Reactive strategies with longer memory

Nikoleta E. Glynatsi, Ethan Akin, Martin Nowak, Christian Hilbe

1 Formal Model

We consider the infinitely repeated games among two players, player p and player q. Each round, they engage in the donation game with payoff matrix

$$\begin{pmatrix} b-c & -c \\ b & 0 \end{pmatrix}. \tag{1}$$

Here b and c denote the benefit and the cost of cooperation, respectively. We assume $b > c > 0$ throughout. Therefore, the payoff matrix (1) is a special case of a prisoner's dilemma with payoff matrix,

$$\begin{pmatrix} R & S \\ T & P \end{pmatrix}, \tag{2}$$

with $T > R > S > P$ and $2R > T + S$.

Memory-n strategies. We assume in the following, that the players' decisions only depend on the outcome of the previous n rounds. To this end, an *n-history for player p* is a string $h^p = (a^p_{-1}, \ldots, a^p_{-n}) \in \{C, D\}^n$. An entry a^p_{-k} corresponds to player p's action k rounds ago. Let H^p denote the space of all n-histories of player p. Analogously, we define H^q as the set of n-histories h^q of player q. A pair $h = (h^p, h^q)$ is called an *n-history of the game*. We use $H = H^p \times H^q$ to denote the space of all such histories. This set contains $|H| = 2^{2n}$ elements.

A *memory-n* strategy is a vector $\mathbf{m} = (m_h)_{h \in H} \in [0, 1]^{2n}$. Each entry m_h corresponds to the player's cooperation probability in the next round, depending on the outcome of the previous n rounds.

If the two players use memory-n strategies $\mathbf{m}$ and $\mathbf{m}'$, one can represent the interaction as a Markov chain with a $2^{2n} \times 2^{2n}$ transition matrix M. Let $\mathbf{v} = (v_h)_{h \in H}$ be an invariant distribution of this Markov chain. Based on the invariant distribution $\mathbf{v}$, we can also compute the players' payoffs. To this end, let $\mathbf{S}^k = (S^k_h)_{h \in H}$ denote the vector that returns for each h the one-shot payoff that player p obtained k rounds ago,

$$S^k_h = \begin{cases} b-c & \text{if } a^p_{-k} = C \text{ and } a^q_{-k} = C \\ -c & \text{if } a^p_{-k} = C \text{ and } a^q_{-k} = D \\ b & \text{if } a^p_{-k} = D \text{ and } a^q_{-k} = C \\ 0 & \text{if } a^p_{-k} = D \text{ and } a^q_{-k} = D \end{cases} \tag{3}$$

Then we can define player p's repeated-game payoff $s_{\mathbf{m},\mathbf{m}'}$ as

$$s_{\mathbf{m},\mathbf{m}'} = \mathbf{v} \cdot \mathbf{S}^1 = \mathbf{v} \cdot \mathbf{S}^2 = \ldots = \mathbf{v} \cdot \mathbf{S}^n. \tag{4}$$

The equalities $\mathbf{v} \cdot \mathbf{S}^1 = \ldots = \mathbf{v} \cdot \mathbf{S}^n$ correspond to the intuition that it does not matter which of the past n rounds we use to define average payoffs. The payoff $s_{\mathbf{m}',\mathbf{m}}$ of player q can be defined analogously.

Nash Strategies. A strategy $\mathbf{m}$ for player p, is a *Nash strategy*, if the co-player never receives a payoff higher than that of the mutual cooperation payoff. Irrespective of the co-player's strategy. Namely if,

$$s_{\mathbf{m}',\mathbf{m}} \leq (b - c) \; \forall \; m'. \tag{5}$$

Partner Strategies. A player's strategy is *nice*, if the player is never the first to defect. A *partner strategy* for player p is a nice strategy such that,

$$s_{\mathbf{m}',\mathbf{m}} \geq (b - c) \;\Rightarrow\; s_{\mathbf{m}',\mathbf{m}} = s_{\mathbf{m},\mathbf{m}'} = (b - c). \tag{6}$$

irrespective of the co-player's strategy.

Thus, if a player uses a partner strategy, both players can share the rewards fairly. However, if a co-player prefers an unfair approach, they will receive a reduced payoff as a consequence. Partner strategies, by definition, are best responses to themselves, making them Nash equilibria Hilbe et al. [2015].

Previously the work of [Akin, 2016] characterized all partner strategies for $n = 1$, and work by [Hilbe et al., 2017] characterized subsets of partner strategies for higher memory ($n > 1$). In this work, we aim to characterize all partner strategies for $n = 2, n = 3$ for reactive strategies. We formally introduce reactive strategies and present the results from section 3 onwards. In the next section, we will discuss a series of results for the general case of memory$-n$.

2 An Extension of Akin's Theorem

The work of Akin focuses on the case of memory-1 strategies. Thus, $n = 1$, and a memory-1 strategy of player p is the vector $m = (m_1, m_2, m_3, , m_4)$, and against a co-player m' the stationary distribution is of $v = (v_1, v_2, v_3, v_4)$. Akin's lemma is the following:

Theorem 2.1 (Akin's Lemma)**.** Assume that player p uses the memory-one strategy $\mathbf{m} = (m_1, m_2, m_3, m_4)$, 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 $\tilde{\mathbf{m}} = \mathbf{m} - \mathbf{e}_{12}$ where $\mathbf{e}_{12} = (1, 1, 0, 0)$. Then,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbf{v}^{(k)} \cdot \tilde{\mathbf{m}} = 0, \text{ and therefore } \mathbf{v} \cdot \tilde{\mathbf{m}} = 0. \tag{7}$$

One special case of such a memory-n strategy is the *round-k-repeat strategy*. Player p uses a *round-k-repeat strategy* $\mathbf{m}^{k-\text{Rep}}$ if in any given round, the player chooses the same action as k rounds ago. That is, if the game's n-history is such that $a^p_{-k} = C$, then $m^{k-\text{Rep}}_h = 1$; otherwise $m^{k-\text{Rep}}_h = 0$.

With the same method as in [Akin, 2016], one can show *Akin's Lemma*: For each k with $1 \leq k \leq n$, the invariant distribution $\mathbf{v}$ satisfies the following relationship,

$$\mathbf{v} \cdot (\mathbf{m} - \mathbf{m}^{k-\text{Rep}}) = \sum_{h \in H} v_h (m_h - m_h^{k-\text{Rep}}) = 0. \tag{8}$$

The intuition for this result is that $\mathbf{v} \cdot \mathbf{m}$ and all $\mathbf{v} \cdot \mathbf{m}^{k-\text{Rep}}$ are just different (but equivalent) expressions for player p's average cooperation rate. For example, $\mathbf{v} \cdot \mathbf{m}$ corresponds to a setup in which one first draws a history h according to the invariant distribution $\mathbf{v}$; then one takes player p's probability m_h to cooperate in the next round; the expectation of this procedure is $\sum_{h \in H} v_h m_h$.

Zero-determinant strategies. Based on Akin's Lemma, we can derive a theory of zero-determinant strategies analogous to the case of memory-one strategies. In the following, we say a memory-n strategy $\mathbf{m}$ is a zero-determinant strategy if there are k_1, k_2, k_3 and α, β, γ such that $\mathbf{m}$ can be written as

$$\mathbf{m} = \alpha \mathbf{S}^{k_1} + \beta \tilde{\mathbf{S}}^{k_2} + \gamma \mathbf{1} + \mathbf{m}^{k-\text{Rep}}, \tag{9}$$

where $\mathbf{1}$ is the vector for which every entry is 1. By Akin's Lemma and the definition of payoffs,

$$0 = \mathbf{v} \cdot (\mathbf{m} - \mathbf{m}^{k-\text{Rep}}) = \mathbf{v} \cdot (\alpha \mathbf{S}^{k_1} + \beta \tilde{\mathbf{S}}^{k_2} + \gamma \mathbf{1}) = \alpha s_{\mathbf{m},\mathbf{m}'} + \beta s_{\mathbf{m}',\mathbf{m}} + \gamma. \tag{10}$$

That is, payoffs satisfy a linear relationship.

One interesting special case arises if $k_1 = k_2 = k_3 =: k$ and $\alpha = -\beta = 1/(b+c)$ and $\gamma = 0$. In that case, the formula (9) yields the strategy

$$m_h = \begin{cases} 1 & \text{if } \ a_{-k}^q = C \\ 0 & \text{if } \ a_{-k}^q = D \end{cases} \tag{11}$$

That is, this strategy implements Tit-for-Tat (for $k=1$) or delayed versions thereof (for $k>1$). By Eq. (10), the enforced payoff relationship is $s_{\mathbf{p}} = s_{\mathbf{q}}$ (in particular, these strategies are *good*).

Another interesting special case arises if $k_1 = k_2 = k_3 =: k$ and $\alpha = 0$, $\beta = -1/b$, $\gamma = 1 - c/b$. In that case Eq. (9) yields the strategy

$$m_h = \begin{cases} 1 & \text{if } \ a_{-k}^q = C \\ 1 - c/b & \text{if } \ a_{-k}^q = D \end{cases} \tag{12}$$

That is, the generated strategy is GTFT (if $k=1$), or delayed versions thereof (for $k>1$). By Eq. (10), the enforced payoff relationship is $s_{\mathbf{m}',\mathbf{m}} = b - c$. In particular, these strategies are not *partner strategies*, but they satisfy the notion of being *Nash*.

3 Reactive and Partner Strategies

A $n-$ *bit reactive strategy* is a vector $\mathbf{p} = (p_h)_{h \in H^q} \in [0,1]^{2n}$. Each entry p_h corresponds to the player's cooperation probability in the next round, depending on the co-player's action(s) of the previous n rounds.

Thus, n-bit reactive strategies only depend on the co-player's n-history (independent of the focal player's own actions during the past n rounds).

We will differentiate from memory-n strategies by using $\mathbf{p}$ to represent reactive strategies. Studying this subset of strategies leads to a number of interesting results.

By focusing on this specific set of strategies we derive a series of interesting results.

3.1 Sufficiency of self reactive strategies

To characterize all partner n-bit reactive strategies, one would usually need to check against all pure n-memory one strategies McAvoy and Nowak [2019]. However, we demonstrate that when player p employs an n-bit reactive strategy, it is sufficient to check only against $n-$bit self-reactive strategies. This finding aligns with the previous result by Press and Dyson Press and Dyson [2012].

More specifically, the result states that for any memory-n strategy used by player q, player ps' score is exactly the same as if q had played a specific self-reactive memory-n strategy.

A "maybe" example will consider the reactive $\hat{\mathbf{p}} = (0, 1)$ and the memory-1 strategy Pavlov or Win Stay Lose Shift $\mathbf{p} = (1, 0, 0, 1)$.

3.2 $2-$bit partner strategies

For $n = 2$, $\hat{\mathbf{p}} = (\hat{p}_{CC}, \hat{p}_{CD}, \hat{p}_{DC}, \hat{p}_{DD})$, where $\hat{p}_{CC}$ is the probability of cooperating in round t when the co-player cooperated in the last 2 rounds, $\hat{p}_{CD}$ is the probability of cooperating given that the co-player cooperated in the second to last round and defected in the last, and so on. An agreeable 2-bit strategy is represented by the vector $\hat{\mathbf{p}} = (1, \hat{p}_{CD}, \hat{p}_{DC}, \hat{p}_{DD})$:

An agreeable $2-$bit reactive strategy is a partner strategy if the entries of $\hat{\mathbf{p}}$ satisfy:

$$\hat{p}_{DD} < 1 - \frac{c}{b} \quad and \quad \frac{\hat{p}_{CD} + \hat{p}_{DC}}{2} < 1 - \frac{1}{2} \cdot \frac{c}{b}. \tag{13}$$

We have two independent proves of Theorem. The first proves is in line with the work of Akin and the second prove rely on Theorem. Here we present both proves. Suppose player p adopts a 2-bit reactive strategy $\mathbf{p} = (p_{CC}, p_{CD}, p_{DC}, p_{DD})$. Moreover, suppose player q adopts an arbitrary memory-2 strategy. Let $\mathbf{v} = (v_h)_{h \in H}$ be an invariant distribution of the game between the two players. We define the following four marginal distributions with respect to the possible two-histories of player q,

$$\begin{aligned}
v^q_{CC} &= \sum_{h^p \in H^p} v_{(h^p, CC)} \\
v^q_{CD} &= \sum_{h^p \in H^p} v_{(h^p, CD)} \\
v^q_{DC} &= \sum_{h^p \in H^p} v_{(h^p, DC)} \\
v^q_{DD} &= \sum_{h^p \in H^p} v_{(h^p, DD)}.
\end{aligned} \tag{14}$$

These four entries describe how often we observe player q to choose actions CC, CD, DC, DD in two consecutive rounds (irrespective of the actions of player p). Based on the above notation, we can define player q's average cooperation rate $\rho_{\mathbf{q}}$ as

$$\rho_{\mathbf{q}} := v^q_{CC} + v^q_{CD} = v^q_{CC} + v^q_{DC}. \tag{15}$$

Here, the second equality holds because it does not matter whether we define player q's cooperation rate based on the first or the second round of each 2-history. In particular, we can use this equality to conclude

$$v^q_{CD} = v^q_{DC}. \tag{16}$$

Similarly, we can express player p's average cooperation rate $\rho_{\mathbf{p}}$ in terms of v^q_{CC}, v^q_{CD}, v^q_{DC}, v^q_{DD} by noting that

$$\begin{aligned} \rho_{\mathbf{p}} &= v^q_{CC}\, p_{CC} + v^q_{CD}\, p_{CD} + v^q_{DC}\, p_{DC} + v^q_{DD}\, p_{DD} \\ &= v^q_{CC}\, p_{CC} + v^q_{CD}\, (p_{CD}+p_{DC}) + v^q_{DD}\, p_{DD}. \end{aligned} \tag{17}$$

Here, the second equality is due to Eq. (16). Because we consider simple donation games, we note that these two quantities are sufficient to define the payoffs of the two players,

$$\begin{aligned} s_{\mathbf{p}} &= b\,\rho_{\mathbf{q}} - c\,\rho_{\mathbf{p}} \\ s_{\mathbf{q}} &= b\,\rho_{\mathbf{p}} - c\,\rho_{\mathbf{q}}. \end{aligned} \tag{18}$$

Finally, we note that we trivially have the following relationship (since all probabilities need to add up to one),

$$1 = v^q_{CC} + v^q_{CD} + v^q_{DC} + v^q_{DD} = v^q_{CC} + 2v^q_{CD} + v^q_{DD} \tag{19}$$

After these preparations, we can prove our conjecture based on the same method as in Akin [2016].

Proposition 1 ('Main conjecture')**.** Suppose the entries of $\mathbf{p}$ satisfy

$$p_{CC}=1, \qquad \frac{p_{CD}+p_{DC}}{2} < 1-\frac{c}{2b}, \qquad p_{DD} < 1-\frac{c}{b}. \tag{20}$$

Then $\mathbf{p}$ is a good strategy.

Proof. Suppose player q has some strategy $\mathbf{q}$ such that $s_{\mathbf{q}} \geq b-c$. It follows that

$$\begin{aligned} 0 \quad &\leq \quad s_{\mathbf{q}} - (b-c) \\ &\overset{Eq.\ (18)}{=} \quad b\rho_{\mathbf{p}} - c\rho_{\mathbf{q}} - (b-c) \\ &\overset{Eqs.\ (15),(17),(19)}{=} \quad b\Big(v^q_{CC}p_{CC}+v^q_{CD}(p_{CD}+p_{DC})+v^q_{DD}p_{DD}\Big) - c\Big(v^q_{CC}+v^q_{CD}\Big) - (b-c)\Big(v^q_{CC}+2v^q_{CD}+v^q_{DD}\Big) \\ &= \quad v^q_{CC}\, b\,(p_{CC}-1) + v^q_{CD}\Big(b(p_{CD}+p_{DC})+c-2b\Big) + v^q_{DD}\Big(bp_{DD} - (b-c)\Big). \end{aligned} \tag{21}$$

By assumption (20),

$$p_{CC} = 1, \qquad b(p_{CD}+p_{DC})+c-2b < 0, \qquad bp_{DD} - (b-c) < 0. \tag{22}$$

Because any $v^q_{XY} \geq 0$, inequality (21) can only hold if $v^q_{CD} = v^q_{DD} = 0$, which implies $v^q_{DC} = 0$ because of Eq. (16). But then it follows that $v^q_{CC} = 1$. By Eqs. (15) and (17) it follows that $\rho_{\mathbf{q}} = \rho_{\mathbf{p}} = 1$, and hence $s_{\mathbf{q}} = s_{\mathbf{p}} = b-c$. $\square$

3.3 3−bit partner strategies

For $n = 3$, $\hat{\mathbf{p}} = (\hat{p}_{CCC}, \hat{p}_{CCD}, \hat{p}_{CDC}, \hat{p}_{CDD}, \hat{p}_{DCC}, \hat{p}_{DCD}, \hat{p}_{DDC}, \hat{p}_{DDD})$ where $\hat{p}_{CCC}$ is the probability of cooperating in round t when the co-player cooperates in the last 3 rounds, $\hat{p}_{CCD}$ is the probability of cooperating given that the co-player cooperated in the third and second to last rounds and defected in the last, etc. An agreeable 3−bit strategy is of the vector $\hat{\mathbf{p}} = (1, \hat{p}_{CCD}, \hat{p}_{CDC}, \hat{p}_{CDD}, \hat{p}_{DCC}, \hat{p}_{DCD}, \hat{p}_{DDC}, \hat{p}_{DDD})$.

An agreeable 3−bit reactive strategy is a partner strategy if the entries of $\hat{\mathbf{p}}$ satisfy:

$$\frac{\hat{p}_{CCD} + \hat{p}_{CDC} + \hat{p}_{DCC}}{3} < 1 - \frac{1}{3} \cdot \frac{c}{b} \qquad \frac{\hat{p}_{CDD} + \hat{p}_{DCD} + \hat{p}_{DDC}}{3} < 1 - \frac{2}{3} \cdot \frac{c}{b} \qquad \hat{p}_{DDD} < 1 - \frac{c}{b} \tag{23}$$

$$\frac{\hat{p}_{CCD} + \hat{p}_{CDD} + \hat{p}_{DCC} + \hat{p}_{DDC}}{4} < 1 - \frac{1}{2} \cdot \frac{c}{b} \qquad \frac{\hat{p}_{CDC} + \hat{p}_{DCD}}{2} < 1 - \frac{1}{2} \cdot \frac{c}{b} \tag{24}$$

We have two independent proves of Theorem. The first proves is in line with the work of Akin and the second prove rely on Theorem. Here we present both proves.

3.4 $n-$bit counting partner strategies

A special case of reactive strategies are counting reactive strategies.

A special case of 2−bit reactive strategies is the 2−*bit counting reactive* strategies. These are strategies that respond to the action of the co-player, but they do not differentiate between when defection occurs, only if one or two defections occurred. Let r_i be the probability of cooperating given that the co-player cooperated i number of times in the last 2 turns.

Thus, $r_2 = \hat{p}_1, r_1 = \hat{p}_2 = \hat{p}_3, r_0 = \hat{p}_4$ and $\hat{\mathbf{p}} = (r_2 = 1, r_1, r_0)$. Conditions (13) then become:

$$r_1 < 1 - \frac{1}{2} \cdot \frac{c}{b} \;\; \textit{and} \;\; r_0 < 1 - \frac{c}{b}. \tag{25}$$

A special case of 3−bit reactive strategies are the 3−bit counting reactive strategies. Let r_i be the probability of cooperating given that the co-player cooperated i number of times in the last 3 turns. So, $r_3 = \hat{p}_{CCC}, r_2 = \hat{p}_{CCD} = \hat{p}_{CDC} = \hat{p}_{DCC}, r_1 = \hat{p}_{CDD} = \hat{p}_{DCD} = \hat{p}_{DDC}, r_0 = \hat{p}_{CCC}$ and $\hat{\mathbf{p}} = (r_3 = 1, r_2, r_1, r_0)$. Then, conditions (27), the conditions for being a partner strategy become:

$$r_2 < 1 - \frac{1}{3} \cdot \frac{c}{b}, \quad r_1 < 1 - \frac{2}{3} \cdot \frac{c}{b} \;\; \textit{and} \;\; r_0 < 1 - \frac{c}{b}. \tag{26}$$

In the case of counting reactive strategies, we observe a pattern in the conditions they must satisfy to be partner strategies. We show that for an $n-$bit counting reactive strategy to be a partner strategy, the strategy's entries must satisfy the conditions:

$$
\begin{aligned}
r_n =& 1 \\
r_{n-1} \leq& 1 - \frac{(n-1)}{n} \times \frac{c}{b} \\
r_{n-2} \leq& 1 - \frac{(n-2)}{n} \times \frac{c}{b} \\
&\vdots \\
r_0 \leq& 1 - \frac{c}{b}
\end{aligned}
$$

References

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

C. Hilbe, A. Traulsen, and K. Sigmund. Partners or rivals? strategies for the iterated prisoner's dilemma. *Games and economic behavior*, 92:41–52, 2015.

C. Hilbe, L. A. Martinez-Vaquero, K. Chatterjee, and M. A. Nowak. Memory-n strategies of direct reciprocity. *Proceedings of the National Academy of Sciences*, 114(18):4715–4720, 2017.

A. McAvoy and M. A. Nowak. Reactive learning strategies for iterated games. *Proceedings of the Royal Society A*, 475(2223):20180819, 2019.

W. H. Press and F. J. Dyson. Iterated prisoner's dilemma contains strategies that dominate any evolutionary opponent. *Proceedings of the National Academy of Sciences*, 109(26):10409–10413, 2012.