$n-$bits reactive strategies in repeated games

Nikoleta E. Glynatsi, Christian Hilbe, Martin Nowak

1 Introduction

In this work we explore *reactive strategies* in the infinitely repeated Prisoner's Dilemma. The Prisoner's Dilemma is a two person symmetric game that provides a simple model of cooperation. Each of the two players, p and q, simultaneously and independently decide to cooperate (C) or to defect (D). In the one shot game there are four possible outcomes, $xy \in \{CC, CD, DC, DD\}$, where x and y represent p's and q's choices respectively. After making a decision each player receives a payoff. The following (2×2) payoff matrix describes the payoffs of p,

$$\begin{array}{c c} & \begin{array}{cc} \text{cooperate} & \text{defect} \end{array} \\ \begin{array}{c} \text{cooperate} \\ \text{defect} \end{array} & \begin{pmatrix} R & S \\ T & P \end{pmatrix} \end{array} \tag{1}$$

where R is the reward payoff for mutual cooperation, T is the temptation payoff, S is the sucker's payoff and P is the punishment for mutual defection. It is assumed that $T > R > P > S$ and $2R > T + S$. The transpose of the payoff matrix (1) is the payoff of q. Alternatively, we can define the payoff vectors for each player by,

$$\mathbf{S_p} = (R, S, T, P) \quad \text{and} \quad \mathbf{S_q} = (R, T, S, P). \tag{2}$$

A special case of the Prisoner's Dilemma is the donation game. In the donation game each player can chose to cooperate and pay a cost c such that their co-player gets a benefit b. Given this, the payoff matrix (1) is re-written as,

$$\begin{array}{c c} & \begin{array}{cc} \text{cooperate} & \text{defect} \end{array} \\ \begin{array}{c} \text{cooperate} \\ \text{defect} \end{array} & \begin{pmatrix} b-c & -c \\ c & 0 \end{pmatrix} \end{array} \tag{3}$$

where $b, c > 0$ and $b > c$, and equivalently,

$$\mathbf{S_p} = (b-c, -c, b, 0) \quad \text{and} \quad \mathbf{S_q} = (b-c, b, -c, 0). \tag{4}$$

We study the infinitely repeated Prisoner's Dilemma where players are infinitely interacting, choosing between C and D at each round. The long term payoffs of the players are denoted as s_p and s_q.

1.1 Strategies

There are infinitely many strategies for the repeated Prisoner's Dilemma. A strategy is a mapping from the entire history of play to an action of the stage game. We focus on *reactive* strategies, a set of strategies that map the previous actions of the co-player to an action. Reactive strategies are a special case of memory-one strategy, which are a set of strategies well studied in the literature. Though reactive strategies have gained some attention, the majority of work focuses on the case of memory size of one. Thus, they focus on reactive strategies that only consider the last action of the co-player. These will be reefer to as *one-bit reactive strategies*, and so, n-bit reactive strategies, are strategies that consider the last n rounds of the co-player.

The aim of this work is to extensively study reactive strategies of higher memory. In section 2.1 we analytically characterize reactive strategies that are good and of Nash type. In section we explore reactive strategies that can sustain pure Nash equilibria in an environment with noise. In section we perform an evolutionary analysis, and investigate which strategies evolve.

2 Results

2.1 Good two-bit reactive strategies

In [1], Akin defined what it means for s strategy to be *good* and of *Nash type*. His original work focused on memory-one strategies. With the outcomes listed in order as CC, CD, DC, DD a memory-one strategy for p is a vector $p = (p_1, p_2, p_3, p_4)$ where p_i is the probability of playing C when the i^{th} outcome occurred in the previous round. Initially, Akin introduces the notion of *agreeable* strategies. A strategy is agreeable if it always cooperates following a mutual cooperation, in the case of memory-one strategies $p_1 = 1$. Give this the definition of good and Nash type strategies is as follows.

Definition 2.1. A strategy for p is called good if it is agreeable and if for any general strategy chosen by q against it, the expected payoffs satisfy:

$$s_q \geq (b - c) \Rightarrow s_q = s_p = (b - c). \tag{5}$$

The strategy is called of Nash type if it is agreeable and if the expected payoffs against any q general strategy satisfy:

$$s_q \geq R \Rightarrow s_q = (b - c). \tag{6}$$

where s_p and s_q are the long term payoffs of p and q respectively.

Hence, a good strategy is a strategy for which the co-player can achieve the reward payoff in the long run if and only if the player receives the reward payoff as well. A strategy that is of Nash type, is a strategy which reassures that the co-player can never receive a payoff higher than $b - c$. As we can see, the definitions make no assumptions regarding the type of strategies the players need to play, and thus, these definitions are extendable to $n-$ bit reactive strategies. The definition of agreeable strategies needs to be properly defined in the case of higher memory, but we discuss this in the section that follows.

Akin proceeds to proof an interesting result regarding the long term expected states of the game. A play between two memory-one strategies, $p = (p_1, p_2, p_3, p_4)$ and $q = (q_1, q_2, q_3, q_4)$, follows a Markov chain with four states corresponding to the possible outcomes, and the transition matrix M^1. The stationary distribution of the Markov process, denoted as $\mathbf{v}^1$, is the solution to $\mathbf{v}^1 M^1 = \mathbf{v}^1$. Note that $\sum_{i=1}^{4} u_i = 1$. $\mathbf{v}^1$ gives the probability of the strategies being in each of the four possible outcomes at the end of the game.

$$M_1 = \begin{bmatrix} p_1 q_1 & p_1(1-q_1) & q_1(1-p_1) & (1-p_1)(1-q_1) \\ p_2 q_3 & p_2(1-q_3) & q_3(1-p_2) & (1-p_2)(1-q_3) \\ p_3 q_2 & p_3(1-q_2) & q_2(1-p_3) & (1-p_3)(1-q_2) \\ p_4 q_4 & p_4(1-q_4) & q_4(1-p_4) & (1-p_4)(1-q_4) \end{bmatrix}. \tag{7}$$

The, given the stationary distribution vector $\mathbf{v}^1$, Akin proved Theorem 2.1.

Theorem 2.1. Akin's Theorem. Assume that $\tilde{p}$ uses the strategy $\tilde{p} = (\tilde{p}_1, \tilde{p}_2, \tilde{p}_3, \tilde{p}_4)$ where $\tilde{p}_1 = p_1 - 1$ and $\tilde{p}_2 = p_2 - 1$, and $\mathbf{v}^1$ is the stationary distribution for a given co-player q then,

$$< \mathbf{v}\tilde{p} >= v_1\tilde{p_1} + v_2\tilde{p_2} + v_3\tilde{p_3} + v_4\tilde{p_4} = 0 \Leftrightarrow \tag{8}$$

$$v_1(p_1 - 1) + v_2(p_2 - 1) + v_3 p_3 + v_4 p_4 = 0. \tag{9}$$

All of the above results are extendable to higher memory strategies. Here we demonstrate the special of two-bit reactive strategies.

2.2 Two-bit reactive strategies

In the case of *two-bit reactive strategies*, players base their decisions on the actions of the co-player in the previous two rounds. Since for a single round there are 4 possible outcomes, for two rounds there will be 16 (4×4). We denote the states as $E_x E_y | F_x F_y$ $(E_x, E_y, F_x, F_y \in \{C, D\})$ where the outcome of the previous round is $E_x E_y$ and the outcome of the current round is $F_x F_y$.

With the two previous actions of the co-player listed in order as CC, CD, DC, DD a two-bit reactive strategy p is a vector $p = (p_1, p_2, p_3, p_4, p_3, p_4 p_1, p_2, p_3, p_4, p_3, p_4)$ where p_i is the probability of playing C when the i^{th} ordered actions were played by the co-player in the previous two rounds. The play between two two-bits reactive strategies can be described by a Markov process with the transition matrix M^2.

$$M^2 = \begin{pmatrix}
p_1q_1 & p_1(1-q_1) & (1-p_1)q_1 & (1-p_1)(1-q_1) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & p_2q_1 & p_2(1-q_1) & (1-p_2)q_1 & (1-p_2)(1-q_1) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_1q_2 & p_1(1-q_2) & (1-p_1)q_2 & (1-p_1)(1-q_2) & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_2q_2 & p_2(1-q_2) & (1-p_2)q_2 & (1-p_2)(1-q_2) \\
p_3q_1 & p_3(1-q_1) & (1-p_3)q_1 & (1-p_3)(1-q_1) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & p_4q_1 & p_4(1-q_1) & (1-p_4)q_1 & (1-p_4)(1-q_1) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_3q_2 & p_3(1-q_2) & (1-p_3)q_2 & (1-p_3)(1-q_2) & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_4q_2 & p_4(1-q_2) & (1-p_4)q_2 & (1-p_4)(1-q_2) \\
p_1q_3 & p_1(1-q_3) & (1-p_1)q_3 & (1-p_1)(1-q_3) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & p_2q_3 & p_2(1-q_3) & (1-p_2)q_3 & (1-p_2)(1-q_3) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_1q_4 & p_1(1-q_4) & (1-p_1)q_4 & (1-p_1)(1-q_4) & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_2q_4 & p_2(1-q_4) & (1-p_2)q_4 & (1-p_2)(1-q_4) \\
p_3q_3 & p_3(1-q_3) & (1-p_3)q_3 & (1-p_3)(1-q_3) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & p_4q_3 & p_4(1-q_3) & (1-p_4)q_3 & (1-p_4)(1-q_3) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_3q_4 & p_3(1-q_4) & (1-p_3)q_4 & (1-p_3)(1-q_4) & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_4q_4 & p_4(1-q_4) & (1-p_4)q_4 & (1-p_4)(1-q_4)
\end{pmatrix}.$$

The transition matrix M^2 is a (16×16) matrix where

$$M(v_{n+1}^2 = G_x G_y | H_x H_y | v_n^2 = E_x E_y | F_x F_y = 0) \text{ if } G_x G_y \neq F_x F_y, \tag{10}$$

so in each row of the matrix there will be at most four nonzero elements. The stationary distribution, denoted as $\mathbf{v}^2$, is the solution to $\mathbf{v}^2 M^2 = \mathbf{v}^2$ where $\sum_{i=1}^{16} u_i = 1$.

Because of the nature of the transition matrix M^2, and more specifically, the condition (10), the following holds for the stationary distribution,

$$\sum_{i,j\in\{C,D\}} v_{i,j|CD} = \sum_{i,j\in\{C,D\}} v_{CD|ij} \Leftrightarrow \tag{11}$$
$$v_1 + v_5 + v_9 + v_{13} = v_1 + v_2 + v_3 + v_4, \tag{12}$$
$$v_2 + v_4 + v_6 + v_{14} = v_5 + v_6 + v_7 + v_8, \tag{13}$$
$$v_3 + v_7 + v_{11} + v_{15} = v_9 + v_{10} + v_{11} + v_{12} \text{ and} \tag{14}$$
$$v_4 + v_8 + v_{12} + v_{16} = v_{13} + v_{14} + v_{15} + v_{16}. \tag{15}$$

The extension to Akin's Theorem (Theorem 2.1) is give by Lemma 2.2.

Lemma 2.2. Assume that $\tilde{p}$ uses the strategy $\tilde{p} = (\tilde{p}_1, \tilde{p}_2, \tilde{p}_1, \tilde{p}_2, \dots, \tilde{p}_3, \tilde{p}_4)$ where $\tilde{p}_1 = p_1 - 1$ and $\tilde{p}_2 = p_1 - 1$, and $\mathbf{v}^2$ is the stationary distribution for a given co-player $q = (q_1, q_2, q_1, q_2, \dots, q_3, q_4)$ then,

$$< \mathbf{v}\tilde{p} >= v_1\tilde{p_1} + v_2\tilde{p_2} + v_3\tilde{p_1} + v_4\tilde{p_2}, \dots + v_{13}\tilde{p_3} + v_{14}\tilde{p_4} + v_{15}\tilde{p_3} + v_{16}\tilde{p_4} = 0 \Leftrightarrow \tag{16}$$
$$(v_1 + v_9)(p_1 - 1) + (v_2 + v_{10})(p_2 - 1) + (v_5 + v_{13})(p_3 - 1) + (v_6 + v_{14})(1 - p_4) \tag{17}$$
$$+(v_3 + v_{11})p_1 + (v_4 + v_{12})p_2 + (v_7 + v_{15})p_3 + (v_8 + v_{16})p_4 = 0. \tag{18}$$

The long run steady state probability vector $\mathbf{v}^2$ can be combined with the payoff vectors to give the expected payoffs for each player, s_p and s_q. In the case of the two-bit reactive strategies, there can be two set of payoff vectors which depends if payoffs are defined based on the outcome of the last round, or of the second to last. Thus payoff vectors can be one of the following,

$$\begin{aligned} S_p &= (b-c, -c, b, 0, b-c, -c, b, 0, b-c, -c, b, 0, b-c, -c, b, 0) \text{ and} \\ S_q &= (b-c, b, -c, 0, b-c, b, -c, 0, b-c, b, -c, 0, b-c, b, -c, 0) \end{aligned} \tag{19}$$

or

$$\begin{aligned} S'_p &= (b-c, b-c, b-c, b-c, -c, -c, -c, -c, b, b, b, b, 0, 0, 0, 0) \text{ and} \\ S'_q &= (b-c, b-c, b-c, b-c, b, b, b, b, -c, -c, -c, -c, 0, 0, 0, 0). \end{aligned} \tag{20}$$

Note that $s_p = \mathbf{v}^2 \times S_p = \mathbf{v}^2 \times S'_p$ and $s_q = \mathbf{v}^2 \times S_q = \mathbf{v}^2 \times S'_q$.

2.3 Good Nash

We are interested in which two-bit reactive strategies can sustain a Nash equilibrium, and more specifically, a cooperative one. An agreeable strategy in the case of two-bit reactive strategies is a play that always cooperates after two consecutive cooperations of the co-player, thus $p_1 = 0$. We know that p is of Nash type if against a co-player q the long term payoff of the co-player $s_q \leq (b - c)$.

$$s_q - (b-c) = \mathbf{v}^2 \times S_q - (b-c)\sum_{i=1}^{16} u_i \tag{21}$$
$$= (v_2 + v_6 + v_{10} + v_{14})c + (c-b)(v_4 + v_8 + v_{12} + v_{16}) - b(v_3 + v_7 + v_{11} + v_{15}).$$

This is derived given that payoff vectors are of Eq. (19), if the payoff vectors of Eq. (20) are used instead,

$$s_q - \leq (b-c) = \tag{22}$$
$$= (v_5 + v_6 + v_7 + v_8)c + (c-b)(v_{13} + v_{14} + v_{15} + v_{16}) - b(v_9 + v_{10} + v_{11} + v_{12}).$$

Note that Eq. (21) and Eq. (21) are equivalent if we substitute Eq. (12).

2.3.1 Special case $p_3 = p_2$

In the special case of $p_3 = p_2$ a player does not care about when a defection occurred, but that a defection happened.

Theorem 2.3. Let $p = (p_1, p_2, p_1, p_2, p_3, p_4, p_3, p_4, p_1, p_2, p_1, p_2, p_3, p_4, p_3, p_4)$ be an agreeable plan, that is, $p_1 = 1$. Strategy p is of Nash type iff the following inequalities hold.

$$p_2 \geq p_4 \text{ and } p_2 \leq 1 - \frac{c}{b} \text{ and } p_4 \leq 1 - \frac{c}{b}.$$

Strategy p is good iff, in addition, both inequalities are strict.

Proof. □

A Two-bit reactive strategies cheat sheet.

References

[1] Ethan Akin. The iterated prisoner's dilemma: good strategies and their dynamics. *Ergodic Theory, Advances in Dynamical Systems*, pages 77–107, 2016.

History/State	State number	Coop/Def	Coop. probability M.	Coop. probability	Reaction to
$CC\|CC$	1	Coop	p_1	p_1	CC
$CC\|CD$	2	Coop	p_2	p_2	CD
$CC\|DC$	3	Def	p_3	p_1	CC
$CC\|DD$	4	Def	p_4	p_2	CD
$CD\|CC$	5	Coop	p_5	p_3	DC
$CD\|CD$	6	Coop	p_6	p_4	DD
$CD\|DC$	7	Def	p_7	p_3	DC
$CD\|DD$	8	Def	p_8	p_4	DD
$DC\|CC$	9	Coop	p_9	p_1	CC
$DC\|CD$	10	Coop	p_{10}	p_2	CD
$DC\|DC$	11	Def	p_{11}	p_1	CC
$DC\|DD$	12	Def	p_{12}	p_2	CD
$DD\|CC$	13	Coop	p_{13}	p_3	DC
$DD\|CD$	14	Coop	p_{14}	p_4	DD
$DD\|DC$	15	Def	p_{15}	p_3	DC
$DD\|DD$	16	Def	p_{16}	p_4	DD

Table 1: **Cheat Sheet** for two-bit reactive strategies.