Supporting Information

Reactive strategies with longer memory

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

1 Model and basic results

1.1 The repeated prisoner's dilemma

We consider the infinitely repeated prisoner's dilemma between two players, player 1 and player 2. Each round, each player can either cooperate (C) or defect (D). The resulting payoffs are given by the matrix

$$\begin{array}{c c} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} R & S \\ T & P \end{pmatrix} \end{array}. \tag{1}$$

Here, R is the reward payoff of mutual cooperation, T is the temptation to defect, S is the sucker's payoff, and P is the punishment payoff for mutual defection. For the game to be a prisoner's dilemma, we require

$$T > R > S > P \quad \text{and} \quad 2R > T+S. \tag{2}$$

That is, mutual cooperation is the best outcome to maximize the players' total payoffs, but each player's dominant action is to defect. For some of our results, we focus on a special case of the prisoner's dilemma, the donation game. This game only depends on two free parameters, the benefit b and the cost c of cooperation. The payoff matrix of the donation game takes the form

$$\begin{array}{c c} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} b-c & -c \\ b & 0 \end{pmatrix} \end{array}. \tag{3}$$

For this game to satisfy the conditions (2) of a prisoner's dilemma, we assume $b > c > 0$ throughout.

Players interact in the repeated prisoner's dilemma for infinitely many rounds, and future payoffs are not discounted. A strategy σ^i for player i is a rule that tells the player what to in any given round, depending on the outcome of all previous rounds. Given the player's strategies σ^1 and σ^2, one can compute each player i's expected payoff $\pi^i_{\sigma^1,\sigma^2}(t)$ in round t. For the entire repeated game, we define the players' payoffs as the expected payoff per round,

$$\pi^i(\sigma^1, \sigma^2) = \lim_{\tau\to\infty} \frac{1}{\tau} \sum_{t=1}^{\tau} \pi^i_{\sigma^1,\sigma^2}(t). \tag{4}$$

For general strategies σ^1 and σ^2, the above limit may not always exist. Problems may arise, for example, if one of the players cooperates in the first round, defects in the two subsequent rounds, cooperates in the four rounds thereafter, etc., which prevents the time averages from converging. However, in the following, we focus on strategies with finite memory. When both players adopt such a strategy, the existence of the limit (4) is guaranteed, as we discuss further below.

1.2 Finite-memory strategies

In this study, we focus on strategies that ignore all events that happened more than n rounds ago. To define these strategies, we need some notation. An *n-history for player i* is a string $\mathbf{h^i} = (a^i_{-n}, \dots, a^i_{-1}) \in \{C, D\}^n$. We interpret the string's entry a^i_{-k} as player i's action k rounds ago. We denote the space of all n-histories for player i as H^i. This space contains $|H^i| = 2^n$ elements. A pair $\mathbf{h} = (\mathbf{h^1}, \mathbf{h^2})$ is called an *n-history of the game*. We use $H = H^1 \times H^2$ to denote the space of all such histories, which contains $|H| = 2^{2n}$ elements.

Memory-n strategies. Based on this notation, a *memory-n strategy* for player i as a tuple $\mathbf{m} = (m_{\mathbf{h}})_{\mathbf{h}\in H}$. Each input $\mathbf{h} = (h^i, h^{-i})$ refers to a possible n-history, where now $\mathbf{h^i}$ and $\mathbf{h^{-i}}$ refer to the n-histories of the focal player and the co-player, respectively. The corresponding output $m_{\mathbf{h}} \in [0, 1]$ is the focal player's cooperation probability in the next round, contingent on the outcome of the previous n rounds. We refer to the set of all memory-n strategies as

$$\mathcal{M}_n := \Big\{\mathbf{m} = (m_{\mathbf{h}})_{\mathbf{h}\in H} \;\Big|\; 0 \le m_{\mathbf{h}} \le 1 \text{ for all } \mathbf{h} \in H\Big\} = [0, 1]^{2^{2n}}. \tag{5}$$

This definition leaves the strategy's actions during the first n rounds unspecified, for which no complete n-history is yet available. However, because we consider infinitely repeated games without discounting, these first n rounds are usually irrelevant for the long-run dynamics, as we show further below. In the following, we therefore only specify a strategy's move during the first n rounds when necessary.

Among all memory-n spaces $\mathcal{M}_n$, the space with $n = 1$ is the one most frequently studied. Memory-1 strategies take the form $\mathbf{m} = (m_{CC}, m_{CD}, m_{DC}, m_{DD})$, where the first index refers to the focal player's last action (1-history) and the second index refers to the co-player's last action. As one example of a well-known memory-1 strategy, we mention Win-Stay Lose-Shift [1], $\mathbf{m} = (1, 0, 0, 1)$. However, there are many others [2].

Reactive-n strategies. For our following analysis, two particular subsets of memory-n strategies will play an important role. The first subset is the set of *reactive-n strategies*,

$$\mathcal{R}_n := \Big\{\mathbf{m} \in \mathcal{M}_n \;\Big|\; m_{(\mathbf{h^i},\mathbf{h^{-i}})} = m_{(\tilde{\mathbf{h}}^{\mathbf{i}},\mathbf{h^{-i}})} \text{ for all } \mathbf{h^i}, \tilde{\mathbf{h}}^{\mathbf{i}} \in H^i \text{ and } \mathbf{h^{-i}} \in H^{-i}\Big\}. \tag{6}$$

That is, reactive-n strategies are independent of the focal player's own n-history. The space of reactive-n strategies can be naturally identified with the space of all 2^n-dimensional vectors

$$\mathbf{p}=(p_{\mathbf{h^{-i}}})_{\mathbf{h^{-i}}\in H^{-i}} \text{ with } 0\le p_{\mathbf{h^{-i}}}\le 1 \text{ for all } \mathbf{h^{-i}}\in H^{-i}. \tag{7}$$

In this reduced representation, each entry $p_{\mathbf{h^{-i}}}$ corresponds to the player's cooperation probability in the next round based on the co-player's actions in the previous n rounds. Again, the most studied case of reactive-n strategies is when $n=1$. Here, the reduced representation according to Eq. (7) takes the form $\mathbf{p}=(p_C, p_D)$. The probably best-known example of a reactive-1 strategy is Tit-for-Tat, `TFT` [3]. `TFT` cooperates if and only if the co-player cooperated in the previous round. Hence, its memory-1 representation is $\mathbf{m}=(1,0,1,0)$, whereas its reduced representation is $\mathbf{p}=(1,0)$. Another example is the strategy Generous Tit-for-Tat, `GTFT` [4, 5]. `GTFT` occasionally cooperates even if the co-player defected. For that strategy, the memory-1 representation is $\mathbf{m}=(1,p_D^*,1,p_D^*)$, and the reduced representation is $\mathbf{p}=(1,p_D^*)$, where

$$p_D^* := \min\big\{1-(T-R)/(R-S),\ (R-P)/(T-P)\big\}. \tag{8}$$

In the special case that payoffs are given by the donation game, this condition simplifies to $p_D^*=1-c/b$.

Self-reactive-n strategies. The other important subspace of memory-n strategies is the set of self-reactive-n strategies,

$$\mathcal{S}_n := \Big\{\mathbf{m}\in\mathcal{M}_n \;\Big|\; m_{(\mathbf{h^i},\mathbf{h^{-i}})}=m_{(\mathbf{h^i},\tilde{\mathbf{h}}^{-\mathbf{i}})} \text{ for all } \mathbf{h^i}\in H^i \text{ and } \mathbf{h^{-i}}, \tilde{\mathbf{h}}^{-\mathbf{i}}\in H^{-i}\Big\}. \tag{9}$$

These strategies only depend on the focal player's own decisions during the last n rounds, independent of the co-player's decisions. Again, we can identify any self-reactive-n strategies with a 2^n-dimensional vector,

$$\tilde{\mathbf{p}} = (\tilde{p}_{\mathbf{h^i}})_{\mathbf{h^i}\in H^i} \text{ with } 0\le\tilde{p}_{\mathbf{h^i}}\le 1 \text{ for all } \mathbf{h^i}\in H^i. \tag{10}$$

Each entry $\tilde{p}_{\mathbf{h^i}}$ corresponds to the player's cooperation probability in the next round, contingent on the player's own actions in the previous n rounds. A special subset of self-reactive strategies is given by the round-k-repeat strategies, for some $1\le k\le n$. In any given round, players with a *round-k-repeat strategy* $\tilde{\mathbf{p}}^{k-\text{Rep}}$ choose the same action as they did k rounds ago. Formally, the entries of $\tilde{\mathbf{p}}^{k-\text{Rep}}$ are defined by

$$p_{\mathbf{h^i}}^{k-\text{Rep}} = \begin{cases} 1 & \text{if } a_{-k}^i = C \\ 0 & \text{if } a_{-k}^i = D. \end{cases}$$

From this point forward, we will use the notations $\mathbf{m}$, $\mathbf{p}$, and $\tilde{\mathbf{p}}$ to denote memory-n, reactive-n, and self-reactive-n strategies, respectively. We say these strategies are *pure* or *deterministic* if all conditional cooperation probabilities are either zero or one. If all cooperation probabilities are strictly between zero and one, we say the strategy is *strictly stochastic*. When it is convenient to represent the self-reactive repeat strategies as elements of the memory-n strategy space, we write $\mathbf{m}^{k-\text{Rep}}\in[0,1]^{2^{2n}}$ instead of $\tilde{\mathbf{p}}^{k-\text{Rep}}\in[0,1]^{2^n}$.

1.3 Computing the payoffs of finite-memory strategies

A Markov chain representation. The interaction between two players with memory-n strategies $\mathbf{m^1}$ and $\mathbf{m^2}$ can be represented as a Markov chain. The states of the Markov chain are the possible n-histories $\mathbf{h}\in H$. To compute the transition probabilities from one state to another within a single round, suppose players currently have the n-history $\mathbf{h}=(\mathbf{h^1},\mathbf{h^2})$ in memory. Then the transition probability that the state after one round is $\tilde{\mathbf{h}}=(\tilde{\mathbf{h}}^\mathbf{1},\tilde{\mathbf{h}}^\mathbf{2})$ is a product of two factors,

$$M_{\mathbf{h},\tilde{\mathbf{h}}} = x^1 \cdot x^2, \tag{11}$$

The two factors represent the (independent) decisions of the two players,

$$x^i = \begin{cases} m^i_{(\mathbf{h^i},\mathbf{h^{-i}})} & \text{if } \tilde{a}^i_{-1}=C, \quad \text{and} \quad \tilde{a}^i_{-t}=a^i_{-t+1} \quad \text{for } \; t\in\{2,\dots,n\} \\ 1-m^i_{(\mathbf{h^i},\mathbf{h^{-i}})} & \text{if } \tilde{a}^i_{-1}=D, \quad \text{and} \quad \tilde{a}^i_{-t}=a^i_{-t+1} \quad \text{for } \; t\in\{2,\dots,n\} \\ 0 & \text{if } \tilde{a}^i_{-t}\neq a^i_{-t+1} \quad \text{for some } \; t\in\{2,\dots,n\}. \end{cases} \tag{12}$$

The resulting $2^{2n}\times 2^{2n}$ transition matrix $M=(M_{\mathbf{h},\tilde{\mathbf{h}}})$ fully describes the dynamics among the two players after the first n rounds. More specifically, suppose $\mathbf{v}(t)=\big(v_\mathbf{h}(t)\big)_{\mathbf{h}\in H}$ is the probability distribution of observing state $\mathbf{h}$ after players made their decisions for round $t\geq n$. Then the respective probability distribution after round $t+1$ is given by $\mathbf{v}(t+1)=\mathbf{v}(t)\cdot M$. The long-run dynamics is particularly simple to describe when the matrix M is primitive (which happens, for example, when the two strategies are $m^i_\mathbf{h}$ are strictly stochastic). In that case, it follows by the theorem of Perron and Frobenius that $\mathbf{v}(t)$ converges to some $\mathbf{v}$ as $t\to\infty$. As a result, also the respective time average exists and converges,

$$\mathbf{v} = \lim_{\tau\to\infty}\frac{1}{\tau}\sum_{t=n}^{n+\tau-1}\mathbf{v}(t). \tag{13}$$

This limiting distribution $\mathbf{v}$ can be computed as the unique solution of the system $\mathbf{v}=\mathbf{v}M$, with the additional constraint that the entries of $\mathbf{v}$ need to sum up to one.

But even when M is not ergodic, $\mathbf{v}(t)$ still converges to an invariant distribution $\mathbf{v}$ that satisfies $\mathbf{v}=\mathbf{v}M$. However, in that case, the system $\mathbf{v}=\mathbf{v}M$ no longer has a unique solution. Instead, the limiting distribution $\mathbf{v}$ depends on the very first n-history after the first n rounds, $\mathbf{v}(n)$, which in turn depends on the players' moves during the first n rounds.

A formula for the payoffs among memory-n players. Based on the above considerations, we can derive an explicit formula for the payoffs according to Eq. (4) when players use memory-n strategies $\mathbf{m^1}$ and $\mathbf{m^2}$. To this end, we introduce a 2^{2n}-dimensional vector $\mathbf{g^i}(k)=\big(g^i_\mathbf{h}(k)\big)_{\mathbf{h}\in H}$, that takes an n-history $\mathbf{h}$ as an input and returns player i's payoff k rounds ago, for $k \leq n$. That is,

$$g^i_{\mathbf{h}}(k) = \begin{cases} R & \text{if } a^i_{-k} = C \text{ and } a^{-i}_{-k} = C \\ S & \text{if } a^i_{-k} = C \text{ and } a^{-i}_{-k} = D \\ T & \text{if } a^i_{-k} = D \text{ and } a^{-i}_{-k} = C \\ P & \text{if } a^i_{-k} = D \text{ and } a^{-i}_{-k} = D. \end{cases} \tag{14}$$

Now for a given $t \geq n$, given that $\mathbf{v}(t)$ captures the state of the system after round t, we can write player i's expected payoff in that round as

$$\pi^i_{\mathbf{m^1},\mathbf{m^2}}(t) = \big\langle\, \mathbf{v}(t),\, \mathbf{g^i}(1) \,\big\rangle = \sum_{\mathbf{h} \in H} v_{\mathbf{h}}(t) \cdot g^i_{\mathbf{h}}(1). \tag{15}$$

As a result, we obtain for the player's average payoff across all rounds

$$\begin{aligned} \pi^i(\mathbf{m^1}, \mathbf{m^2}) &\overset{(4)}{=} \lim_{\tau \to \infty} \frac{1}{\tau} \sum_{t=1}^{\tau} \pi^i_{\mathbf{m^1},\mathbf{m^2}}(t) \;=\; \lim_{\tau \to \infty} \frac{1}{\tau} \sum_{t=n}^{n+\tau-1} \pi^i_{\mathbf{m^1},\mathbf{m^2}}(t) \\ &\overset{(15)}{=} \lim_{\tau \to \infty} \frac{1}{\tau} \sum_{t=n}^{n+\tau-1} \big\langle\, \mathbf{v}(t),\, \mathbf{g^i}(1) \,\big\rangle \;=\; \Big\langle \lim_{\tau \to \infty} \frac{1}{\tau} \sum_{t=n}^{n+\tau-1} \mathbf{v}(t)\,,\, \mathbf{g^i}(t) \Big\rangle \\ &\overset{(13)}{=} \big\langle\, \mathbf{v}, \mathbf{g^i}(1) \,\big\rangle. \end{aligned} \tag{16}$$

That is, given we know the invariant distribution $\mathbf{v}$ that captures the game's long-run dynamics, it is straightforward to compute payoffs by taking the scalar product with the vector $\mathbf{g^i}(1)$. With a similar approach as in Eq. (16), one can also show

$$\big\langle\, \mathbf{v}, \mathbf{g^i}(1) \,\big\rangle \;=\; \big\langle\, \mathbf{v}, \mathbf{g^i}(2) \,\big\rangle \;=\; \ldots \;=\; \big\langle\, \mathbf{v}, \mathbf{g^i}(n) \,\big\rangle. \tag{17}$$

That is, to compute player i's expected payoff, it does not matter whether one refers to the last round of an n-history or to an earlier round of an n-history. All rounds k with $1 \leq k \leq n$ are equivalent.

1.4 An Extension of Akin's Lemma

The above Markov chain approach allows us to analyze games when both players adopt memory-n strategies. But even if only one player adopts a memory-n strategy (and the other player's strategy is arbitrary), one can still derive certain constraints on the game's long-run dynamics. One such constraint was first described by Akin [6]: if player 1 adopts a memory-1 strategy $\mathbf{m}$ against an arbitrary opponent, and if the time average $\mathbf{v}$ defined by the right hand side of Eq. (13) exists, then

$$\big\langle\, \mathbf{v}\,, \mathbf{m} - \mathbf{m}^{1-\text{Rep}} \,\big\rangle = 0. \tag{18}$$

That is, the limiting distribution $\mathbf{v}$ needs to be orthogonal to the vector $\mathbf{m} - \mathbf{m}^{1-\text{Rep}}$. This result has been termed *Akin's Lemma* [7]. With similar methods as in Ref. [6], one can generalize this result to the context of memory-n strategies.

Lemma 1 (A generalized version of Akin's Lemma)
Let player 1 *use a memory-*n *strategy, and let player* 2 *use an arbitrary strategy. For the resulting game and all* $t \geq n$*, let* $\mathbf{v}(t) = \big(v_{\mathbf{h}}(t)\big)_{\mathbf{h}\in H}$ *denote the probability distribution of observing each possible* n*-history* $\mathbf{h} \in H$ *after players made their decisions for round* t*. Moreover, suppose the respective time average* $\mathbf{v}$ *according to Eq.* (13) *exists. Then for each* k *with* $1 \leq k \leq n$*, we obtain*

$$\big\langle\, \mathbf{v}\,, \mathbf{m} - \mathbf{m}^{k-Rep} \,\big\rangle = 0. \tag{19}$$

All proofs are presented in the Appendix. Here we provide an intuition. The expression $\langle \mathbf{v}, \mathbf{m} \rangle = \sum_{\mathbf{h}} v_{\mathbf{h}} m_{\mathbf{h}}$ can be interpreted as player 1's average cooperation rate across all rounds of the repeated game. To compute that average cooperation rate, one first draws an n-history $\mathbf{h}$ (with probability $v_{\mathbf{h}}$), and then one computes how likely player 1 would cooperate in the subsequent round (with probability $m_{\mathbf{h}}$). Alternatively, one could compute the average cooperation rate by drawing an n-history $\mathbf{h}$ and then checking how likely player 1 was to cooperate k rounds ago, according that n-history. That second interpretation leads to the expression $\langle \mathbf{v}, \mathbf{m}^{k-\text{Rep}} \rangle$. According to Eq. (19), both interpretations are equivalent.

2 Characterizing the partner strategies among the reactive-n strategies

2.1 Partner strategies

In this study, we are interested in identifying strategies that can sustain full cooperation in a Nash equilibrium. Strategies with these properties have been termed as being of *Nash type* by Akin [6], or as *partner strategies* by Hilbe *et al* [8]. In the following, we formally define them.

Definition (Partner strategies)

(i) A strategy σ for the repeated prisoner's dilemma is a *Nash equilibrium* if it is a best response to itself. That is, we require $\pi^1(\sigma, \sigma)$ to exist and

$$\pi^1(\sigma,\sigma) \geq \pi^1(\sigma',\sigma) \;\; \text{for all other strategies } \sigma' \text{ for which } \pi^1(\sigma',\sigma) \text{ exists.} \tag{20}$$

(ii) A player's strategy is *nice*, if the player is never the first to defect.

(iii) A *partner strategy* is a strategy that is both nice and a Nash equilibrium.

Several remarks are in order. First, we note that when two players with nice strategies interact, they both cooperate in every round. Partner strategies thus sustain mutual cooperation in a Nash equilibrium.

Second, if a memory-n strategy $\mathbf{m} = (m_{\mathbf{h}})_{\mathbf{h}\in H}$ is to be nice, it needs to cooperate after n rounds of mutual cooperation. In other words, if $\mathbf{h_C} = (\mathbf{h^i_C}, \mathbf{h^{-i}_C})$ is the n-history that consists of mutual cooperation for the past n rounds, then the strategy needs to respond by cooperating with certainty, $m_{\mathbf{h_C}}=1$. Similarly, a nice reactive-n strategy needs to satisfy $p_{\mathbf{h_C^{-i}}}=1$.

Third, in general it is a difficult task to verify that any given strategy σ is a Nash equilibrium. After all, one needs to verify that it yields the highest payoff according to Eq. (20) among all (uncountably) many alternative strategies σ'. Fortunately, the situation is somewhat simpler if the strategy under consideration is a memory-n strategy. In that case, it follows from an argument by Press and Dyson [9] that one only needs to compare the strategy to all other memory-n strategies. However, this still leaves us with uncountably many strategies to check. In fact, it is one aim of this paper to show that for reactive-n strategies, it suffices to check finitely many alternative strategies.

2.2 Tit For Tat and Generous Tit For Tat with arbitrary memory lengths

Zero-determinant strategies with n rounds memory. Before we provide a general algorithm to identify reactive-n partner strategies, we first generalize some of the well-known reactive-1 partner strategies, `TFT` and `GTFT`, to the case of memory-n. To this end, we use Lemma 1 to develop a theory of zero-determinant strategies within the class of memory-n strategies, see also Refs. [11, 12]. In the following, we say a memory-n strategy $\mathbf{m}$ is a *zero-determinant strategy* if there are integers $k_1, k_2, k_3 \leq n$ and real numbers α, β, γ such that $\mathbf{m^i}$ can be written as

$$\mathbf{m^i} = \alpha\, \mathbf{g}^i(k_1) + \beta\, \mathbf{g}^{-i}(k_2) + \gamma\, \mathbf{1} + \mathbf{m}^{k_3-\text{Rep}}. \tag{21}$$

In this expression, $\mathbf{g}^i(k)$ is the vector that returns player i's payoff k rounds ago, as defined by Eq. (14), $\mathbf{m}^{k-\text{Rep}}$ is the memory-n strategy that repeats player i's own move k rounds ago, and $\mathbf{1}$ is the 2^{2n}-dimensional vector for which every entry is one. Using the generalized version of Akin's Lemma, we obtain

$$\begin{aligned}
0 \quad &\overset{(19)}{=} \quad \left\langle \mathbf{v}, \mathbf{m} - \mathbf{m}^{k_3-\text{Rep}} \right\rangle \\
&\overset{(21)}{=} \quad \left\langle \mathbf{v}, \alpha \mathbf{g}^i(k_1) + \beta \mathbf{g}^{-i}(k_2) + \gamma \mathbf{1} \right\rangle \\
&= \quad \alpha \left\langle \mathbf{v}, \mathbf{g}^i(k_1) \right\rangle + \beta \langle \mathbf{v}, \mathbf{g}^{-i}(k_2) \rangle + \gamma \left\langle \mathbf{v}, \mathbf{1} \right\rangle \\
&\overset{(16),(17)}{=} \quad \alpha\, \pi^i(\mathbf{m^i}, \sigma^{-i}) + \beta\, \pi^{-i}(\mathbf{m^i}, \sigma^{-i}) + \gamma.
\end{aligned} \tag{22}$$

That is, a player with a zero-determinant strategy enforces a linear relationship between the player's payoffs, irrespective of the co-player's strategy. Remarkably, the parameters α, β, and γ of that linear relationship are entirely under player i's control.

Generalized versions of Tit-for-tat. One interesting special case arises if $k_1 = k_2 = k_3 =: k$ and $\alpha = -\beta = 1/(T - S)$, $\gamma = 0$. In that case, formula (21) yields the strategy with entries

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

Therefore, a player with that strategy cooperates if and only if the co-player cooperated k rounds ago. Thus, the strategy implements TFT (for $k=1$) or delayed versions thereof (for $k>1$). By Eq. (22), the strategy enforces equal payoffs against any co-player,

$$\pi^i(\mathbf{m^i}, \sigma^{-i}) = \pi^{-i}(\mathbf{m^i}, \sigma^{-i}). \tag{23}$$

Moreover, this strategy is nice if we additionally require it to unconditionally cooperate during the first k rounds. Given this additional requirement, the payoff of $\mathbf{m^i}$ against itself is R. Moreover, the strategy is a Nash equilibrium. To see why, suppose to the contrary that there is a strategy σ^{-i} with $\pi^{-i}(\mathbf{m}^i, \sigma^{-i}) > R$. Then it follows from (23) that $\pi^i(\mathbf{m^i}, \sigma^{-i}) + \pi^{-i}(\mathbf{m^i}, \sigma^{-i}) > 2R$. That is, the total payoff per round exceeds $2R$, which is incompatible with the basic assumptions on a prisoner's dilemma, Eq. (2). We conclude that all these versions of TFT are nice and they are Nash equilibria. Hence, they are partner strategies.

Generalized versions of Generous Tit-for-Tat. Another interesting special case arises in the donation game if $k_1 = k_2 = k_3 =: k$ and $\alpha = 0$, $\beta = -1/b$, $\gamma = 1 - c/b$. In that case Eq. (21) yields the strategy with entries

$$m_{\mathbf{h}} = \begin{cases} 1 & \text{if } a_{-k}^{-i} = C \\ 1 - c/b & \text{if } a_{-k}^{-i} = D \end{cases}$$

That is, the generated strategy is GTFT (if $k=1$), or a delayed version thereof (for $k>1$). By Eq. (22), the enforced payoff relationship is $\pi^{-i}(\mathbf{m^i}, \sigma^{-i}) = b - c$. That is, the co-player always obtains the mutual cooperation payoff, irrespective of the co-player's strategy. In particular, all these versions of GTFT are nice and they are Nash equilibria. Hence, they are partner strategies.

2.3 An algorithm to check whether a reactive-n strategy is a Nash equilibrium

Sufficiency of pure self-reactive strategies. After discussing these particular cases, we would like to derive a general algorithm that allows us to verify whether a given reactive-n strategy is a Nash equilibrium. In principle, this requires us to check the payoff of any other strategy (including strategies that have a much longer memory length than n). Fortunately, however, some simplifications are possible when we use an insight by Press and Dyson [9]. They discussed the case where one player uses a memory-1 strategy and the other player employs a longer memory strategy. They demonstrated that the payoff of the player with the longer memory is exactly the same as if the player had employed a specific shorter-memory strategy, disregarding any history beyond what is shared with the short-memory player. Here we show a result that follows a similar intuition. If there is a part of the game's history that one player does not take into account, then the co-player gains nothing by considering that part of the history.

Lemma 2 (Against reactive strategies, any feasible payoff can be generated with self-reactive strategies)
Let $\mathbf{p} \in \mathcal{R}_n$ be a reactive strategy for player 1. Moreover, suppose player 2 adopts some strategy σ such that for the resulting game, the time average $\mathbf{v}$ according to Eq. (13) exists. Then there is a self-reactive-n strategy $\tilde{\mathbf{p}} \in \mathcal{S}_n$ such that $\pi^i(\mathbf{p}, \sigma) = \pi^i(\mathbf{p}, \tilde{\mathbf{p}})$ for $i \in \{1, 2\}$.

If we are to verify that some given reactive-$\mathbf{n}$ strategy $\mathbf{p}$ is a Nash equilibrium, Lemma 2 simplifies our task considerably. Instead of checking condition (20) for all possible strategies σ', we only need to check it for all self-reactive strategies $\tilde{\mathbf{p}} \in \mathcal{S}_n$. Any payoff that could be obtained with some general strategy can already be obtained with a self-reactive strategy. The following result simplifies our task even further.

Theorem 1 (To any reactive strategy, there is a best response among the pure self-reactive strategies)
For any reactive strategy $\mathbf{p} \in \mathcal{R}_n$ *there is some pure self-reactive strategy* $\tilde{\mathbf{p}} \in \mathcal{S}_n$ *such that*

$$\pi^1(\tilde{\mathbf{p}}, \mathbf{p}) \geq \pi^1(\sigma', \mathbf{p}) \;\; \textit{for all other strategies } \sigma' \textit{ for which } \pi^1(\sigma', \mathbf{p}) \textit{ exists.} \tag{24}$$

This result implies that we only need to check finitely many other strategies if we are to verify that some given reactive-n strategy is a Nash equilibrium.

Corollary 1 (An algorithm to check whether a reactive-n strategy is a Nash equilibrum)
A reactive strategy $\mathbf{p} \in \mathcal{R}_n$ *is a Nash equilibrium if and only if* $\pi^1(\mathbf{p}, \mathbf{p}) \geq \pi^1(\tilde{\mathbf{p}}, \mathbf{p})$ *for all pure self-reactive strategies* $\tilde{\mathbf{p}} \in \mathcal{S}_n$.

In particular, to verify that some given reactive-n strategy is a Nash equilibrium, one needs to check at most 2^n possible deviations.

A more efficient way to calculate payoffs. For the remainder of this section, we thus assume that player 1 uses a self-reactive-n strategy $\tilde{\mathbf{p}} = (\tilde{p}_{\mathbf{h}^{\mathbf{i}}})_{\mathbf{h}^{\mathbf{i}} \in H^i}$, whereas player 2 uses a self-reactive-n strategy $\mathbf{p} = (p_{\mathbf{h}^{-\mathbf{i}}})_{\mathbf{h}^{-\mathbf{i}} \in H^{-i}}$. Our algorithm to compute payoffs for the two players in Section 1.3 would require us to interpret the two strategies as memory-n strategies. We would thus compute a left eigenvector of a $2^{2n} \times 2^{2n}$ transition matrix. in the following, however, we show that for games between reactive and self-reactive players, it suffices to consider a $2^n \times 2^n$ transition matrix. This efficiency gain is possible because both players only consider player 1's past actions. Instead of taking the space of all of the game's n-histories $H = H^1 \times H^2$ as the state space, we can thus take the space H^1. Let $h^1 = (a^1_{-n}, \ldots, a^1_{-1})$ be the state in the current round. Then we obtain the following probability that state after one round is $\tilde{h}^1 = (\tilde{a}^1_{-n}, \ldots, \tilde{a}^1_{-1})$,

$$\tilde{M}_{\mathbf{h}^1, \tilde{\mathbf{h}}^1} = \begin{cases} \tilde{p}_{\mathbf{h}^1} & \text{if } \tilde{a}^1_{-1} = C, \;\text{ and }\; \tilde{a}^1_{-t} = a^1_{-t+1} \;\text{ for all } t \in \{2, \ldots, n\} \\ 1 - \tilde{p}_{\mathbf{h}^1} & \text{if } \tilde{a}^1_{-1} = D, \;\text{ and } \tilde{a}^1_{-t} = a^1_{-t+1} \;\text{ for all } t \in \{2, \ldots, n\} \\ 0 & \text{if } \tilde{a}^1_{-t} \neq a^1_{-t+1} \;\text{ for some } t \in \{2, \ldots, n\}. \end{cases} \tag{25}$$

Let $\tilde{\mathbf{v}} = (\tilde{v}_{\mathbf{h}^1})_{\mathbf{h}^1 \in H^1}$ be the limiting distribution of the resulting game (which only in exceptional cases depends on player 1's behavior during the first n rounds). Then the players' payoffs are given by

$$\begin{aligned} \pi^1(\tilde{\mathbf{p}}, \mathbf{p}) &= \sum_{\mathbf{h}^1 \in H^1} \tilde{v}_{\mathbf{h}^1} \Big(\tilde{\mathbf{p}}_{\mathbf{h}^1} \mathbf{p}_{\mathbf{h}^1} \cdot R + \tilde{\mathbf{p}}_{\mathbf{h}^1} (1 - \mathbf{p}_{\mathbf{h}^1}) \cdot S + (1 - \tilde{\mathbf{p}}_{\mathbf{h}^1}) \mathbf{p}_{\mathbf{h}^1} \cdot T + (1 - \tilde{\mathbf{p}}_{\mathbf{h}^1})(1 - \mathbf{p}_{\mathbf{h}^1}) \cdot P \Big), \\ \pi^2(\tilde{\mathbf{p}}, \mathbf{p}) &= \sum_{\mathbf{h}^1 \in H^1} \tilde{v}_{\mathbf{h}^1} \Big(\tilde{\mathbf{p}}_{\mathbf{h}^1} \mathbf{p}_{\mathbf{h}^1} \cdot R + \tilde{\mathbf{p}}_{\mathbf{h}^1} (1 - \mathbf{p}_{\mathbf{h}^1}) \cdot T + (1 - \tilde{\mathbf{p}}_{\mathbf{h}^1}) \mathbf{p}_{\mathbf{h}^1} \cdot S + (1 - \tilde{\mathbf{p}}_{\mathbf{h}^1})(1 - \mathbf{p}_{\mathbf{h}^1}) \cdot P \Big). \end{aligned} \tag{26}$$

Example: Payoffs and best responses with one-round memory. To illustrate the above results, we consider the case $n=1$. Assume player 1's self-reactive strategy is $\tilde{\mathbf{p}}^{\mathbf{1}}=(\tilde{p}^1_C,\tilde{p}^1_D)$ and player 2's reactive strategy is $\mathbf{p}^{\mathbf{2}}=(p^2_C,p^2_D)$. If we use the algorithm in Section 1.3, we first formally represent these strategies as memory-1 strategies, $\mathbf{m}^{\mathbf{1}}=(\tilde{p}^1_C,\tilde{p}^1_C,\tilde{p}^1_D,\tilde{p}^1_D)$ and $\mathbf{m}^{\mathbf{2}}=(p^2_C,p^2_D,p^2_C,p^2_D)$. The respective transition matrix according to Eq. (11) is

$$M=\begin{pmatrix} \tilde{p}^1_C p^2_C & \tilde{p}^1_C(1-p^2_C) & (1-\tilde{p}^1_C)p^2_C & (1-\tilde{p}^1_C)(1-p^2_C)\\ \tilde{p}^1_C p^2_C & \tilde{p}^1_C(1-p^2_C) & (1-\tilde{p}^1_C)p^2_C & (1-\tilde{p}^1_C)(1-p^2_C)\\ \tilde{p}^1_D p^2_D & \tilde{p}^1_D(1-p^2_D) & (1-\tilde{p}^1_D)p^2_D & (1-\tilde{p}^1_D)(1-p^2_D)\\ \tilde{p}^1_D p^2_D & \tilde{p}^1_D(1-p^2_D) & (1-\tilde{p}^1_D)p^2_D & (1-\tilde{p}^1_D)(1-p^2_D) \end{pmatrix}. \tag{27}$$

Assuming player 1's strategy is different from the one-round repeat strategy, $\tilde{\mathbf{p}}^{\mathbf{1}}\neq(1,0)$, this transition matrix has a unique invariant distribution,

$$\mathbf{v}=\left(\frac{\tilde{p}^1_D\left(\tilde{p}^1_C(p^2_C-p^2_D)+p^2_D\right)}{1-(\tilde{p}^1_C-\tilde{p}^1_D)},\frac{\tilde{p}^1_D\left(1-\tilde{p}^1_C(p^2_C-p^2_D)-p^2_D\right)}{1-(\tilde{p}^1_C-\tilde{p}^1_D)},\frac{(1-\tilde{p}^1_C)\left(\tilde{p}^1_D(p^2_C-p^2_D)+p^2_D\right)}{1-(\tilde{p}^1_C-\tilde{p}^1_D)},\frac{(1-\tilde{p}^1_C)\left(1-\tilde{p}^1_D(p^2_C-p^2_D)-p^2_D\right)}{1-(\tilde{p}^1_C-\tilde{p}^1_D)}\right).$$

According to Eq. (15), Player 1's payoff is the scalar product

$$\pi^1(\tilde{\mathbf{p}}^{\mathbf{1}},\mathbf{p}^{\mathbf{2}})=\big\langle\mathbf{v}\,,(R,S,T,P)\big\rangle. \tag{28}$$

Following Corollary 1, we can use these observations to characterize under which conditions a nice reactive strategy $\mathbf{p}^{\mathbf{2}}=(1,p^2_D)$ is a partner. To this end, we compute player 1's payoff for all pure self-reactive strategies $\tilde{\mathbf{p}}^{\mathbf{1}}=(\tilde{p}^1_C,\tilde{p}^1_D)$. These are `ALLC` $=(1,1)$, `ALLD` $=(0,0)$, and `Alternator` $=(0,1)$; we can ignore the one-round repeat strategy (1,0), because depending on the strategy's first round-behavior it is either equivalent to `ALLC` or to `ALLD`. The payoffs of these three strategies are

$$\begin{aligned} \pi^1(\texttt{ALLC},\mathbf{p}^{\mathbf{2}}) &= R,\\ \pi^1(\texttt{ALLD},\mathbf{p}^{\mathbf{2}}) &= p^2_D\cdot T+(1-p^2_D)\cdot P\\ \pi^1(\texttt{Alternator},\mathbf{p}^{\mathbf{2}}) &= p^2_D/2\cdot R+(1-p^2_D)/2\cdot S+1/2\cdot T. \end{aligned} \tag{29}$$

We conclude that player 2's reactive strategy $\mathbf{p}^{\mathbf{2}}$ is a Nash equilibrium (and hence a partner) if none of these three payoffs exceeds the mutual cooperation payoff R. This requirement yields the condition

$$p^2_D\leq\min\big\{1-(T-R)/(R-S),\ (R-P)/(T-P)\big\}. \tag{30}$$

That is, as one may expect $\mathbf{p}^{\mathbf{2}}$ is a partner if and only if its generosity p^2_D does not exceed the generosity of `GTFT`, see Eq. (8).

Instead of computing the 4×4 matrix M in (27), we could also consider the simplified 2×2 transition matrix (25). Here, the two possible states are $\mathbf{h}^{\mathbf{1}}\in\{C,D\}$, and hence the matrix is

$$\tilde{M}=\begin{pmatrix} \tilde{p}^1_C & 1-\tilde{p}^1_C\\ \tilde{p}^1_D & 1-\tilde{p}^1_D \end{pmatrix}. \tag{31}$$

Again, for $\tilde{\mathbf{p}}^{\mathbf{1}} \neq (1,0)$, this transition matrix has a unique invariant distribution,

$$\tilde{\mathbf{v}} = (\tilde{v}_C, \tilde{v}_D) = \left(\frac{\tilde{p}_D^1}{1-(\tilde{p}_C^1 - \tilde{p}_D^1)}, \frac{1-\tilde{p}_C^1}{1-(\tilde{p}_C^1 - \tilde{p}_D^1)} \right). \tag{32}$$

If we take this invariant distribution and compute player 1's payoff according to Eq. (26), we recover the same expression as in Eq. (28), as expected.

2.4 Reactive partner strategies in the donation game

Just as in the previous example with $n = 1$, we can use the results of the previous section to characterize the partner strategies for reactive-2 and reactive 3-strategies. For simplicity, we first consider the case of the donation game. Results for the general prisoner's dilemma follow in the next section.

Reactive-2 partner strategies. We first consider the case $n=2$. The resulting reactive-2 strategies can be represented as a vector $\mathbf{p} = (p_{CC}, p_{CD}, p_{DC}, p_{DD})$. The entries $p_{\mathbf{h}^{-\mathbf{i}}}$ are the player's cooperation probability, depending on the co-player's actions in the previous two rounds, $\mathbf{h}^{-\mathbf{i}} = (a_{-2}^{-i}, a_{-1}^{-i})$. For the strategy to be nice, we require $p_{CC}=1$. Based on Corollary 1, we obtain the following characterization of partners.

Theorem 2 (Reactive-2 partner strategies)
A nice reactive-2 strategy $\mathbf{p}$*, is a partner strategy if and only if its entries satisfy the conditions*

$$p_{CC}=1, \qquad \frac{p_{CD}+p_{DC}}{2} \leq 1 - \frac{1}{2}\cdot\frac{c}{b}, \qquad p_{DD} \leq 1 - \frac{c}{b}. \tag{33}$$

The resulting conditions can be interpreted as follows: For each time a co-player has defected during the past two rounds, the reactive player's cooperation probability needs to decrease by $c/(2b)$. This reduced cooperation probability is sufficient to incentivize the co-player to cooperate. Interestingly, for the strategy to be a partner, the middle condition in (33) suggests that the exact timing of a co-player's defection is irrelevant. As long as *on average*, the respective cooperation probabilities p_{CD} and p_{DC} are below the required threshold $1-c/(2b)$, the strategy is a Nash equilibrium.

The conditions for a partner become even more simple if we consider *reactive-n counting strategies*. To define these strategies, let $|\mathbf{h}^{-\mathbf{i}}|$ denote the number of C's in a given n-history of the co-player. We say a reactive-n strategy $\mathbf{p} = (p_{h^{-i}})_{\mathbf{h}^{-\mathbf{i}} \in \mathbf{H}^{-\mathbf{i}}}$ is a counting strategy if

$$|\mathbf{h}^{-i}| = |\tilde{\mathbf{h}}^{-\mathbf{i}}| \quad \Rightarrow \quad p_{\mathbf{h}^{-\mathbf{i}}} = p_{\tilde{\mathbf{h}}^{-\mathbf{i}}} \tag{34}$$

That is, the reactive player's cooperation probability only depends on the number of cooperative acts during the past n rounds and not on its timing. Such reactive-n counting strategies can be written as n+1-dimensional vectors $\mathbf{r} = (r_k)_{k \in \{n,\dots,1\}}$, where r_i is the player's cooperation probability if the co-player cooperated k times during the past n rounds. In particular, for reactive-2 counting strategies, we associate $r_2 = p_{CC}$, $r_1 = p_{CD} = p_{DC}$, and $r_0 = p_{DD}$. The following characterization of partners among the reactive-2 counting strategies then follows immediately from Theorem 2.

Corollary 2 (Partners among the reactive-2 counting strategies)
A nice reactive-2 counting strategy $\mathbf{r} = (r_2, r_1, r_0)$ *is a partner strategy if and only if*

$$r_2 = 1, \qquad r_1 \leq 1 - \frac{1}{2} \cdot \frac{c}{b}, \qquad r_0 \leq 1 - \frac{c}{b}. \tag{35}$$

Reactive-3 Partner Strategies. Now we focus on the case of $n = 3$. Reactive-3 strategies can be represented as a vector $\mathbf{p} = (p_{CCC}, p_{CCD}, p_{CDC}, p_{CDD}, p_{DCC}, p_{DCD}, p_{DDC}, p_{DDD})$. Again, each entry $p_{\mathbf{h^{-i}}}$ refers to the player's cooperation probability, depending on the co-player's previous three actions, $\mathbf{h^{-i}} = (a_{-3}^{-i}, a_{-2}^{-i}, a_{-1}^{-i})$. For the respective partner strategies, we obtain the following characterization.

Theorem 3 (Reactive-3 partner strategies)
A nice reactive-3 strategy $\mathbf{p}$ *is a partner strategy if and only if its entries satisfy the conditions*

$$\begin{aligned} p_{CCC} &= 1 \\ \frac{p_{CDC} + p_{DCD}}{2} &\leq 1 - \frac{1}{2} \cdot \frac{c}{b} \\ \frac{p_{CCD} + p_{CDC} + p_{DCC}}{3} &\leq 1 - \frac{1}{3} \cdot \frac{c}{b} \\ \frac{p_{CDD} + p_{DCD} + p_{DDC}}{3} &\leq 1 - \frac{2}{3} \cdot \frac{c}{b} \\ \frac{p_{CCD} + p_{CDD} + p_{DCC} + p_{DDC}}{4} &\leq 1 - \frac{1}{2} \cdot \frac{c}{b} \\ p_{DDD} &\leq 1 - \frac{c}{b} \end{aligned} \tag{36}$$

As before, the average of certain cooperation probabilities need to be below specific thresholds. However, compared to the case of $n = 2$, the respective conditions are now somewhat more difficult to interpret. These conditions again become more straightforward to interpret if we further restrict attention to reactive-3 counting strategies.

Corollary 3 (Partners among the reactive-3 counting strategies)
A reactive-3 counting strategy $\mathbf{r} = (r_3, r_2, r_1, r_0)$ *is a partner strategy if and only if*

$$r_3 = 1 \qquad r_2 \leq 1 - \frac{1}{3} \cdot \frac{c}{b}, \qquad r_1 \leq 1 - \frac{2}{3} \cdot \frac{c}{b}, \qquad r_0 \leq 1 - \frac{c}{b}. \tag{37}$$

As in the case of $n = 2$ we observe here that with each additional defection of the opponent in memory, the focal player reduces its conditional cooperation probability by a constant, in this case $c/(3b)$.

Partners among the memory-n counting strategies. Using the same methods as before, one can in principle also characterize the partners among the reactive-4 or the reactive-5 strategies. However, the respective conditions quickly become unwieldy. In case of the counting strategies, however, the simple pattern in

Corollaries 2 and 3 does generalize to arbitrary memory lengths.

Theorem 4 (Partners among the reactive-n counting strategies)
A nice reactive-n counting strategy $\mathbf{r}=(r_i)_{i\in\{n,n-1,\ldots,0\}}$, is a partner strategy if and only if

$$r_n=1 \qquad \textit{and} \qquad r_{n-k} \leq 1-\frac{k}{n}\cdot\frac{c}{b} \;\text{ for }\; k\in\{1,2,\ldots,n\}. \tag{38}$$

2.5 Reactive partner strategies in the general prisoner's dilemma

In the previous section, we have characterized the reactive partner strategies for a special case of the prisoner's dilemma, the donation game. In the following, we apply the same methods based on Section 2.3 to analyze the general prisoner's dilemma. For the case of reactive-2 strategies, we obtain the following characterization.

Theorem 5 (Partners among the reactive-2 strategies in the prisoner's dilemma)
A nice reactive-2 strategy $\mathbf{p}$ is a partner strategy if and only if its entries satisfy the conditions

$$\begin{array}{rcl}
p_{CC} & = & 1,\\
(T-P)\,p_{DD} & \leq & R-P,\\
(R-S)\,(p_{CD}+p_{DC}) & \leq & 3R-2S-T,\\
(T-P)\,p_{DC}+(R-S)\,p_{CD} & \leq & 2R-S-P,\\
(T-P)\,(p_{CD}+p_{DC})+(R-S)\,p_{DD} & \leq & 3R+S-2\,P,\\
(T-P)\,p_{CD}+(R-S)\,(p_{CD}+p_{DD}) & \leq & 4\,R-2\,S-P-T.
\end{array} \tag{39}$$

Compared to the donation games, there are now more conditions that need to be met, and the conditions are somewhat more difficult to interpret. Reassuringly, however, the conditions simplify to the conditions (33) in the special case that the payoff values satisfy $R=b-c$, $S=-c$, $T=b$, and $P=0$. For the case of reactive-3 strategies, the characterization is as follows.

Theorem 6
A nice reactive-3 strategy $\mathbf{p}$, is a partner strategy if and only if its entries satisfy the conditions in Table 1.

Given the complexity of the conditions in Table 1, we do not pursue deriving conditions for $n\geq 2$, even though the same methods are generally applicable.

$$
\begin{aligned}
p_{CCC} &= 1, \\
(T-P)\,(p_{CDD}+p_{DCD}+p_{DDC})+(R-S)\,p_{DDD} &\leq 4R{-}3P{-}S \\
(T-P)\,p_{CDC}+(R-S)\,p_{DCD} &< 2R{-}P{-}S \\
(T-P)\,p_{DDD} &\leq R{-}P \\
(T-P)\,(p_{CCD}+p_{CDD}+p_{DDC})+(R-S)\,(p_{CDC}+p_{DCC}+p_{DCD}+p_{DDD}) &\leq 8R{-}3P{-}4S{-}T \\
(T-P)\,p_{DCC}+(R-S)\,(p_{CCD}+p_{CDC}) &\leq 3R{-}P{-}2S \\
(T-P)\,(p_{CCD}+p_{DCC}+p_{DDC})+(R-S)\,(p_{CDC}+p_{CDD}+p_{DCD}) &\leq 6R{-}3P{-}3S \\
(T-P)\,(p_{CCD}+p_{DDC})+(R-S)\,(p_{CDC}+p_{CDD}+p_{DCC}+p_{DCD}) &\leq 7R{-}2P{-}4S{-}T \\
(T-P)\,(p_{CCD}+p_{CDD}+p_{DCC})+(R-S)\,(p_{DDC}+p_{DDD}) &\leq 5R{-}3P{-}2S \\
(T-P)\,(p_{DCD}+p_{DDC})+(R-S)\,p_{CDD} &\leq 3R{-}2P{-}S \\
(T-P)\,p_{CCD}+(R-S)\,(p_{CDD}+p_{DCC}+p_{DDC}) &\leq 5R{-}P{-}3S{-}T \\
(T-P)\,(p_{CCD}+p_{DCC})+(R-S)\,(p_{CDD}+p_{DDC}) &\leq 4R{-}2P{-}2S \\
(T-P)\,(p_{CDC}+p_{DCD})+(R-S)\,(p_{CCD}+p_{CDD}+p_{DCC}+p_{DDC}) &\leq 7R{-}2P{-}4S{-}T \\
(T-P)\,(p_{CDC}+p_{CDD}+p_{DCD})+(R-S)\,(p_{CCD}+p_{DCC}+p_{DDC}+p_{DDD}) &\leq 8R{-}3P{-}4S{-}T \\
(T-P)\,(p_{CDC}+p_{DCC}+p_{DCD})+(R-S)\,(p_{CCD}+p_{CDD}+p_{DDC}) &\leq 6R{-}3P{-}3S \\
(T-P)\,(p_{CCD}+p_{CDD}+p_{DCC}+p_{DDC})+(R-S)\,(p_{CDC}+p_{DCD}+p_{DDD}) &\leq 7R{-}4P{-}3S \\
(R-S)\,(p_{CCD}+p_{CDC}+p_{DCC}) &\leq 4R{-}3S{-}T \\
(T-P)\,(p_{CCD}+p_{CDD})+(R-S)\,(p_{DCC}+p_{DDC}+p_{DDD}) &\leq 6R{-}2P{-}3S{-}T \\
(T-P)\,(p_{CDC}+p_{CDD}+p_{DCC}+p_{DCD})+(R-S)\,(p_{CCD}+p_{DDC}+p_{DDD}) &\leq 7R{-}4P{-}3S
\end{aligned}
$$

Table 1: Necessary and sufficient conditions for a nice reactive-3 strategy to be a partner in the prisoner's dilemma.

Proof of Lemma 1. Let player 1 use a memory-1 strategy $\mathbf{m}$ and player 2 an arbitrary memory-n strategy. The probability that player 1 cooperated in the n^{th} round be denoted as v_{C}^n. Let v_{C}^n be defined as the probability that player 1 played C, k $(1 \leq k \leq n)$ rounds ago. Then,

$$v_{\mathrm{C}}^n = \sum_{h \in H} y_h, \quad \text{where} \quad y_h = \begin{cases} u_h & \text{if } \alpha_{-k}^1 = C \\ 0 & \text{if } \alpha_{-k}^1 = D. \end{cases}$$

Equivalently,

$$v_{\mathrm{C}}^n = \mathbf{v}^n \cdot \mathbf{m}^{k-\text{Rep}}.$$

Let k be fixed to $k = 1$ then,

$$v_{\mathrm{C}}^n = \mathbf{v}^n \cdot \mathbf{m}^{1-\text{Rep}}.$$

Moreover, the probability that player 1 cooperates in the $(n+1)^{th}$ round, denoted by $v_{\mathrm{C}}^{n+1} = \mathbf{v}^n \cdot \mathbf{m}$. Hence,

$$v_{\mathrm{C}}^{n+1} - v_{\mathrm{C}}^n = \mathbf{v}^n \cdot \mathbf{m} - \mathbf{v}^n \cdot \mathbf{m}^{1-\text{Rep}} = \mathbf{v}^n \cdot (\mathbf{m} - \mathbf{m}^{1-\text{Rep}}).$$

This implies,

$$\sum_{t=1}^{n} \mathbf{v}^t \cdot (\mathbf{m} - \mathbf{m}^{1-\text{Rep}}) = \sum_{t=1}^{n} v_{\mathrm{C}}^{t+1} - v_{\mathrm{C}}^t \quad \Rightarrow \quad \sum_{t=1}^{n} \mathbf{v}^t \cdot (\mathbf{m} - \mathbf{m}^{1-\text{Rep}}) = v_{\mathrm{C}}^{n+1} - v_{\mathrm{C}}^1.$$

As the right side has absolute value at most 1,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} \mathbf{v}^t \cdot (\mathbf{m} - \mathbf{m}^{1-\text{Rep}}) = 0.$$

Repeat for $1 < k \leq n$.

□

Sketch of the proof of Theorem 1. By Lemma 2, we know there exists a best response to $\mathbf{p}$ within the self-reactive n strategies. It thus remains to show that this best response $\tilde{\mathbf{p}}$ can be chosen to be pure. The proof follows from a series of auxiliary results:

Auxiliary result 1: Let $\tilde{\mathbf{p}} \in \mathcal{S}_n$ such that its transition matrix $\tilde{M}$ according to Eq. (27) has a unique invariant distribution $\tilde{\mathbf{v}} = (\tilde{v}_{\mathbf{h}^1})_{\mathbf{h}^1 \in H^1}$. Moreover, for a given history $\mathbf{h}' \in H^1$, let $\tilde{M}_{\mathbf{h}'}$ be the $(2^n - 1) \times (2^n - 1)$ matrix one obtains from $\tilde{M}$ after subtracting the $2^n \times 2^n$ identity matrix, and then deleting the row and the column associated to $\mathbf{h}'$. Then $(\tilde{v}_{\mathbf{h}'})_{\mathbf{h}' \in H^i}$ is up to a normalization factor equal to $\big(\det(\tilde{M}_{\mathbf{h}'})\big)_{\mathbf{h}' \in H^i}$.

Auxiliary result 2: On its domain, the payoff function $\pi^1(\tilde{\mathbf{p}}, \mathbf{p})$ is a rational function, and both its numerator and denominator can be chosen to be affine linear in each entry $\tilde{p}_{\mathbf{h^i}}$, for all $\mathbf{h^i} \in H^i$.

Auxiliary result 3: Suppose $g, h : [0,1]^{2^n} \to \mathbb{R}$ and suppose both $g(\mathbf{x})$ and $h(\mathbf{x})$ are affine linear in each component of $\mathbf{x} = (x_1, \ldots, x_{2^n})$. If the rational function $f := g/h$ is continuous on $[0,1]^{2^n}$ then it has a maximum in one of the corners. That is, there is some $\mathbf{x}^* \in \{0,1\}^{2^n}$ such that $f(\mathbf{x}^*) \geq f(\mathbf{x})$ for all $\mathbf{x} \in [0,1]^{2^n}$. □

A Proofs for Theorems Based on Pure Self-reactive Strategies Result

A.1 Proof of Theorem 2

Suppose player 1 adopts a nice reactive-2 strategy $\mathbf{p} = (1, p_{CD}, p_{DC}, p_{DD})$. For $\mathbf{p}$ to be a Nash strategy,

$$s_{\tilde{\mathbf{p}},\mathbf{p}} \leq (b - c),$$

must hold against all $\tilde{\mathbf{p}} \in \tilde{P}$, where $\tilde{P}$ is the set of all pure self-reactive-2 strategies. In the case of $n = 2$, the set contains 16 strategies.

Proof. Suppose player 1 plays a nice reactive-2 strategy $\mathbf{p} = (1, p_{CD}, p_{DC}, p_{DD})$, and suppose the co-player 2 plays a pure self-reactive-2 strategy $\tilde{\mathbf{p}}$. The possible payoffs for $\tilde{\mathbf{p}} \in \{\tilde{\mathbf{p}}^0, \ldots, \tilde{\mathbf{p}}^{16}\}$ are:

$$
\begin{aligned}
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= \quad b \cdot p_{DD} && for \;\; i \in \quad \{0, 2, 4, 6, 8, 10, 12, 14\} \\
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= \quad \tfrac{b \cdot (p_{CD} + p_{DC} + p_{DD})}{3} - \tfrac{1}{3} \cdot c && for \;\; i \in \quad \{1, 9\} \\
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= \quad \tfrac{b \cdot (p_{CD} + p_{DC} + p_{DD} + 1)}{4} - \tfrac{1}{2} \cdot c && for \;\; i \in \quad \{3\} \\
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= \quad \tfrac{b \cdot (p_{CD} + p_{DC})}{2} - \tfrac{1}{2} \cdot c && for \;\; i \in \quad \{4, 5, 12, 13\} \\
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= \quad \tfrac{b \cdot (p_{CD} + p_{DC} + 1)}{3} - \tfrac{2}{3} \cdot c && for \;\; i \in \quad \{6, 7\} \\
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= \quad b - c && for \;\; i \in \quad \{8, 9, 10, 11, 12, 13, 14, 15\}
\end{aligned}
$$

Setting the payoff expressions of $s_{\tilde{\mathbf{p}}^i,\mathbf{p}}$ to smaller or equal to $(b - c)$ we get the following unique conditions,

$$p_{DD} \leq 1 - \frac{c}{b} \tag{40}$$

$$\frac{p_{CD} + p_{DC}}{2} \leq 1 - \frac{1}{2} \cdot \frac{c}{b} \tag{41}$$

$$\frac{p_{CD} + p_{DC} + p_{DD}}{3} \leq 1 - \frac{2}{3} \cdot \frac{c}{b} \tag{42}$$

Notice that only conditions (40) and (41) are necessary.

$\square$

A.2 Proof of Theorem 3

Consider all the pure self-reactive-3 strategies. There is a total of 256 such strategies. The payoff expression for each of these strategies against a nice reactive-3 strategies can be calculated explicitly. We use these expressions to obtain the conditions for partner strategies similar to the previous section.

Proof. The payoff expressions for a nice reactive-3 strategy $\mathbf{p}$ against all pure self-reactive-3 strategies are as follows,

$$
\begin{aligned}
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= b\, p_{DDD} && for\ \ i \in && \{0, 2, 4, 6, \ldots, 250, 252, 254\} \\
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= \tfrac{b\cdot(p_{CDD}+p_{DCD}+p_{DDC}+p_{DDD})}{4} - \tfrac{1}{4}\cdot c && for\ \ i \in && \{1, 9, 33, 41, 65, 73, 97, 105, 129, 137, 161, \\
&&&&& 169, 193, 201, 225, 233\} \\
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= \tfrac{b\cdot(p_{CCD}+p_{CDD}+p_{DCC}+p_{DDC}+p_{DDD})}{5} - \tfrac{2}{5}\cdot c && for\ \ i \in && \{3, 7, 35, 39, 131, 135, 163, 167\} \\
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= \tfrac{b\cdot(p_{CDC}+p_{DCD})}{2} - \tfrac{1}{2}\cdot c && for\ \ i \in && \{4-7, 12-15, 20-23, 28-31, 68-71, \\
&&&&& 76-79, 84-87, 92-95, 132-135, \\
&&&&& 140-143, 148-151, 156-159, \\
&&&&& 196-199, 204-207, 212-215, 220-223\} \\
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= \tfrac{b\cdot(p_{CCD}+p_{CDD}+p_{DCC}+p_{DDC}+p_{DDD}+1)}{6} - \tfrac{1}{2}\cdot c && for\ \ i \in && \{11, 15, 43, 47\} \\
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= \tfrac{b\cdot(p_{CDD}+p_{DCD}+p_{DDC})}{3} - \tfrac{1}{3}\cdot c && for\ \ i \in && \{16, 17, 24, 25, 48, 49, 56, 57, 80, 81, 88, \\
&&&&& 89, 112, 113, 120, 121, 144, 145, 152, 153, \\
&&&&& 176, 177, 184, 185, 208, 209, 216, 217, \\
&&&&& 240, 241, 248, 249\} \\
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= \tfrac{b\cdot(p_{CCD}+p_{CDD}+p_{DCC}+p_{DDC})}{4} - \tfrac{1}{2}\cdot c && for\ \ i \in && \{18, 19, 22, 23, 50, 51, 54, 55, 146, 147, \\
&&&&& 150, 151, 178, 179, 182, 183\} \\
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= \tfrac{b\cdot(p_{CCD}+p_{CDD}+p_{DCC}+p_{DDC}+1)}{5} - \tfrac{3}{5}\cdot c && for\ \ i \in && \{26, 27, 30, 31, 58, 59, 62, 63\} \\
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= \tfrac{b\cdot(p_{CCD}+p_{CDC}+p_{CDD}+p_{DCC}+p_{DCD}+p_{DDC}+p_{DDD})}{7} - \tfrac{3}{7}\cdot c && for\ \ i \in && \{37, 67, 165, 195\} \\
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= \tfrac{b\cdot(p_{CCD}+p_{CDC}+p_{CDD}+p_{DCC}+p_{DCD}+p_{DDC}+p_{DDD}+1)}{8} - \tfrac{1}{2}\cdot c && for\ \ i \in && \{45, 75\} \\
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= \tfrac{b\cdot(p_{CCD}+p_{CDC}+p_{CDD}+p_{DCC}+p_{DCD}+p_{DDC})}{6} - \tfrac{1}{2}\cdot c && for\ \ i \in && \{52, 53, 82, 83, 180, 181, 210, 211\} \\
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= \tfrac{b\cdot(p_{CCD}+p_{CDC}+p_{CDD}+p_{DCC}+p_{DCD}+p_{DDC}+1)}{7} - \tfrac{4}{7}\cdot c && for\ \ i \in && \{60, 61, 90, 91\} \\
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= \tfrac{b\cdot(p_{CCD}+p_{CDC}+p_{DCC})}{3} - \tfrac{2}{3}\cdot c && for\ \ i \in && \{96-103, 112-119, 224-231, 240-247\} \\
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= \tfrac{b\cdot(p_{CCD}+p_{CDC}+p_{DCC}+1)}{4} - \tfrac{3}{4}\cdot c && for\ \ i \in && \{104-111, 120-127\} \\
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= (b-c) && for\ \ i \in && \{128, 129, 130, \ldots, 255\}
\end{aligned}
\tag{43}
$$

Setting these to smaller or equal than the mutual cooperation payoff $(b-c)$ give the following ten conditions,

$$p_{DDD} \leq 1 - \frac{c}{b}, \quad \frac{p_{CDC} + p_{DCD}}{2} \leq 1 - \frac{1}{2} \cdot \frac{c}{b}, \quad \frac{p_{CDD} + p_{DCD} + p_{DDC}}{3} \leq 1 - \frac{2}{3} \cdot \frac{c}{b}, \tag{44}$$

$$\frac{p_{CCD} + p_{CDC} + p_{DCC}}{3} \leq 1 - \frac{1}{3} \cdot \frac{c}{b}, \quad \frac{p_{CCD} + p_{CDD} + p_{DCC} + p_{DDC}}{4} \leq 1 - \frac{1}{2} \cdot \frac{c}{b} \tag{45}$$

$$\frac{p_{CDD} + p_{DCD} + p_{DDC} + p_{DDD}}{4} \leq 1 - \frac{3}{4} \cdot \frac{c}{b}, \tag{46}$$

$$\frac{p_{CCD} + p_{CDC} + p_{CDD} + p_{DCC} + p_{DCD} + p_{DDC} + p_{DDD}}{7} \leq 1 - \frac{4}{7} \cdot \frac{c}{b}, \tag{47}$$

$$\frac{p_{CCD} + p_{CDD} + p_{DCC} + p_{DDC} + p_{DDD}}{5} \leq 1 - \frac{3}{5} \cdot \frac{c}{b}, \tag{48}$$

$$\frac{p_{CCD} + p_{CDC} + p_{CDD} + p_{DCC} + p_{DCD} + p_{DDC}}{6} \leq 1 - \frac{1}{2} \cdot \frac{c}{b} \tag{49}$$

Notice that only the conditions of Eq. (44) and (45) are necessary. The remaining conditions can be derived from the sums of conditions in Eq. (44) and (45). □

A.3 Proof of Theorem 5

There are 16 pure-self reactive strategies in $n = 2$. We use calculate the explicit payoff expressions for each pure self-reactive strategy against a nice reactive-2 strategy as given by Eq. (26). This gives the following payoff expressions:

$$\begin{aligned}
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= && P(1 - p_{DD}) + T p_{DD} && for \;\; i \in \{0, 2, 4, 6, 8, 10, 12, 14\} \\
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= && \tfrac{-P(p_{CD}+p_{DC}-2)+Rp_{DD}-S(p_{DD}-1)+T(p_{CD}+p_{DC})}{3} && for \;\; i \in \{1, 9\} \\
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= && \tfrac{P(1-p_{CD})+R(p_{DC}+p_{DD})-S(p_{DC}+p_{DD}-2)+T(p_{CD}+1)}{4} && for \;\; i \in \{3\} \\
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= && \tfrac{P(1-p_{DC})+Rp_{CD}-S(p_{CD}-1)+Tp_{DC}}{2} && for \;\; i \in \{4, 5, 12, 13\} \\
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= && \tfrac{R(p_{CD}+p_{DC})-S(p_{CD}+p_{DC}-2)+T}{3} && for \;\; i \in \{6, 7\} \\
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= && R && for \;\; i \in \{8, 9, 10, 11, 12, 13, 14, 15\}
\end{aligned}$$

Setting the above expressions to $\leq R$ gives the following conditions,

$$\begin{aligned}
(T - P)\, p_{DD} &< R - P, \\
(R - S)\, (p_{CD} + p_{DC}) &< 3R - 2S - T, \\
(T - P)\, p_{DC} + (R - S)\, p_{CD} &< 2R - S - P, \\
(T - P)\, (p_{CD} + p_{DC}) + (R - S)\, p_{DD} &< 3R + S - 2\,P, \\
(T - P)\, p_{CD} + (R - S)\, (p_{CD} + p_{DD}) &< 4\,R - 2\,S - P - T.
\end{aligned}$$

A.4 Proof of Theorem 6

Previously as in the previous subsection we calculate the explicit payoff expressions for each $\tilde{\mathbf{p}} \in \tilde{P}$ against a nice reactive-3. The set of pure self-reactive strategies $\tilde{P}$ in $n = 3$ contains 256 elements. The expressions for each strategy are given below,

$$s_{\tilde{\mathbf{p}}^i,\mathbf{p}} = \frac{(T-P)\,(p_{CDD}+p_{DCD}+p_{DDC})+3\,P+(R-S)\,p_{DDD}+S}{4} \qquad for\ \ i \in \{1, 9, 33, 41, 65, 73, 97, 10\ 129, 137, 161, 169, 193, 20\ 225, 233\}$$

$$s_{\tilde{\mathbf{p}}^i,\mathbf{p}} = \frac{(T-P)\,p_{CDC}+P+(R-S)\,p_{DCD}+S}{2} \qquad for\ \ i \in \{4-7, 12-15, 20-23, 28-31, 68-71, 76-79, 84-87, 92-95, 132-135\ 140-143, 148-151, 156\ 196-199, 204-207, 212-\ 220-223\}$$

$$s_{\tilde{\mathbf{p}}^i,\mathbf{p}} = -P\,(p_{DDD}-1)+Tp_{DDD} \qquad for\ \ i \in \{0, 2, 4, \ldots, 252, 254\}$$

$$s_{\tilde{\mathbf{p}}^i,\mathbf{p}} = \frac{(T-P)\,(p_{CCD}+p_{CDD}+p_{DDC})+3\,P+(R-S)\,(p_{CDC}+p_{DCC}+p_{DCD}+p_{DDD})+4\,S+T}{8} \qquad for\ \ i \in \{45\}$$

$$s_{\tilde{\mathbf{p}}^i,\mathbf{p}} = \frac{(T-P)\,p_{DCC}+P+(R-S)\,(p_{CDC}+p_{CCD})+2\,S}{3} \qquad for\ \ i \in \{96-103, 112-119, 224-231, 240-247\}$$

$$s_{\tilde{\mathbf{p}}^i,\mathbf{p}} = \frac{(T-P)\,(p_{CCD}+p_{DCC}+p_{DDC})+3\,P+(R-S)\,(p_{CDC}+p_{CDD}+p_{DCD})+3\,S}{6} \qquad for\ \ i \in \{52, 53, 180, 181\}$$

$$s_{\tilde{\mathbf{p}}^i,\mathbf{p}} = \frac{(T-P)\,(p_{CCD}+p_{DDC})+2\,P+T+(R-S)\,(p_{CDC}+p_{CDD}+p_{DCC}+p_{DCD})+4\,S}{7} \qquad for\ \ i \in \{60, 61\}$$

$$s_{\tilde{\mathbf{p}}^i,\mathbf{p}} = \frac{(T-P)\,(p_{CCD}+p_{CDD}+p_{DCC})+3\,P+(R-S)\,(p_{DDC}+p_{DDD})+2\,S}{5} \qquad for\ \ i \in \{3, 7, 35, 39, 131, 135, 163$$

$$s_{\tilde{\mathbf{p}}^i,\mathbf{p}} = \frac{(T-P)\,(p_{DCD}+p_{DDC})+2\,P+(R-S)\,p_{CDD}+S}{3} \qquad for\ \ i \in \{16, 17, 24, 25, 48, 49, 56, 57, 80, 81, 88, 89, 112, 113\ 120, 121, 144, 145, 152, 15\ 176, 177, 184, 185, 208, 20\ 216, 217, 240, 241, 248, 24$$

$$s_{\tilde{\mathbf{p}}^i,\mathbf{p}} = R \qquad for\ \ i \in \{128, 129, \ldots, 255\}$$

$$s_{\tilde{\mathbf{p}}^i,\mathbf{p}} = \frac{(T-P)\,p_{CCD}+P+T+(R-S)\,(p_{CDD}+p_{DCC}+p_{DDC})+3S}{5} \qquad for\ \ i \in \{26, 27, 30, 31, 58, 59, 62,$$

$$s_{\tilde{\mathbf{p}}^i,\mathbf{p}} = \frac{(T-P)\,(p_{CCD}+p_{DCC})+2\,P+(R-S)\,(p_{CDD}+p_{DDC})+2\,S}{4} \qquad for\ \ i \in \{18, 19, 22, 23, 50, 51, 54,\ 146, 147, 150, 151, 178, 17\ 182, 183\}$$

$$s_{\tilde{\mathbf{p}}^i,\mathbf{p}} = \frac{(T-P)\,(p_{CDC}+p_{DCD})+2\,P+T+(R-S)\,(p_{CCD}+p_{CDD}+p_{DCC}+p_{DDC})+4\,S}{7} \qquad for\ \ i \in \{90, 91\}$$

$$s_{\tilde{\mathbf{p}}^i,\mathbf{p}} = \frac{(T-P)\,(p_{CDC}+p_{CDD}+p_{DCD})+3\,P+T+(R-S)\,(p_{CCD}+p_{DCC}+p_{DDC}+p_{DDD})+4\,S}{8} \qquad for\ \ i \in \{75\}$$

$$s_{\tilde{\mathbf{p}}^i,\mathbf{p}} = \frac{(T-P)\,(p_{CDC}+p_{DCC}+p_{DCD})+3\,P+(R-S)\,(p_{CCD}+p_{CDD}+p_{DDC})+3\,S}{6} \qquad for\ \ i \in \{82, 83, 210, 211\}$$

$$s_{\tilde{\mathbf{p}}^i,\mathbf{p}} = \frac{(T-P)\,(p_{CCD}+p_{CDD}+p_{DCC}+p_{DDC})+4\,P+(R-S)\,(p_{CDC}+p_{DCD}+p_{DDD})+3\,S}{7} \qquad for\ \ i \in \{37, 165\}$$

$$s_{\tilde{\mathbf{p}}^i,\mathbf{p}} = \frac{T+(R-S)\,(p_{CCD}+p_{CDC}+p_{DCC})+3\,S}{4} \qquad for\ \ i \in \{104-111, 120-127\}$$

$$s_{\tilde{\mathbf{p}}^i,\mathbf{p}} = \frac{(T-P)\,(p_{CCD}+p_{CDD})+2\,P+T+(R-S)\,(p_{DCC}+p_{DDC}+p_{DDD})+3\,S}{6} \qquad for\ \ i \in \{11, 15, 43, 47\}$$

$$s_{\tilde{\mathbf{p}}^i,\mathbf{p}} = \frac{(T-p)\,(p_{CDC}+p_{CDD}+p_{DCC}+p_{DCD})+4\,P+(R-S)\,(p_{CCD}+p_{DDC}+p_{DDD})+3\,S}{7} \qquad for\ \ i \in \{67, 195\}$$

Setting the above expressions to $\leq R$ gives the following conditions,

$$
\begin{aligned}
(T-P)\left(p_{CDD}+p_{DCD}+p_{DDC}\right)+(R-S)\,p_{DDD} &< 4R-3P-S\\
(T-P)\,p_{CDC}+(R-S)\,p_{DCD} &< 2R-P-S\\
(T-P)\,p_{DDD} &< R-P\\
(T-P)\left(p_{CCD}+p_{CDD}+p_{DDC}\right)+(R-S)\left(p_{CDC}+p_{DCC}+p_{DCD}+p_{DDD}\right) &< 8R-3P-4S-T\\
(T-P)\,p_{DCC}+(R-S)\left(p_{CCD}+p_{CDC}\right) &< 3R-P-2S\\
(T-P)\left(p_{CCD}+p_{DCC}+p_{DDC}\right)+(R-S)\left(p_{CDC}+p_{CDD}+p_{DCD}\right) &< 6R-3P-3S\\
(T-P)\left(p_{CCD}+p_{DDC}\right)+(R-S)\left(p_{CDC}+p_{CDD}+p_{DCC}+p_{DCD}\right) &< 7R-2P-4S-T\\
(T-P)\left(p_{CCD}+p_{CDD}+p_{DCC}\right)+(R-S)\left(p_{DDC}+p_{DDD}\right) &< 5R-3P-2S\\
(T-P)\left(p_{DCD}+p_{DDC}\right)+(R-S)\,p_{CDD} &< 3R-2P-S\\
(T-P)\,p_{CCD}+(R-S)\left(p_{CDD}+p_{DCC}+p_{DDC}\right) &< 5R-P-3S-T\\
(T-P)\left(p_{CCD}+p_{DCC}\right)+(R-S)\left(p_{CDD}+p_{DDC}\right) &< 4R-2P-2S\\
(T-P)\left(p_{CDC}+p_{DCD}\right)+(R-S)\left(p_{CCD}+p_{CDD}+p_{DCC}+p_{DDC}\right) &< 7R-2P-4S-T\\
(T-P)\left(p_{CDC}+p_{CDD}+p_{DCD}\right)+(R-S)\left(p_{CCD}+p_{DCC}+p_{DDC}+p_{DDD}\right) &< 8R-3P-4S-T\\
(T-P)\left(p_{CDC}+p_{DCC}+p_{DCD}\right)+(R-S)\left(p_{CCD}+p_{CDD}+p_{DDC}\right) &< 6R-3P-3S\\
(T-P)\left(p_{CCD}+p_{CDD}+p_{DCC}+p_{DDC}\right)+(R-S)\left(p_{CDC}+p_{DCD}+p_{DDD}\right) &< 7R-4P-3S\\
(R-S)\left(p_{CCD}+p_{CDC}+p_{DCC}\right) &< 4R-3S-T\\
(T-P)\left(p_{CCD}+p_{CDD}\right)+(R-S)\left(p_{DCC}+p_{DDC}+p_{DDD}\right) &< 6R-2P-3S-T\\
(T-P)\left(p_{CDC}+p_{CDD}+p_{DCC}+p_{DCD}\right)+(R-S)\left(p_{CCD}+p_{DDC}+p_{DDD}\right) &< 7R-4P-3S
\end{aligned}
$$

B Proofs for Theorems Based on Generalized Akin's Lemma

B.1 Further Notation

In this section, we introduce some additional notation that will be used in the following subsection to prove our theorems.

Once again, we assume the setup in which player 1 adopts a reactive-n strategy $\mathbf{p}$, and player 2 adopts a self-reactive-n strategy $\tilde{\mathbf{p}}$. We define the following marginal distributions with respect to the possible n-histories of player 2:

$$v^2_{h^2} = \sum_{h^1 \in H^1} v_{(h^1,h^2)}. \tag{50}$$

These entries describe how often we observe player 2 to choose actions h^2, in n consecutive rounds (irrespective of the actions of player 1). Note that,

$$\sum_{h \in H^2} v_h^2 = 1. \tag{51}$$

Let $\mathbf{p}^{\mathbf{k}-\mathrm{Rep}}$ be a reactive round-k-repeat strategy. Then the cooperation rate of player 2, denoted as $\rho_{\tilde{\mathbf{p}}}$, and based on Lemma 1 is given by,

$$\rho_{\tilde{\mathbf{p}}} = \sum_{h^2 \in H^2} v_{h^2}^2 \cdot p_h^{1-\mathrm{Rep}} \;=\; \sum_{h^2 \in H^2} v_{h^2}^2 \cdot p_h^{2-\mathrm{Rep}} \;=\; \ldots \;=\; \sum_{h^2 \in H^2} v_{h^2}^2 \cdot p_h^{n-\mathrm{Rep}}. \tag{52}$$

Player's 1 cooperation rate can also be defined in a similar manner. However, here we define the cooperation rate of player 1 as,

$$\rho_{\mathbf{p}} \;=\; \sum_{h^2 \in H^2} v_{h^2}^2 \cdot p_{h^2}. \tag{53}$$

B.2 Proof of Theorem 2

Suppose player 1 adopts a reactive-2 strategy $\mathbf{p} = (p_{CC}, p_{CD}, p_{DC}, p_{DD})$. Moreover, suppose player 2 adopts an arbitrary memory-2 strategy $\mathbf{m}$. Let $\mathbf{v} = (v_h)_{h \in H}$ be an invariant distribution of the game between the two players.

The cooperation rate of player 2 given by 52 in the case of $n = 2$ is given by,

$$\rho_{\mathbf{m}} := v_{CC}^2 + v_{CD}^2 = v_{CC}^2 + v_{DC}^2. \tag{54}$$

We can use this equality to conclude that

$$v_{CD}^2 = v_{DC}^2. \tag{55}$$

Moreover the cooperation rate of player 1 based on Eq. 53 is given by,

$$\begin{aligned} \rho_{\mathbf{p}} &= v_{CC}^2\, p_{CC} + v_{CD}^2\, p_{CD} + v_{DC}^2\, p_{DC} + v_{DD}^2\, p_{DD} \\ &= v_{CC}^2\, p_{CC} + v_{CD}^2\, (p_{CD} + p_{DC}) + v_{DD}^2\, p_{DD}. \end{aligned} \tag{56}$$

Here, the second equality is due to Eq. (55).

Proof. ($\Rightarrow$)A reactive-2 strategy $\mathbf{p} = (p_{CC}, p_{CD}, p_{DC}, p_{DD})$ 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

$$\text{AllD} = (0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0) \text{ and Alternator} = (0,0,1,1,0,0,1,1,0,0,1,1,0,0,1,1).$$

Thus, $\mathbf{p}$ can only form a Nash equilibrium if

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

or equivalently, if

$$p_{DD} \leq 1 - \frac{c}{b} \quad \text{and} \quad p_{CD} + p_{DC} \leq 1 + \frac{b-c}{c}. \tag{57}$$

($\Leftarrow$) Now, suppose player 2 has some strategy $\mathbf{m}$ such that $s_{\mathbf{m},\mathbf{p}} > b - c$. It follows that

$$\begin{aligned}
0 \quad &< \quad s_{\mathbf{m},\mathbf{p}} - (b-c) \\
&\overset{Eq.\ (\mathbf{??})}{=} \quad b\rho_{\mathbf{p}} - c\rho_{\mathbf{m}} - (b-c) \\
&\overset{Eqs.\ (54),(56),(51)}{=} \quad b\left(v^2_{CC}p_{CC} + v^2_{CD}(p_{CD}+p_{DC}) + v^2_{DD}p_{DD}\right) - c\left(v^2_{CC}+v^2_{CD}\right) - (b-c)\left(v^2_{CC}+2v^2_{CD}+v^2_{DD}\right) \\
&= \quad v^2_{CC}\, b\,(p_{CC}-1) + v^2_{CD}\Big(b(p_{CD}+p_{DC})+c-2b\Big) + v^2_{DD}\Big(bp_{DD}-(b-c)\Big).
\end{aligned} \tag{58}$$

Condition (58) can hold only if,

$$b\,(p_{CD}+p_{DC})+c-2b > 0, \qquad b\,p_{DD} - (b-c) > 0. \tag{59}$$

Thus, Eq. (57) reassures that $\mathbf{p}$ is Nash strategy, and given that $p_{CC} = 1$, it is a partner strategy. $\square$

B.3 Proof of Theorem 3

Suppose player 1 adopts a reactive-3 strategy $\mathbf{p}$, and suppose player 2 adopts an arbitrary memory-3 strategy $\mathbf{m}$. Let $\mathbf{v} = (v_h)_{h \in H}$ be an invariant distribution of the game between the two players.

The average cooperation rate $\rho_{\mathbf{m}}$ of player 2 (Eq. 52) for $n = 3$ is given by,

$$\rho_{\mathbf{m}} := v^2_{CCC}+v^2_{CCD}+v^2_{DCC}+v^2_{DCD} = v^2_{CCC}+v^2_{DCC}+v^2_{CDC}+v^2_{DDC} = v^2_{CCC}+v^2_{CCD}+v^2_{CDC}+v^2_{CDD}. \tag{60}$$

We can use this equality to conclude that

$$v^2_{CCD} = v^2_{DCC} \tag{61}$$

$$v^2_{DDC} = v^2_{CDD} \tag{62}$$

$$v^2_{CCD} + v^2_{DCD} = v^2_{CDC} + v^2_{DDC} \Rightarrow v^2_{CCD} = v^2_{CDC} + v^2_{CDD} - v^2_{DCD} \tag{63}$$

The average cooperation rate of 1's (Eq. (53)) for $n = 3$ is given by,

$$
\begin{aligned}
\rho_{\mathbf{p}} \quad &= \quad v^2_{CCC}\, p_{CCC} + v^2_{CCD}\, p_{CCD} + v^2_{CDC}\, p_{CDC} + v^2_{CDD}\, p_{CDD} + v^2_{DCD}\, p_{DCD} + \\
& \qquad + v^2_{DDC}\, p_{DDC} + v^2_{DDD}\, p_{DDD} \\
&\overset{Eqs.\ (61),(62)}{=} \quad v^2_{CCC}\, p_{CCC} + v^2_{CCD}\,(p_{CCD} + p_{DCC}) + v^2_{CDC}\, p_{CDC} + v^2_{CDD}\,(p_{CDD} + p_{DDC}) + \\
& \qquad v^2_{DCD}\, p_{DCD} + v^2_{DDD}\, p_{DDD}
\end{aligned}
\tag{64}
$$

Proof. ($\Rightarrow$) A reactive-3 strategy $\mathbf{p}$ can only be a Nash equilibrium if *no* other strategy yields a larger payoff, in particular neither AllD nor the following self-reactive-3 strategies,

$$
\begin{aligned}
\tilde{\mathbf{p}}^{\mathbf{15}} &= (0,0,0,0,1,1,1,1) \\
\tilde{\mathbf{p}}^{\mathbf{17}} &= (0,0,0,1,0,0,0,1) \\
\tilde{\mathbf{p}}^{\mathbf{51}} &= (0,0,1,1,0,0,1,1) \\
\tilde{\mathbf{p}}^{\mathbf{119}} &= (0,1,1,1,0,1,1,1).
\end{aligned}
$$

The above strategies are alternating strategies.For instance, $\tilde{\mathbf{p}}^{\mathbf{15}}$ and $\tilde{\mathbf{p}}^{\mathbf{51}}$ are delayed alternating strategies. $\tilde{\mathbf{p}}^{\mathbf{15}}$ cooperates if and only if defected three rounds ago, and $\tilde{\mathbf{p}}^{\mathbf{15}}$ cooperates after defecting 2 rounds ago. $\tilde{\mathbf{p}}^{\mathbf{17}}$ and $\tilde{\mathbf{p}}^{\mathbf{119}}$ alternate between cooperating and defecting after given sequences occur. Namely, $\tilde{\mathbf{p}}^{\mathbf{17}}$ cooperates after DD sequence has occurred, and $\tilde{\mathbf{p}}^{\mathbf{119}}$ defects after CCC sequence has occurred.

$\mathbf{p}$ can only form a Nash equilibrium if

$$
\pi(\text{AllD}, \mathbf{p}) \le b - c \quad and \quad \pi(\tilde{\mathbf{p}}^i, \mathbf{p}) \le b - c \text{ for } i \in \{15, 17, 51, 102\}.
$$

or equivalently, if

$$
\begin{aligned}
\frac{p_{CCD} + p_{CDC} + p_{DCC}}{3} &< 1 - \frac{1}{3} \cdot \frac{c}{b} \\
\frac{p_{CDD} + p_{DCD} + p_{DDC}}{3} &< 1 - \frac{2}{3} \cdot \frac{c}{b} \\
p_{DDD} &< 1 - \frac{c}{b} \\
\frac{p_{CCD} + p_{CDD} + p_{DCC} + p_{DDC}}{4} &< 1 - \frac{1}{2} \cdot \frac{c}{b} \\
\frac{p_{CDC} + p_{DCD}}{2} &< 1 - \frac{1}{2} \cdot \frac{c}{b}
\end{aligned}
\tag{65}
$$

($\Leftarrow$) Now, suppose player 2 has some strategy $\mathbf{m}$ such that $s_{\mathbf{m},\mathbf{p}} > b - c$. It follows that

$$
\begin{aligned}
0 &\leq s_{\mathbf{m},\mathbf{p}} - (b-c) \\
&\overset{Eq.\ (??)}{=} b\rho_{\mathbf{p}} - c\rho_{\mathbf{m}} - (b-c) \\
&\overset{Eqs.\ (64),(51)}{=} b\left(v^2_{CCC}p_{CCC} + v^2_{CCD}(p_{CCD}+p_{DCC}) + v^2_{CDC}p_{CDC} + v^2_{DDC}(p_{CDD}+p_{DDC}) + v^2_{DCD}p_{DCD} + v^2_{DDD}p_{DDD}\right. \\
&\quad -c\left(v^2_{CCC}+2v^2_{CCD}+v^2_{DCD}\right) - (b-c)\left(v^2_{CCC}+2v^2_{CCD}+v^2_{CDC}+2v^2_{DDC}+v^2_{DCD}+v^2_{DDD}\right) \\
&= b\,v^2_{CCC}(p_{CCC}-1) + v^2_{CCD}(b\,(p_{CCD}+p_{DCC}-2)) + v^2_{CDC}(b\,(p_{CDC}-1)+c) + \\
&\quad v^2_{CDD}(b\,(p_{CDD}+p_{DDC}-2)+2\,c) + v^2_{DCD}(b\,(p_{DCD}-1)) + v^2_{DDD}(b\,(p_{DDD}-1)+c) \\
&\overset{Eq.\ (63)}{=} b\,v^2_{CCC}(p_{CCC}-1) + v^2_{DDD}(b\,(p_{DDD}-1)+c) + v^2_{CDC}(b\,(p_{CCD}+p_{DCC}+p_{CDC}-3)+c) + \\
&\quad v^2_{CDD}(b\,(p_{CDD}+p_{DDC}+p_{CCD}+p_{DCC}-4)+2\,c) + v^2_{DCD}(b\,(p_{DCD}-1) - b\,(p_{CCD}+p_{DCC}) - 2)
\end{aligned}
\tag{66}
$$

Condition (66) holds only for,

$$
b\,(p_{DDD}-1)+c<0, \quad b\,(p_{CCD}+p_{DCC}+p_{CDC}-3)+c
$$

$$
b\,(p_{CDD}+p_{DDC}+p_{CCD}+p_{DCC}-4)+2\,c<0 \Rightarrow -b\,(p_{CCD}+p_{DCC}-2) > b\,(p_{CDD}+p_{DDC}-2)+2\,c
$$

$$
b\,(p_{DCD}-1) - b\,(p_{CCD}+p_{DCC}) - 2 < 0 \Rightarrow b\,(p_{DCD}+p_{CDD}+p_{DDC}-3)+2\,c<0.
$$

Thus, conditions Eq. (65) reassure that $\mathbf{p}$ is Nash strategy, and given that $p_{CC} = 1$, it is a partner strategy. □

C Proof of Theorem 4

To prove Theorem 4, we need to introduce some additional notation. We introduce the vector $\mathbf{w} = (w_i)_{i\in\{0,1,\ldots,n\}}$, where the entry w_i represents the probability that, in the long-term outcome, the co-player cooperates i times.

An element of $\mathbf{w}$ is the sum of one or more of the marginal distributions $u^2_{h^2}$ for $h^2 \in H^2$. Specifically, let,

$$
H^2_i = \{h^2 : |a_C(h^2)| = i \quad \forall \quad h^2 \in H^2\},
$$

where

$$
a_C(h^2) = \{a^2_{-t} : a^2_{-t} = C \quad \forall \quad a^2_{-t} \in h^2\}.
$$

Then we define w_i as

$$w_i = \sum_{h \in H_i^2} v_h.$$

Please note that

$$\sum_{i=0}^{r} w_i = 1. \tag{67}$$

The cooperation rate of the reactive player is given by

$$\rho_{\mathbf{p}} = \sum_{i=0}^{n} r_i \cdot w_i. \tag{68}$$

The co-player can use any self-reactive-n strategy, and thus the co-player differentiates between when the last cooperation/defection occurred. However, we can still express the co-player's cooperation rate as a function of w_i. More specifically, the co-player's cooperation rate is

$$\rho_{\tilde{\mathbf{p}}} = \sum_{i=0}^{n} \frac{i}{n} \cdot w_i. \tag{69}$$

We will also define the self-reactive counting k-round repeat strategies. These strategies start by playing a sequence of cooperation in the first n moves until they reach a total of i cooperations, after which they defect for $n - i$ rounds. Thereafter, they repeat their a_{-n}^i move. We denote this set of strategies as $A = \{\mathbf{A^0}, \mathbf{A^1}, \dots, \mathbf{A^n}\}$.

The payoff of an alternating self-reactive-n against a counting-reactive-n $\mathbf{r}$ is given by

$$s_{\mathbf{A^i},\mathbf{r}} = b \cdot r_i - \frac{i}{n} \cdot c \;\; for \;\; i \in [0, n]. \tag{70}$$

The intuition behind Eq. (70) is that, in the long term, the strategies end up in a state where $\mathbf{A^i}$ has cooperated i times in the last n turns. Thus, the co-player will cooperate and provide the benefit b with a probability r_i, while in return, the alternating strategy has cooperated $\frac{i}{n}$ times and pays the cost.

With this, we have all the required tools to prove the following theorem.

Proof. ($\Rightarrow$) As discussed previously, a strategy can only be a Nash equilibrium if the payoff of the co-player does not exceed $(b - c)$. Therefore, for $\mathbf{p}$ to be a Nash equilibrium against each strategy in set A (for $i \in [0, n]$),

$$\begin{aligned} s_{\mathbf{A^i},\mathbf{r}} &\leq b - c \\ b \cdot r_i - \frac{i}{n} \cdot c &\leq b - c \\ r_i &\leq 1 - \frac{i}{n} \cdot \frac{c}{b}. \end{aligned}$$

Now, suppose player 2 has some strategy $\mathbf{m}$ and player 1 has a reactive-counting strategy such that $s_{\mathbf{m},\mathbf{p}} > b - c$. It follows that

$$\begin{aligned}
0 &\leq s_{\mathbf{m},\mathbf{p}} - (b-c) \\
&\overset{Eq.\ (??)}{=} b\rho_{\mathbf{p}} - c\rho_{\mathbf{m}} - (b-c) \\
&\overset{Eqs.\ (67),(68),(69)}{=} b\ \sum_{k=0}^{n} r_{n-k} \cdot u_{n-k} - c\ \sum_{k=0}^{n} \tfrac{n-k}{n} \cdot u_{n-k} - (b-c)\ \sum_{k=0}^{n} u_{n-k} \\
&\quad u_n\Big(b\,(r_n - 1)\Big) + \sum_{k=1}^{n} u_{n-k}\Big(b\ \sum_{k=1}^{n} r_{n-k} - c\ \sum_{k=0}^{n-1} \tfrac{n-k}{n} - (b-c)\Big)
\end{aligned} \tag{71}$$

This condition holds only if

$$\begin{aligned}
\Big(b\, r_{n-k} - c\,\frac{n-k}{n} - (b-c)\Big) < 0 \Rightarrow \\
b\,(r_{n-k} - 1) + (1 - \frac{n-k}{n})\, c < 0 \Rightarrow \\
r_{n-k} < 1 - \frac{n}{k}
\end{aligned}$$

$\square$

References

[1] Nowak, M. & Sigmund, K. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game. *Nature* **364**, 56–58 (1993).

[2] Hilbe, C., Chatterjee, K. & Nowak, M. A. Partners and rivals in direct reciprocity. *Nature human behaviour* **2**, 469–477 (2018).

[3] Axelrod, R. & Hamilton, W. D. The evolution of cooperation. *science* **211**, 1390–1396 (1981).

[4] Nowak, M. A. & Sigmund, K. Tit for tat in heterogeneous populations. *Nature* **355**, 250–253 (1992).

[5] Molander, P. The optimal level of generosity in a selfish, uncertain environment. *Journal of Conflict Resolution* **29**, 611–618 (1985).

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

[7] Hilbe, C., Wu, B., Traulsen, A. & Nowak, M. A. Cooperation and control in multiplayer social dilemmas. *Proceedings of the National Academy of Sciences USA* **111**, 16425–16430 (2014).

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

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

[10] Levinsky, R., Neyman, A. & Zeleny, M. Should i remember more than you? On the best response to factor-based strategies. *International Journal of Game Theory* **49**, 1105–1124 (2020).

[11] Stewart, A. J. & Plotkin, J. B. Small groups and long memories promote cooperation. *Scientific reports* **6**, 1–11 (2016).

[12] Ueda, M. Memory-two zero-determinant strategies in repeated games. *Royal Society open science* **8**, 202186 (2021).