Good reactive strategies in the iterated prisoner's dilemma

Nikoleta E. Glynatsi, Christian Hilbe, Martin Nowak

The aim of this work is to characterize all good Nash reactive strategies, with memory two ($n = 2$), in the infinitely repeated prisoner's dilemma. We refer to these strategies as two-bit reactive strategies.

1 Two-bits reactive strategies

In the case of $n = 2$ there are 16 possible outcomes. We denote the possible outcomes as $E_pE_q|F_pF_q$ ($E_p, E_q, F_p, F_q \in \{C, D\}$) where the outcome of the previous round is E_pE_q and the outcome of the current round is F_pF_q. With the outcomes listed in order as $CC|CC, CC|CD, \dots, DD|DC, DD|DD$ a two-bit reactive strategy for p is a vector $\mathbf{p} = (p_1, p_2, p_1, p_2, p_3, p_4, p_3, p_4, p_1, p_2, p_1, p_2, p_3, p_4, p_3, p_4)$ where p_1 is the probability cooperating when the last two actions of the co-player were C and C, p_2 is the probability cooperating when the last two actions of the co-player were C and D, and so on. For simplicity, we denote a two-bit reactive strategy for p as $\hat{\mathbf{p}} = (\hat{p}_1, \hat{p}_2, \hat{p}_3, \hat{p}_4)$.

Definition 1.1. We call a two-bit reactive strategy **agreeable** if it is never the first to defect, and if it always cooperates with a probability 1 if the co-player has consecutively cooperated in that last two rounds, thus $\hat{p}_1 = 1$.

The play between a pair of two-bit reactive strategies can be described by a Markov process with the transition matrix M.

$$M = \left(\begin{smallmatrix}
\hat p_1\hat q_1 & \hat p_1(1-\hat q_1) & (1-\hat p_1)\hat q_1 & (1-\hat p_1)(1-\hat q_1) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \hat p_2\hat q_1 & \hat p_2(1-\hat q_1) & (1-\hat p_2)\hat q_1 & (1-\hat p_2)(1-\hat q_1) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \hat p_1\hat q_2 & \hat p_1(1-\hat q_2) & (1-\hat p_1)\hat q_2 & (1-\hat p_1)(1-\hat q_2) & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \hat p_2\hat q_2 & \hat p_2(1-\hat q_2) & (1-\hat p_2)\hat q_2 & (1-\hat p_2)(1-\hat q_2) \\
\hat p_3\hat q_1 & \hat p_3(1-\hat q_1) & (1-\hat p_3)\hat q_1 & (1-\hat p_3)(1-\hat q_1) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \hat p_4\hat q_1 & \hat p_4(1-\hat q_1) & (1-\hat p_4)\hat q_1 & (1-\hat p_4)(1-\hat q_1) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \hat p_3\hat q_2 & \hat p_3(1-\hat q_2) & (1-\hat p_3)\hat q_2 & (1-\hat p_3)(1-\hat q_2) & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \hat p_4\hat q_2 & \hat p_4(1-\hat q_2) & (1-\hat p_4)\hat q_2 & (1-\hat p_4)(1-\hat q_2) \\
\hat p_1\hat q_3 & \hat p_1(1-\hat q_3) & (1-\hat p_1)\hat q_3 & (1-\hat p_1)(1-\hat q_3) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \hat p_2\hat q_3 & \hat p_2(1-\hat q_3) & (1-\hat p_2)\hat q_3 & (1-\hat p_2)(1-\hat q_3) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \hat p_1\hat q_4 & \hat p_1(1-\hat q_4) & (1-\hat p_1)\hat q_4 & (1-\hat p_1)(1-\hat q_4) & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \hat p_2\hat q_4 & \hat p_2(1-\hat q_4) & (1-\hat p_2)\hat q_4 & (1-\hat p_2)(1-\hat q_4) \\
\hat p_3\hat q_3 & \hat p_3(1-\hat q_3) & (1-\hat p_3)\hat q_3 & (1-\hat p_3)(1-\hat q_3) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \hat p_4\hat q_3 & \hat p_4(1-\hat q_3) & (1-\hat p_4)\hat q_3 & (1-\hat p_4)(1-\hat q_3) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \hat p_3\hat q_4 & \hat p_3(1-\hat q_4) & (1-\hat p_3)\hat q_4 & (1-\hat p_3)(1-\hat q_4) & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \hat p_4\hat q_4 & \hat p_4(1-\hat q_4) & (1-\hat p_4)\hat q_4 & (1-\hat p_4)(1-\hat q_4)
\end{smallmatrix}\right).$$

The invariant distribution $\mathbf{v}$ is the solution to $\mathbf{v}M = \mathbf{v}$. We define expected payoffs, denoted as $\mathbf{s}_p$ and $\mathbf{s}_q$, based on the outcome of the last round. Thus,

$$\mathbf{s}_p = \mathbf{v} \cdot \mathbf{S}_p \quad \text{and} \quad \mathbf{s}_q = \mathbf{v} \cdot \mathbf{S}_q.$$

where

$$\begin{aligned}
\mathbf{S}_p &= (b-c, \ -c, \ b, \ 0, \ b-c, \ -c, \ b, \ 0, \ b-c, \ -c, \ b, \ 0, \ b-c, \ -c, \ b, \ 0)\,, \\
\mathbf{S}_q &= (b-c, \ b, \ -c, \ 0, \ b-c, \ b, \ -c, \ 0, \ b-c, \ b, \ -c, \ 0, \ b-c, \ b, \ -c, \ 0),
\end{aligned} \tag{1}$$

and $b > c > 0$.

From [Akin, 2016], a strategy for p is called **good** if (i) it is agreeable, and (ii) if for any general strategy chosen by q against it the expected payoffs satisfy:

$$s_{\mathbf{q}} \geq b-c \qquad \Rightarrow \qquad s_{\mathbf{q}} = s_{\mathbf{p}} = b-c, \tag{2}$$

and the strategy is of **Nash type** if (i) it is agreeable and (ii) if the expected payoffs against any general strategy used by q satisfy:

$$s_{\mathbf{q}} \geq b-c \qquad \Rightarrow \qquad s_{\mathbf{q}} = b-c. \tag{3}$$

Hence, a strategy is good if the co-player achieves the reward payoff if and only if the focal player does as well, and a Nash type strategy reassures that the co-player can never receive a payoff higher than $b-c$ (the payoff for mutual cooperation).

[Akin, 2016] derives an interesting relationship between a player's memory-one strategy (Theorem 1.3) and the resulting invariant distribution of the repeated game. This relationship allows him to characterize all memory-one strategies that are of *Nash type* and *good*. In order to characterize all two-bit reactive strategies that are good we initially extend Theorem 1.3 from [Akin, 2016] to two-bit reactive strategies (Lemma 1.1). We then use Lemma 1.1 to prove Theorem 1.2.

Lemma 1.1. Assume that player p uses a two-bit reactive strategy $\hat{\mathbf{p}}$, and q uses a strategy that leads to a sequence of distributions $\{\mathbf{v}^{(n)}, n = 1, 2, ...\}$ with $\mathbf{v}^{(k)}$ representing the distribution over the states in the k^{th} round of the game. Let $\mathbf{v}$ be the associated stationary distribution, and let $\tilde{\mathbf{p}} = \hat{\mathbf{p}} - \hat{\mathbf{e}}_{12}$ where $\hat{\mathbf{e}}_{12} = (1, 1, 1, 1, 0, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 0)$. Then,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbf{v}^{(k)} \cdot \tilde{\mathbf{p}} = 0, \text{ and therefore } \mathbf{v} \cdot \tilde{\mathbf{p}} = 0.$$

$$\begin{aligned}\mathbf{v}^{(n)} \cdot \tilde{\mathbf{p}} = 0 \Rightarrow & \\ &(v_1 + v_9)(1 - \hat{p}_1) + (v_2 + v_{10})(1 - \hat{p}_2) + (v_5 + v_{13})(1 - \hat{p}_3) + (v_6 + v_{14})(1 - \hat{p}_4) \\ &+ (v_3 + v_{11})\hat{p}_1 + (v_4 + v_{12})\hat{p}_2 + (v_7 + v_{15})\hat{p}_3 + (v_8 + v_{16})\hat{p}_4 = 0.\end{aligned} \tag{4}$$

Proof. The probability that p cooperates in the n^{th} round, denoted by $v_{\text{C}}^{(n)}$, is $v_{\text{C}}^{(n)} = v_1^{(n)} + v_2^{(n)} + v_5^{(n)} + v_6^{(n)} + v_9^{(n)} + v_{10}^{(n)} + v_{13}^{(n)} + v_{14}^{(n)} = v \cdot \hat{\mathbf{e}}_{12}$ where $\hat{\mathbf{e}}_{12} = (1, 1, 1, 1, 0, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 0)$. The probability that p cooperates in the $(n + 1)^{th}$ round, denoted by $v_{\text{C}}^{(n+1)} = v^{(n)} \cdot \hat{\mathbf{p}}$. Hence,

$$v_{\text{C}}^{(n+1)} - v_{\text{C}}^{(n)} = \mathbf{v}^{(\mathbf{n})} \cdot \hat{\mathbf{p}} - v \cdot \hat{\mathbf{e}}_{12} = \mathbf{v}^{(\mathbf{n})} \cdot (\hat{\mathbf{p}} - \hat{\mathbf{e}}_{12}) = v^{(n)} \cdot \tilde{\mathbf{p}}.$$

This implies,

$$\sum_{k=1}^{n} v^{(k)} \cdot \tilde{\mathbf{p}} = \sum_{k=1}^{n} v_{\text{C}}^{(k+1)} - v_{\text{C}}^{(k)} \quad \Rightarrow \quad \sum_{k=1}^{n} v^{(k)} \cdot \tilde{\mathbf{p}} = v_{\text{C}}^{(n+1)} - v_{\text{C}}^{(1)}. \tag{5}$$

As the right side has absolute value at most 1,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} v^{(k)} \cdot \tilde{\mathbf{p}} = 0. \tag{6}$$

□

Theorem 1.2. Let the two-bit reactive strategy $\hat{\mathbf{p}} = (\hat{p}_1, \hat{p}_2, \hat{p}_3, \hat{p}_4)$ be an **agreeable strategy**; that is $\hat{p}_1 = 1$. Strategy $\hat{\mathbf{p}}$ is **Nash** if the following inequalities hold:

$$\hat{p}_4 \leq 1 - \frac{c}{b} \qquad \hat{p}_2 \leq \hat{p}_4 \qquad \hat{p}_3 \leq 1 \qquad 1 + \hat{p}_2 \leq \frac{b}{c} - \hat{p}_4 \frac{b-c}{c}$$

The agreeable strategy $\hat{\mathbf{p}}$ is good if the inequalities above are strict.

Proof. We first eliminate the possibility $\hat{p}_4 = 1$. If $\hat{p}_4 = 1$, then $\hat{\mathbf{p}} = (1, \hat{p}_2, \hat{p}_3, 1)$. If against this q plays AllD $= (0, 0, 0, 0)$, then $v_6 = 1$. So the strategies end up in the outcome $CD|CD$ with with $s_{\mathbf{q}} = b$ and $s_{\mathbf{p}} = -c$. Hence, $\hat{\mathbf{p}}$ is not of Nash type.

We now assume $1 - \hat{p}_4 > 0$. Observe that

$$\begin{aligned} s_{\mathbf{q}} - (b-c) &= \mathbf{v} \times \mathbf{S}_q - (b-c) \sum_{i=1}^{16} v_i \\ &= (v_2 + v_6 + v_{10} + v_{14})c + (c - b)(v_4 + v_8 + v_{12} + v_{16}) - b(v_3 + v_7 + v_{11} + v_{15}). \end{aligned} \tag{7}$$

Multiplying by the positive quantity $(1 - \hat{p}_4)$ and collecting terms, we have

$$\begin{aligned} &s_{\mathbf{q}} - (b-c) \geq 0 \Rightarrow \\ &(1 - \hat{p}_4)(v_6 + v_{14})c \geq -c(1 - \hat{p}_4)(v_2 + v_{10}) + (1 - \hat{p}_4)(-c + b)(v_4 + v_8 + v_{12} + v_{16}) + b(1 - \hat{p}_4)(v_3 + v_7 + v_{11} + v_{15}). \end{aligned} \tag{8}$$

Since $\tilde{p_1} = 0$, equation (4) implies

$$(1-\hat{p}_2)(v_{10}+v_2)+(1-\hat{p}_3)(v_{13}+v_5)+(1-\hat{p}_4)(v_{14}+v_6)-\hat{p}_2(v_{12}+v_4)-\hat{p}_3(v_{15}+v_7)-\hat{p}_4(v_{16}+v_8)-v_{11}-v_3 = 0,$$

and so,

$$(1-\hat{p}_4)(v_{14}+v_6) = -((1-\hat{p}_2)(v_{10}+v_2)+(1-\hat{p}_3)(v_{13}+v_5)-\hat{p}_2(v_{12}+v_4)-\hat{p}_3(v_{15}+v_7)-\hat{p}_4(v_{16}+v_8)-v_{11}-v_3).$$

Substituting this in the above inequality and collecting terms we get,

$$A(v_{10} + v_2) + B(v_{12} + v_4) + C(v_{13} + v_5) + D(v_{15} + v_7) + E(v_{11} + v_{16} + v_3 + v_8) \geq 0 \tag{9}$$

with

$$A = c(\hat{p}_2 - \hat{p}_4), \qquad B = c(1 + \hat{p}_2 - \hat{p}_4) + b(-1 + \hat{p}_4), \qquad C = c(-1 + \hat{p}_3),$$
$$D = c\hat{p}_3 + b(-1 + \hat{p}_4), \qquad E = c + b(-1 + \hat{p}_4).$$

In the case where A, B, C, D and E are strictly smaller than 0, condition (9) holds iff $v_2, v_3, v_4, v_5, v_7, v_8, v_{10}, v_{11}, v_{12}, v_{13}, v_{15}, v_{16} = 0$. This implies, that $(v_1 + v_9)(1 - \hat{p}_1) + (v_6 + v_{14})(1 - \hat{p}_4) = 0$. $\hat{p}_4$ can not be 1, thus $v_6, v_{14} = 0$. This means $(v_1 + v_9) = 1$, so both players receive the reward payoff and $\hat{\mathbf{p}}$ is good.

For $A, B, C, D, E \leq 0$ we derive the following conditions,

$$\hat{p}_4 \leq 1 - \frac{c}{b} \tag{10}$$

$$\hat{p}_2 \leq \hat{p}_4 \tag{11}$$

$$\hat{p}_3 \leq 1 \tag{12}$$

$$1 + \hat{p}_2 \leq \frac{b}{c} - \hat{p}_4 \cdot \frac{b-c}{c} \tag{13}$$

□

1.1 Numerical Evaluation

To verify the result of Theorem 1.2 we explored which agreeable strategies are Nash numerically. Namely, for a given agreeable two-bit strategy, and we checked if condition $\pi(\mathbf{q}, \hat{\mathbf{p}}) \leq (b-c)$ was satisfied against all pure memory-two strategies ($\mathbf{q} \in \{0, 1\}^{16}$). We recorded if the strategy was Nash or not, and against which the pure strategies the condition for Nash was not satisfied. We repeated this step for 10,000 random strategies, for parameter values ($b = 2$ and $c = 1$). The results are shown in Figure 1. From Figure 1**B)** we can conclude that the inequalities (10) are sufficient for a point to be Nash but not necessary.

Since a two-bit reactive strategy $\hat{\mathbf{p}} = (\hat{p}_1, \hat{p}_2, \hat{p}_3, \hat{p}_4)$ can only be a Nash equilibrium if *no* other strategy yields a larger payoff, in particular neither AllD nor the Alternator strategy must yield a larger payoff, where AllD$= (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$ and Alternator$= (0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1)$. We conclude that an agreeable two-bit reactive strategy $\hat{\mathbf{p}} = (\hat{p}_1, \hat{p}_2, \hat{p}_3, \hat{p}_4)$ can only form a Nash equilibrium if

$$\pi(\text{AllD}, \hat{\mathbf{p}}) \leq b-c \quad \text{and} \quad \pi(\text{Alternator}, \hat{\mathbf{p}}) \leq b-c,$$

or equivalently, if

$$\hat{p}_4 \leq 1 - \frac{c}{b} \quad \text{and} \quad \hat{p}_2 + \hat{p}_3 \leq 1 + \frac{b-c}{c} \tag{14}$$

In fact, a further numerical analysis suggests the following stronger result.

Conjecture 1.3. An agreeable two-bit reactive strategy $\hat{\mathbf{p}} = (\hat{p}_1, \hat{p}_2, \hat{p}_3, \hat{p}_4)$ is of Nash type if and only if conditions (14) hold.

References

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

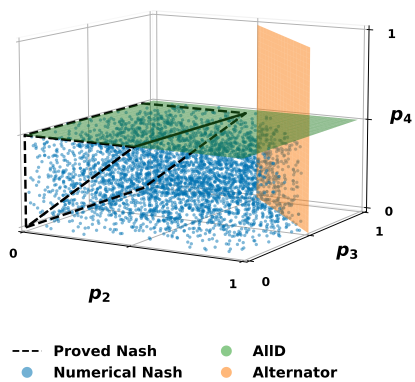

Figure 1: **Nash results for two-bit strategies. A) Proved Nash.** We have shown that if a two-bit reactive strategy is within this space, thus satisfies conditions (10), then it is good Nash. **B) Numerical Nash results.** The results from the numerical evaluation. We evaluated 10^4 points in the space. The numerical results have shown that there are two pure strategies that constrain the Nash space; these are AllD and Alternator. The equations for the planes are obtained by solving $\pi(\mathbf{q}, \hat{\mathbf{p}}) = (b-c)$. The equations are $\hat{p}_4 = 1 - \frac{c}{b}$ and $\hat{p}_3 = 1 + \frac{b-c}{c} - \hat{p}_2$. Parameters: $c = 1, b = 2$. **C) Numerical Nash for high benefit.** We repeat the numerical analysis for a higher value of benefit ($b = 7$) for 10^3 random points. We can see that the strategies AllD and Alternator still constrain the space of possible Nash. Note that we do not plot $\hat{p}_1$ for any of the above plots, since $\hat{p}_1 = 1$.