2\left(1-x_2\right), \qquad x_2\left(0\right)=c_2\end{aligned}$$

the following hold (the indicated limits always exist).

(1) When $c_1=0$ and $c_2\in\left(0,\frac{1}{2}\right)$ we have $\lim_{t\to\infty}x_1(t)=0$ and $\lim_{t\to\infty}x_2(t)=\frac{1}{2}$.
(2) When $c_1=\frac{1}{2}$ and $c_2\in\left(0,\frac{1}{2}\right)$ we have $\lim_{t\to\infty}x_1(t)=\frac{1}{2}$ and $\lim_{t\to\infty}x_2(t)=c_2$.
(3) When $c_1=1$ and $c_2\in\left(0,\frac{1}{2}\right)$ we have $\lim_{t\to\infty}x_1(t)=1$ and $\lim_{t\to\infty}x_2(t)=0$.
(4) When $c_1=0$ and $c_2\in\left(\frac{1}{2},1\right)$ we have $\lim_{t\to\infty}x_1(t)=0$ and $\lim_{t\to\infty}x_2(t)=\frac{1}{2}$.
(5) When $c_1=\frac{1}{2}$ and $c_2\in\left(\frac{1}{2},1\right)$ we have $\lim_{t\to\infty}x_1(t)=\frac{1}{2}$ and $\lim_{t\to\infty}x_2(t)=c_2$.
(6) When $c_1=1$ and $c_2\in\left(\frac{1}{2},1\right)$ we have $\lim_{t\to\infty}x_1(t)=1$ and $\lim_{t\to\infty}x_2(t)=1$.

Proof. The proofs of all cases are straightforward and so only an outline is given. For example, when $c_1=0$ and $c_2\in\left(0,\frac{1}{2}\right)$, the system reduces to

$$\frac{dx_1}{dt}=0, \quad \frac{dx_2}{dt}=2\left(\frac{1}{2}-x_2\right)x_2\left(1-x_2\right). \tag{18}$$

From the first DE of (18) we see that $x_1(t) = 0$ for all t. The second DE of (18) involves only x_2 and can be easily solved; inspection of the solution readily shows that $\lim_{t\to\infty} x_2(t) = \frac{1}{2}$.

The remaining cases are similar; we only give the reduced form of the DE's and leave verification of the claimed results to the reader.

(1) When $c_1 = \frac{1}{2}$ and $c_2 \in \left(0, \frac{1}{2}\right)$ the DE's reduce to
$$\frac{dx_1}{dt} = 0, \quad \frac{dx_2}{dt} = 0.$$
(2) When $c_1 = 1$ and $c_2 \in \left(0, \frac{1}{2}\right)$ the DE's reduce to
$$\frac{dx_1}{dt} = 0, \quad \frac{dx_2}{dt} = -2\left(\frac{1}{2} - x_2\right) x_2 (1 - x_2)$$
(3) When $c_1 = 0$ and $c_2 \in \left(\frac{1}{2}, 1\right)$ the DE's reduce to
$$\frac{dx_1}{dt} = 0, \quad \frac{dx_2}{dt} = 2\left(\frac{1}{2} - x_2\right) x_2 (1 - x_2)$$
(4) When $c_1 = \frac{1}{2}$ and $c_2 \in \left(\frac{1}{2}, 1\right)$ the DE's reduce to
$$\frac{dx_1}{dt} = 0, \quad \frac{dx_2}{dt} = 0$$
(5) When $c_1 = 1$ and $c_2 \in \left(\frac{1}{2}, 1\right)$ the DE's reduce to
$$\frac{dx_1}{dt} = 0, \quad \frac{dx_2}{dt} = -2\left(\frac{1}{2} - x_2\right) x_2 (1 - x_2)$$

This completes the proof. □

We also have a complementary proposition for the cases in which $x_2(0) \in \left\{0, \frac{1}{2}, 1\right\}$. We present it without proof, since it is identical to that of Proposition 11.

Proposition 12. For the problem
$$\frac{dx_1}{dt} = 4\left(\frac{1}{2} - x_1\right)\left(\frac{1}{2} - x_2\right) x_1 (1 - x_1), \qquad x_1(0) = c_1 \tag{19}$$
$$\frac{dx_2}{dt} = 4\left(\frac{1}{2} - x_1\right)\left(\frac{1}{2} - x_2\right) x_2 (1 - x_2), \qquad x_2(0) = c_2 \tag{20}$$
the following hold (the indicated limits always exist).

(1) When $c_1 \in \left(0, \frac{1}{2}\right)$ and $c_2 = 0$ we have $\lim_{t\to\infty} x_1(t) = \frac{1}{2}$ and $\lim_{t\to\infty} x_2(t) = 0$.
(2) When $c_1 \in \left(0, \frac{1}{2}\right)$ and $c_2 = \frac{1}{2}$ we have $\lim_{t\to\infty} x_1(t) = c_1$ and $\lim_{t\to\infty} x_2(t) = \frac{1}{2}$.
(3) When $c_1 \in \left(0, \frac{1}{2}\right)$ and $c_2 = 1$ we have $\lim_{t\to\infty} x_1(t) = 0$ and $\lim_{t\to\infty} x_2(t) = 1$.
(4) When $c_1 \in \left(\frac{1}{2}, 1\right)$ and $c_2 = 0$ we have $\lim_{t\to\infty} x_1(t) = \frac{1}{2}$ and $\lim_{t\to\infty} x_2(t) = 0$.
(5) When $c_1 \in \left(\frac{1}{2}, 1\right)$ and $c_2 = \frac{1}{2}$ we have $\lim_{t\to\infty} x_1(t) = c_1$ and $\lim_{t\to\infty} x_2(t) = \frac{1}{2}$.
(6) When $c_1 \in \left(\frac{1}{2}, 1\right)$ and $c_2 = 1$ we have $\lim_{t\to\infty} x_1(t) = 1$ and $\lim_{t\to\infty} x_2(t) = 1$.

4.3.6. *Summary and Discussion*

We collect our results in the following table. In each row we indicate: the sets which contain $\mathbf{x}(0) = (x_1(0), x_2(0))$ and $\lim_{t\to\infty} \mathbf{x}(t) = (\lim_{t\to\infty} x_1(t), \lim_{t\to\infty} x_2(t))$ (the indicated limits always exist).

$\mathbf{x}(0)$ belongs to	$\lim_{t\to\infty} \mathbf{x}(t)$ belongs to
$\left(\frac{1}{2},1\right) \times \left(\frac{1}{2},1\right)$	$\{(1,1)\}$
$\left(\frac{1}{2},1\right) \times \left(0,\frac{1}{2}\right)$	$\left\{\frac{1}{2}\right\} \times \left[0,\frac{1}{2}\right)$
$\left(0,\frac{1}{2}\right) \times \left(\frac{1}{2},1\right)$	$\left[0,\frac{1}{2}\right) \times \left\{\frac{1}{2}\right\}$
$\left(0,\frac{1}{2}\right) \times \left(0,\frac{1}{2}\right)$	$\left[\left(0,\frac{1}{2}\right) \times \left\{\frac{1}{2}\right\}\right] \cup \left[\left\{\frac{1}{2}\right\} \times \left(0,\frac{1}{2}\right)\right]$
$\{0\} \times \left(0,\frac{1}{2}\right)$	$\left\{\left(0,\frac{1}{2}\right)\right\}$
$\left\{\frac{1}{2}\right\} \times \left(0,\frac{1}{2}\right)$	$\left\{\left(\frac{1}{2},x_2(0)\right)\right\}$
$\{1\} \times \left(0,\frac{1}{2}\right)$	$\{(1,0)\}$
$\{0\} \times \left(\frac{1}{2},1\right)$	$\left\{\left(0,\frac{1}{2}\right)\right\}$
$\left\{\frac{1}{2}\right\} \times \left(\frac{1}{2},1\right)$	$\left\{\left(\frac{1}{2},x_2(0)\right)\right\}$
$\{1\} \times \left(\frac{1}{2},1\right)$	$\{(1,1)\}$
$\left(0,\frac{1}{2}\right) \times \{0\}$	$\left\{\left(\frac{1}{2},0\right)\right\}$
$\left(0,\frac{1}{2}\right) \times \left\{\frac{1}{2}\right\}$	$\left\{\left(x_1(0),\frac{1}{2}\right)\right\}$
$\left(0,\frac{1}{2}\right) \times \{1\}$	$\{(0,1)\}$
$\left(\frac{1}{2},1\right) \times \{0\}$	$\left\{\left(\frac{1}{2},0\right)\right\}$
$\left(\frac{1}{2},1\right) \times \left\{\frac{1}{2}\right\}$	$\left\{\left(x_1(0),\frac{1}{2}\right)\right\}$
$\left(\frac{1}{2},1\right) \times \{1\}$	$\{(1,1)\}$

Table 1.: Attraction basins of the system (8)-(9).

Several conclusions can be reached from the above table and they can be arranged in analogy to the shape of the attraction basins.

(1) Suppose $(x_1(0), x_2(0)) \in \left(\frac{1}{2},1\right) \times \left(\frac{1}{2},1\right)$. Then, as seen in Table 1 and also by Proposition 6, we have $\lim_{t\to\infty}(x_1(t), x_2(t)) = (1,1)$. This can be quite reasonably interpreted as follows: if both players start with a "better-than-average" tendency to cooperate then they will end up with full cooperation.

(2) The case $(x_1(0), x_2(0)) \in \left(0,\frac{1}{2}\right) \times \left(0,\frac{1}{2}\right)$ is more interesting. As seen in Table 1 and also by Proposition 9, we will now have $\lim_{t\to\infty}(x_1(t), x_2(t)) = (\overline{c}_1, \overline{c}_2)$ where at least one of $\overline{c}_1, \overline{c}_2$ will equal $\frac{1}{2}$. In fact, by looking at either Fig. 1

or at the proof of Proposition 9, we can say more: both $x_1(t)$ and $x_2(t)$ are increasing functions of time (both players' levels of cooperation increase). This can be interpreted as follows: both players realize that defection does not payoff and increase their cooperation levels until at least one of them reaches "average" coperation (i.e. at least one of $\lim_{t\to\infty} x_1(t)$ $\lim_{t\to\infty} x_2(t)$ equals $\frac{1}{2}$). At this point the players are trapped into a situation of "average" cooperation, namely one of the equilibrium sets $\left\{\frac{1}{2}\right\} \times \left(0, \frac{1}{2}\right)$, $\left(0, \frac{1}{2}\right) \times \left\{\frac{1}{2}\right\}$.

(3) The players will also end up into one of the above invariant sets if they start in either $\left(0, \frac{1}{2}\right) \times \left(\frac{1}{2}, 1\right)$ or $\left(\frac{1}{2}, 1\right) \times \left(0, \frac{1}{2}\right)$. In both of these cases cooperation is decreasing over time; a "psychological" interpretation could be the following.

   (a) The initially more cooperative player decreases his cooperation level because he observes it is not reciprocated.
   (b) The initially less cooperative player increases his cooperation level because he observes that his defections have a negative outcome.

   Eventually the two levels of cooperation equilibrate at one of the above equilibrium sets.

Overall, the results are rather encouraging, in the following sense: if two continuous PD players actually update their strategies according to the update equations (8)-(9), then they will indeed reach some level of nonzero (and possibly full) cooperation. This is certainly more hopeful than the mutual defection predicted by the NE analysis of the discrete action PD.

5. Extensions

In this section we introduce several possible extensions of the basic model presented in Section 4. The discussion is brief and further study of the proposed extensions is relegated to future publications.

5.1. *Applicability to General Continuous Action Social Dilemmas*

An important part of the basic model is the determination of the signs of the a_{mn} coefficients appearing in the action update equations (6)-(7). These signs were determined by some arguments (for example reinforcement of mutual cooperation) regarding the effect of combinations of pure cooperation / defection levels in $\{0, 1\}$. While we initially presented these arguments in the context of PD, they also hold (for the same reasons) for the other social dilemmas, i.e. HD and SH. Hence the update equations (8)-(9) can be applied to any continuous action social dilemma extrapolated from the discrete actions PD, HD or SH game; in fact, they same arguments apply even to continuous action social dilemmas obtained independently of the discrete actions ones.

5.2. *Symmetric Update Equations with Unitary Coefficients*

We will now argue that *any* combination of unitary action update coefficients, i.e., any

$$\mathbf{a} = (a_{11}, a_{10}, a_{01}, a_{00}) \in \{-1, 1\}^4$$

can be used in the action update equations (6)-(7). Our argument is the following.

(1) We can justify $a_{11} = 1$ because mutual cooperation reinforces itself; we can justify $a_{11} = -1$ because the opponent's cooperation increases temptation to take advantage by defection.
(2) We can justify $a_{10} = -1$ because unreciprocated cooperation makes one spiteful and uncooperative; we can justify $a_{10} = 1$ because when the opponent defects, one may attempt to lure him back to cooperation by showing "good faith" (this argument holds when the outcome of defection is worse than that of cooperation, e.g., in SH).
(3) We can justify $a_{01} = -1$ because unreciprocated defection is rewarded and hence obstructs cooperation; we can justify $a_{01} = 1$ because when a player defects he may become afraid of causing the opponent to also defect and hence incur the high cost of mutual defection.
(4) We can justify $a_{00} = -1$ because reciprocated defection makes one spiteful and unwilling to cooperate; we can justify $a_{00} = 1$ because reciprocated defection makes one understand that cooperation is better.

Under the assumption of *symmetric* update equations, since there exist $2^4 = 16$ possible coefficient vectors $\mathbf{a}$, there also exist 16 possible update equation systems of the form (6)-(7). The system corresponding to $\mathbf{a} = (1, -1, -1, 1)$ is the one we have studied in Section 4. An equally detailed analysis of all 16 systems requires more space than is available in the current paper; hence we simply present the phase plots of these sixteen systems, in the following Figures 2 and 3. The reader is invited to interpret the observed behaviors for each system, with special attention to the conditions under which either full defection or full cooperation is obtained asymptotically.

5.3. *Asymmetric Update Equations with Unitary Coefficients*

If we relax the symmetry condition we get update equations of the form

$$\frac{dx_1}{dt} = (a_{11}x_1x_2 + a_{10}x_1(1-x_2) + a_{01}(1-x_1)x_2 + a_{00}(1-x_1)(1-x_2))\, x_1(1-x_1) \tag{21}$$

$$\frac{dx_2}{dt} = (b_{11}x_1x_2 + b_{10}(1-x_1)x_2 + b_{01}x_1(1-x_2) + b_{00}(1-x_1)(1-x_2))\, x_2(1-x_2) \tag{22}$$

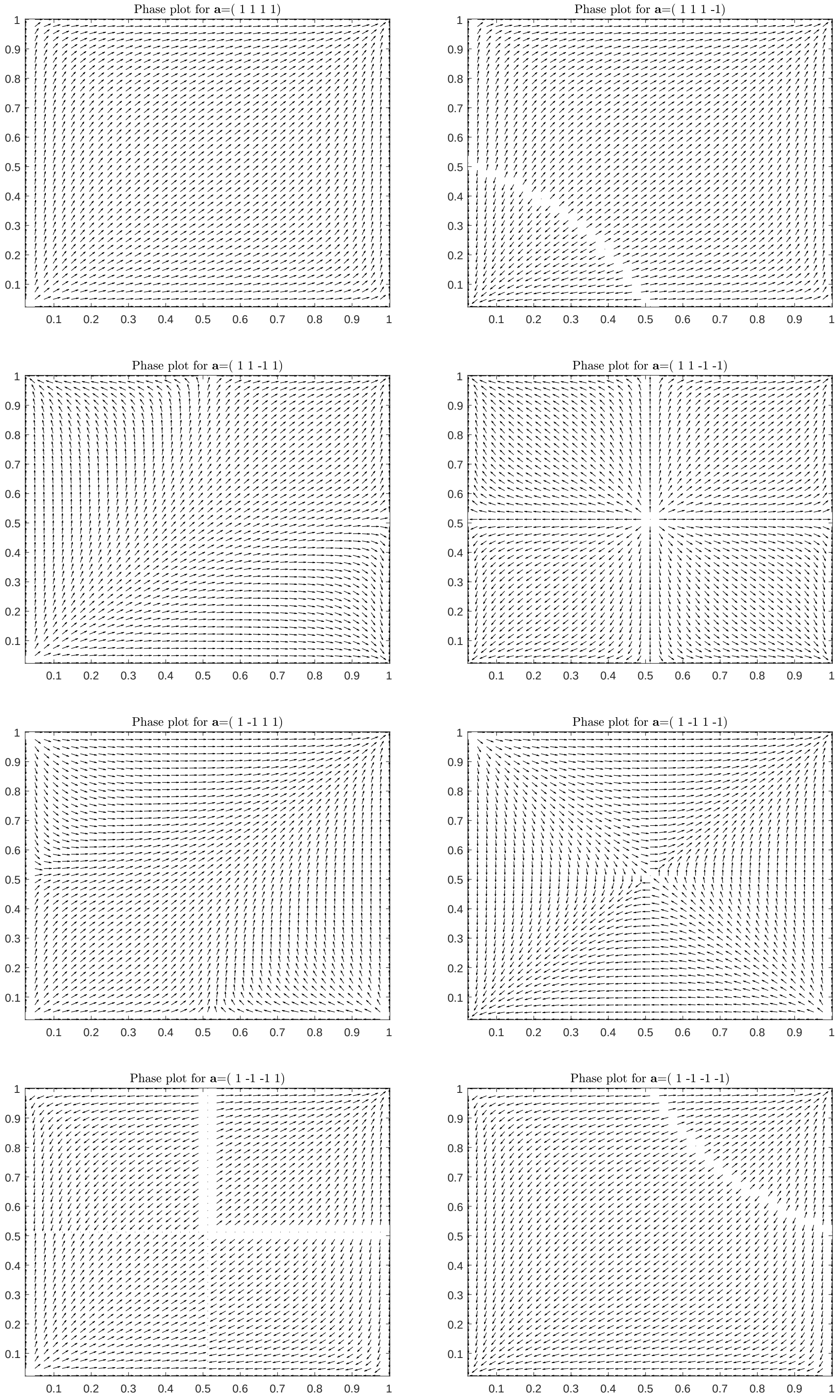
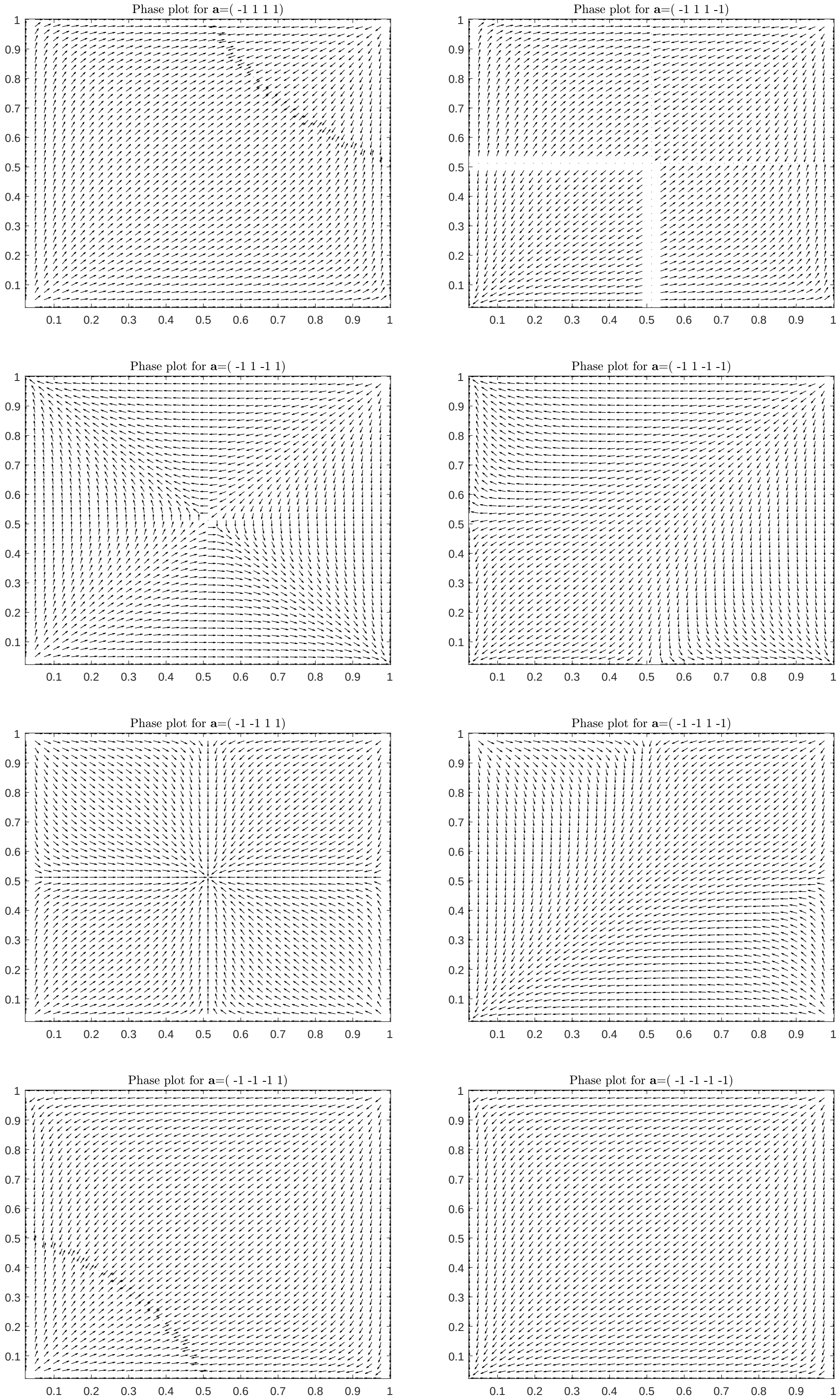


Fig. 2.: Phase plots for systems of the form (6)-(7) and various values of the **a** coefficients.



Phase plot for **a**=(-1 1 1 1)

Phase plot for **a**=(-1 1 1 -1)

Phase plot for **a**=(-1 1 -1 1)

Phase plot for **a**=(-1 1 -1 -1)

Phase plot for **a**=(-1 -1 1 1)

Phase plot for **a**=(-1 -1 1 -1)

Phase plot for **a**=(-1 -1 -1 1)

Phase plot for **a**=(-1 -1 -1 -1)

Fig. 3.: Phase plots for systems of the form (6)-(7) and various values of the **a** coefficients.

where each of $\mathbf{a} = (a_{11}, a_{10}, a_{01}, a_{00})$ and $\mathbf{b} = (b_{11}, b_{10}, b_{01}, b_{00})$ can take any value in $\{-1, 1\}^4$. We now have a total of $16 \times 16 = 256$ possible combinations. We plot some of these in Fig. 4 and observe the appearance of rather interesting dynamics. The interpretation of these in terms of social dilemma behaviors is left to the reader.

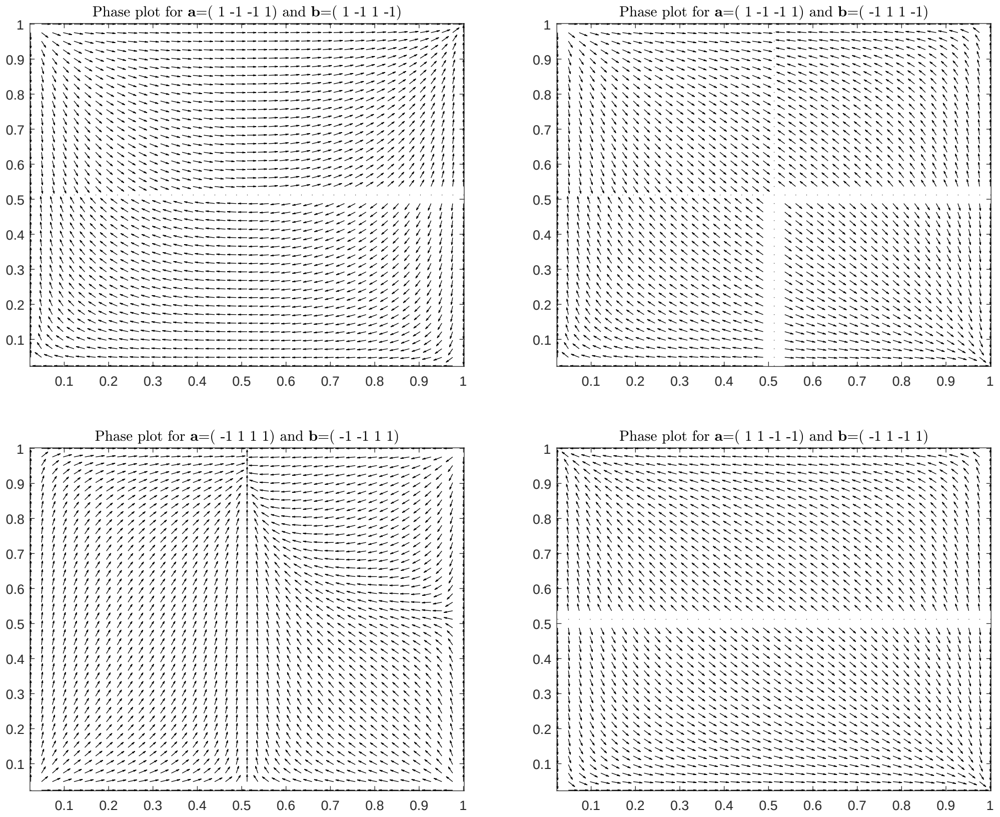


Fig. 4.: Phase plots for systems of the form (6)-(7) and various values of the $\mathbf{a}$ and $\mathbf{b}$ coefficients.

5.4. *General Update Equations*

We can obtain even more general action update equations using the form of (21)-(22) but removing the constraint that $|a_{mn}| = |b_{mn}| = 1$. In this manner we can obtain even more involved dynamics. For example, using

$$\mathbf{a} = (2, -2, 8, 4), \qquad \mathbf{b} = (3, -8, -2, 4)$$

we get the update equations

$$\frac{dx_1}{dt} = (2x_1x_2 - 2x_1(1-x_2) + 8(1-x_1)x_2 + 4(1-x_1)(1-x_2))\,x_1(1-x_1) \tag{23}$$

$$\frac{dx_2}{dt} = (3x_1x_2 - 8(1-x_1)x_2 - 2x_1(1-x_2) + 4(1-x_1)(1-x_2))\,x_2(1-x_2) \tag{24}$$

which have the phase plot of Fig. 5.

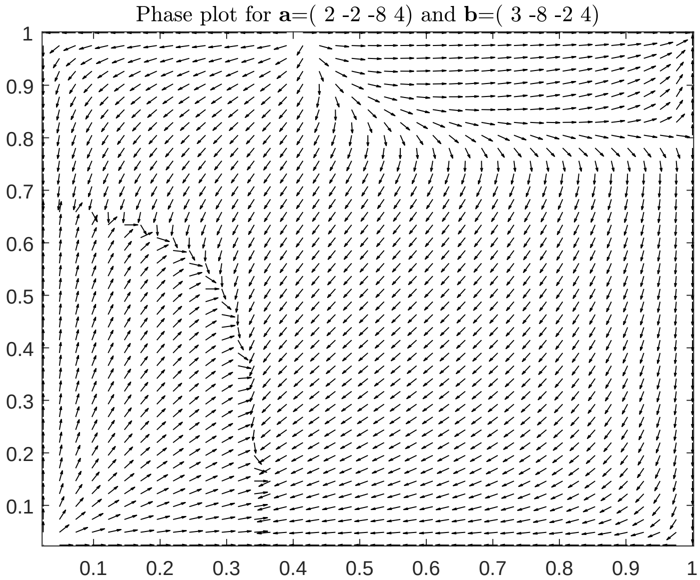


Fig. 5.: Phase plots for the systems of (23)-(24).

5.5. *More than Two Players*

All of the models presented up to this point involved two players. A natural extension is to consider a game played between N players. There are several ways to generalize the update equations (21)-(22). We propose the general form

$$\forall n \in \{1, ..., N\} : \frac{dx_n}{dt} = F_n(\mathbf{x})\, x_n (1 - x_n)$$

and it remains to choose the form of $F_n(\mathbf{x})$. We propose two forms, both of which can be understood as generalizations of the basic two-player model. In both cases we only give the form of $F_1(\mathbf{x})$; the form of the remaining F_n's is obtained similarly.

(1) The first model has the form

$$\forall n \in \{1, ..., N\} : F_1(\mathbf{x}) = a_{11} x_1 \sum_{m \neq n} x_m +$$

$$a_{10} x_n \left(1 - \sum_{m \neq n} x_m\right) + a_{01}(1 - x_n) \sum_{m \neq n} x_m + a_{00}(1 - x_n)\, 1 - \sum_{m \neq n} x_m$$

In this case the change in P_n's cooperation level $\frac{dx_n}{dt}$ depends on x_n itself and on $\sum_{m \neq n} x_m$, the *aggregate* cooperation levels of the other players. Hence $\sum_{m \neq n} x_m$ now plays the role of the single "other" player's cooperation level in equations (21)-(22). For the sake of concereteness, let us consider an example with $N = 3$ and using the symmetric selections $a_{11} = a_{00} = 1$, $a_{10} = a_{01} = -1$. Then, after simplification, the first equation of the first model is

$$\frac{dx_1}{dt} = (4x_1x_2 + 4x_1x_3 - 2x_1 - 2x_2 - 2x_3 + 1)\, x_1 (1 - x_1)$$

By symmetry, the full system is

$$\begin{aligned}
\frac{dx_1}{dt} &= (4x_1 (x_2 + x_3) - 2 (x_1 + x_2 + x_3) + 1)\, x_1 (1 - x_1) \\
\frac{dx_2}{dt} &= (4x_2 (x_3 + x_1) - 2 (x_1 + x_2 + x_3) + 1)\, x_2 (1 - x_2) \\
\frac{dx_3}{dt} &= (4x_3 (x_1 + x_2) - 2 (x_1 + x_2 + x_3) + 1)\, x_3 (1 - x_3)
\end{aligned}$$

(2) The second model has the form:

$$\forall n \in \{1, ..., N\} : F(\mathbf{x}) = a_{11}x_1 \prod_{m \neq n} x_m +$$

$$a_{10}x_n \prod_{m \neq n} (1 - x_m) + a_{01} (1 - x_n) \prod_{m \neq n} x_m + a_{00} (1 - x_n) \prod_{m \neq n} (1 - x_m)$$

In this case, P_1 treats each other player's level of cooperation "individually", rather than "aggregately". For the sake of concreteness, let us consider an example with $N = 3$ and using the symmetric selections $a_{11} = a_{00} = 1$, $a_{10} = a_{01} = -1$. Then the first equation of the second model is

$$\frac{dx_1}{dt} = (2x_1x_2 + 2x_1x_3 + 1 - 2x_1 - x_2 - x_3)\, x_1 (1 - x_1)$$

By symmetry, the full system is

$$\begin{aligned}
\frac{dx_1}{dt} &= (2x_1x_2 + 2x_1x_3 + 1 - 2x_1 - x_2 - x_3)\, x_1 (1 - x_1) \\
\frac{dx_2}{dt} &= (2x_2x_3 + 2x_2x_1 + 1 - 2x_2 - x_3 - x_1)\, x_2 (1 - x_2) \\
\frac{dx_3}{dt} &= (2x_3x_1 + 2x_3x_2 + 1 - 2x_3 - x_1 - x_2)\, x_3 (1 - x_3)
\end{aligned}$$

6. Achieving a Prescribed NE

In our study of the basic model we have seen that, for every $z \in (0, 1)$, the points $\left(z, \frac{1}{2}\right)$ and $\left(\frac{1}{2}, z\right)$ are (unstable) equilibria. It is worth noting that these points are equilibria for *any* payoff values. Now we are going to consider an "inverse" problem: suppose we know that, for a given symmetric continuous social dilemma, we can compute an interior symmetric NE $(b, b) \in [0, 1] \times [0, 1]$. Is there a system of update equations of the form (6)-(7) such that one of its equilibria is (b, b)?

To achieve this goal we must find appropriate values of the coefficients $a_{11}, a_{10}, a_{01}, a_{00}$. Rather than assuming all coefficients take arbitrary values, we will first consider the case $a_{11} = 1$, $a_{10} = -1$, $a_{01} = -1$ and let the only free parameter be $a_{00} = a$. Then we have the following.

Proposition 13. For any $b \in (0,1)$ there is a value of a such that the system

$$\frac{dx_1}{dt} = x_1 x_2 - x_1 (1 - x_2) - (1 - x_1) x_2 + a (1 - x_1)(1 - x_2) \tag{25}$$

$$\frac{dx_2}{dt} = x_1 x_2 - x_2 (1 - x_1) - (1 - x_2) x_1 + a (1 - x_1)(1 - x_2) \tag{26}$$

has an equilibrium $\widehat{\mathbf{x}} = (b, b)$.

Proof. At any *symmetric* equilibrium (x, x) of (25)-(26) we will have

$$0 = xx - x (1 - x) - (1 - x) x + a (1 - x)(1 - x).$$

This equation has solutions

$$r_1 (a) = \frac{1 + a + \sqrt{1 - a}}{3 + a}, \quad r_2 (a) = \frac{1 + a - \sqrt{1 - a}}{3 + a}.$$

Hence for any value of a, $(r_1 (a), r_1 (a))$ and $(r_2 (a), r_2 (a))$ are equilibria of (25)-(26).

Now, $r_1 (a)$ is continuous in $(-\infty, 1)$, with $\lim_{a \to -\infty} r_1 (a) = 1$ and $r_1 (1) = \frac{1}{2}$. Hence for any $\widehat{b} \in [\frac{1}{2}, 1)$ there exists some $\widehat{a}$ such that $r_1 (\widehat{a}) = \widehat{b}$. Similarly, $r_2 (a)$ is continuous in $[0, \frac{1}{2}]$, with $r_2 (0) = 0$ and $r_2 (1) = \frac{1}{2}$. Hence for any $\widehat{b} \in \left[0, \frac{1}{2}\right]$ there exists some $\widehat{a}$ such that $r_2 (\widehat{a}) = \widehat{b}$. □

7. Conclusion

We have studied an action update model for two-player continuous actions social dilemma games, such as Prisoner's Dilemma, Hawk and Dove, Stag Hunt etc. The proposed action updates depend on the cooperation levels of the two players but are independent of specific payoffs; consequently the update equations have the same form for each of the abovementioned games. We have presented a particular action update model, which can be generalized and extended in various ways, as indicated in Section 5; these extensions will be further explored in future publications.

References

1. Axelrod, Robert, and William D. Hamilton. "The evolution of cooperation." *Science* 211.4489 (1981): 1390-1396.
2. Axelrod, Robert. "The emergence of cooperation among egoists." *American Political Science Review* 75.2 (1981): 306-318.
3. E.N. Barron, *Game theory: an introduction.* John Wiley & Sons, 2013.
4. Bowling, Michael, and Manuela Veloso. "Multiagent learning using a variable learning rate." *Artificial Intelligence* 136.2 (2002): 215-250.
5. Chicone, Carmen. *Ordinary differential equations with applications* (2006) Springer.
6. Dawes, Robyn M. "Social dilemmas." *Annual Review of Psychology* 31.1 (1980): 169-193.
7. Dawes, Robyn M., and David M. Messick. "Social dilemmas." *International Journal of Psychology* 35.2 (2000): 111-116.

8. Doebeli, Michael, and Nancy Knowlton. "The evolution of interspecific mutualisms." *Proceedings of the National Academy of Sciences* 95.15 (1998): 8676-8680.
9. M. Frean. "The evolution of degrees of cooperation." *Journal of Theoretical Biology*, vol.182.4 (1996): 549-559.
10. Hofbauer, Josef, and Karl Sigmund. *Evolutionary games and population dynamics.* Cambridge university press, 1998.
11. Hofbauer, Josef, and Karl Sigmund. "Evolutionary game dynamics." *Bulletin of the American Mathematical Society* 40.4 (2003): 479-519.
12. Gallo Jr, Philip S., and Charles G. McClintock. "Cooperative and competitive behavior in mixed-motive games." *Journal of Conflict Resolution* 9.1 (1965): 68-78.
13. T. Killingback and M. Doebeli. "Spatial evolutionary game theory: Hawks and Doves revisited." *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 263.1374 (1996): 1135-1144.
14. Van Lange, Paul AM, et al. "The psychology of social dilemmas: A review." *Organizational Behavior and Human Decision Processes* 120.2 (2013): 125-141.
15. Van Lange, Paul AM, Bettina Rockenbach, and Toshio Yamagishi, eds. *Trust in social dilemmas.* Oxford University Press, 2017.
16. Liebrand, Wim BG. "A classification of social dilemma games." *Simulation & Games* 14.2 (1983): 123-138.
17. Nowak, Martin A., et al. "Emergence of cooperation and evolutionary stability in finite populations." *Nature* 428.6983 (2004): 646-650.
18. G. Owen, *Game theory.* Academic Press, 1968.
19. M. Petrovitsch, "Sur une manière d'étendre le théorème de la moyence aux équations différentielles du premier ordre". *Math. Ann.* , vol. 54.3 (1901): 417–436.
20. T. E. S. Raghavan, "Non-zero-sum two-person games." *Handbook of Game Theory with Economic Applications,* vol. 3 (2002): 1687-1721.
21. Rapoport, Anatol, Albert M. Chammah, and Carol J. Orwant. *Prisoner's dilemma: A study in conflict and cooperation.* University of Michigan press, 1965.
22. Rapoport A. and Guyer M., "A Taxonomy of 2×2 Games", *General Systems*, 11 (1966).
23. Rapoport, Anatol. "Exploiter, leader, hero, and martyr: The four archetypes of the $2\times\ 2$ game." *Behavioral Science* 12.2 (1967): 81-84.
24. G. Roberts and T.N. Sherratt. "Development of cooperative relationships through increasing investment." *Nature*, vol.394.6689 (1998): 175-179.
25. Sandholm, William H. (2010). *Population Games and Evolutionary Dynamics. Economic Learning and Social Evolution*, The MIT Press.
26. Singh, Satinder P., Michael J. Kearns, and Yishay Mansour. "*Nash Convergence of Gradient Dynamics in General-Sum Games.*" UAI. 2000.
27. Skyrms, Brian. *The stag hunt and the evolution of social structure.* Cambridge University Press, 2004.
28. Smith, John Maynard. *Evolution and the Theory of Games.* Cambridge university press, 1982.
29. Strogatz, Steven H. *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering.* CRC Press, 2018.
30. Tanimoto, Jun. *Fundamentals of evolutionary game theory and its applications.* Springer Japan, 2015.
31. T. Verhoeff. "The trader's dilemma: A continuous version of the prisoner's dilemma." *Computing Science Notes*, vol. 93.02 (1998) and Computer Science (2005).
32. Wit, Arjaan P., and Henk AM Wilke. "The effect of social categorization on cooperation in three types of social dilemmas." *Journal of Economic Psychology* 13.1 (1992):

135-151.

33. Wubben, Maarten. Social functions of emotions in social dilemmas. No. EPS-2010-187-ORG. 2010.