Reactive strategies with longer memory

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

1 Model

1.1 The repeated prisoner's dilemma

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

$$
\begin{array}{cc}
 & \begin{array}{cc} C & D \end{array} \\
\begin{array}{c} C \\ D \end{array} & \left(\begin{array}{cc} R & S \\ T & P \end{array} \right).
\end{array}
\tag{1}
$$

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 yet each player's dominant action is to defect. For some of our results, we focus on a special case of the prisoner's dilemma that only depends on two parameters, the donation game. For this game, the payoff matrix takes the form

$$
\begin{array}{cc}
 & \begin{array}{cc} C & D \end{array} \\
\begin{array}{c} C \\ D \end{array} & \left(\begin{array}{cc} b - c & -c \\ b & 0 \end{array} \right).
\end{array}
\tag{3}
$$

Here b and c correspond to the benefit and the cost of cooperation, respectively. For this game to satisfy the conditions (2) of a prisoner's dilemma, we assume $b > c > 0$ throughout.

We assume players interact in this game for infinitely many rounds, and that future payoffs are not discounted. A strategy σ^i for player i thus needs to tell 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

average 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, 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. 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 [8].

1.2 Finite-memory strategies

Some important strategy sets. For this study, we focus on those strategies where a player's decision in any round is independent of events that happened more than n rounds ago. To define these strategies formally, we introduce some notation. An *n-history for player* $i \in \{1, 2\}$ is a string $\mathbf{h^i} = (a^i_{-n}, \ldots, a^i_{-1}) \in \{C, D\}^n$. 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.

Based on this notation, we define a *memory-n strategy* for player i as a tuple $\mathbf{m} = (m_{\mathbf{h}})_{\mathbf{h} \in H} \in [0, 1]^{2^{2n}}$. Here, 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 $m_{\mathbf{h}}$ is then 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

$$\mathcal{M}_n := \left\{ \mathbf{m} = (m_{\mathbf{h}})_{\mathbf{h} \in H} \in [0, 1]^{2^{2n}} \; \middle| \; \mathbf{h} \in H \right\}. \tag{5}$$

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 irrelevant for the long-run dynamics of almost all pairs of memory-n strategies, as we show further below. In the following, we 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. The corresponding 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. Perhaps the most well-known memory-1 strategy is Win-Stay Lose-Shift [5], $\mathbf{m} = (1, 0, 0, 1)$.

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*,

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

That is, such reactive-n strategies are independent of the focal player's own n-history. This space of reactive-

n strategies can be naturally identified with the space of all n-dimensional vectors $\mathbf{p} = (p_{\mathbf{h}^{-\mathbf{i}}})_{\mathbf{h}^{-\mathbf{i}} \in H^{-i}} \in [0,1]^{2^n}$. Here, each entry $p_{\mathbf{h}^{-\mathbf{i}}}$ 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 the case of $n=1$, for which $\mathbf{p} = (p_C, p_D)$. Examples of well-known reactive-1 strategies include Tit-for-Tat [2], $\mathbf{p} = (1,0)$, and Generous Tit-for-Tat [6], $\mathbf{p} = (1, p_D^*)$, where $p_D^* := \min\left\{1-(T-R)/(R-S),\ (R-P)/(T-P)\right\}$.

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

$$\mathcal{S}_n := \left\{ \mathbf{m} \in \mathcal{M}_n \ \middle|\ m_{(\mathbf{h}^{\mathbf{i}}, \mathbf{h}^{-\mathbf{i}})} = m_{(\mathbf{h}^{\mathbf{i}}, \tilde{\mathbf{h}}^{-\mathbf{i}})} \ \text{ for all } \mathbf{h}^{\mathbf{i}} \in H^i \text{ and } \mathbf{h}^{-\mathbf{i}}, \tilde{\mathbf{h}}^{-\mathbf{i}} \in H^{-i} \right\}. \tag{7}$$

That is, such strategies only depend on the focal player's own decisions during the last n rounds, whereas they are independent of the co-player's last decisions. Again, we can identify these self-reactive strategies with an n-dimensional vector, $\tilde{\mathbf{p}} = (\tilde{p}_{\mathbf{h}^{\mathbf{i}}})_{\mathbf{h}^{\mathbf{i}} \in H^i} \in [0,1]^{2^n}$. Each entry $\tilde{p}_{\mathbf{h}^{\mathbf{i}}}$ 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 $1 \leq k \leq n$. In any given round, a player with a *round-k-repeat strategy* $\tilde{\mathbf{p}}^{k-\mathrm{Rep}}$ chooses the same action as she did k rounds ago. Formally, the entries of $\tilde{\mathbf{p}}^{k-\mathrm{Rep}}$ are defined by

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

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. When it is convenient to represent the self-reactive repeat strategies as elements of the memory-n strategy space, we write $\mathbf{m}^{k-\mathrm{Rep}} \in [0,1]^{2^{2n}}$ instead of $\tilde{\mathbf{p}}^{k-\mathrm{Rep}} \in [0,1]^{2^n}$.

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

$$M_{\mathbf{h}, \tilde{\mathbf{h}}} = x^1 \cdot x^2 \tag{8}$$

where

$$x^i = \begin{cases} m_{(\mathbf{h}^{\mathbf{i}}, \mathbf{h}^{-\mathbf{i}})}^i & \text{if } \tilde{a}_{-1}^i = C, \quad \text{and} \quad \tilde{a}_{-t}^i = a_{-t+1}^i \quad \text{for } t \in \{2, \ldots, n\} \\ 1 - m_{(\mathbf{h}^{\mathbf{i}}, \mathbf{h}^{-\mathbf{i}})}^i & \text{if } \tilde{a}_{-1}^i = D, \quad \text{and} \quad \tilde{a}_{-t}^i = a_{-t+1}^i \quad \text{for } t \in \{2, \ldots, n\} \\ 0 & \text{if } \tilde{a}_{-t}^i \neq a_{-t+1}^i \quad \text{for some } t \in \{2, \ldots, n\}. \end{cases} \tag{9}$$

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, if $\mathbf{v}(t) = \big(v_{\mathbf{h}}(t)\big)_{\mathbf{h} \in H}$ is the probability distribution of observing state $\mathbf{h}$ after round $t \geq n$, 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 all conditional cooperation probabilities $m_{\mathbf{h}}^i$ are strictly between zero and one). In that case, it follows by the theorem of Perron and Frobenius that $\mathbf{v}(t)$ converges to some $\mathbf{v}$ as $t \to \infty$. In particular, it also follows that the respective means converge,

$$\mathbf{v} = \lim_{\tau \to \infty} \frac{1}{\tau} \sum_{t=n}^{n+\tau-1} \mathbf{v}(t). \tag{10}$$

This respective limiting distribution 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 $\mathbf{g^i}(k) = \big(g_{\mathbf{h}}^i(k) \big)_{\mathbf{h} \in H}$, which captures player i's payoffs $k \leq n$ rounds ago, according to the n-history $\mathbf{h}$,

$$g_{\mathbf{h}}^i(k) = \begin{cases} R & \text{if } a_{-k}^i = C \text{ and } a_{-k}^{-i} = C \\ S & \text{if } a_{-k}^i = C \text{ and } a_{-k}^{-i} = D \\ T & \text{if } a_{-k}^i = D \text{ and } a_{-k}^{-i} = C \\ P & \text{if } a_{-k}^i = D \text{ and } a_{-k}^{-i} = D. \end{cases} \tag{11}$$

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

$$\pi_{\mathbf{m^1},\mathbf{m^2}}^i(t) = \big\langle \mathbf{v}(t), \mathbf{g^i}(1) \big\rangle = \sum_{\mathbf{h} \in H} v_{\mathbf{h}}(t) \cdot g_{\mathbf{h}}^i(1). \tag{12}$$

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

$$\pi^i(\mathbf{m^1}, \mathbf{m^2}) \overset{(4)}{=} \lim_{\tau \to \infty} \frac{1}{\tau} \sum_{t=1}^{\tau} \pi_{\mathbf{m^1},\mathbf{m^2}}^i(t) = \lim_{\tau \to \infty} \frac{1}{\tau} \sum_{t=n}^{n+\tau-1} \pi_{\mathbf{m^1},\mathbf{m^2}}^i(t)$$

$$\overset{(12)}{=} \lim_{\tau \to \infty} \frac{1}{\tau} \sum_{t=n}^{n+\tau-1} \big\langle \mathbf{v}(t), \mathbf{g^i}(1) \big\rangle = \Big\langle \lim_{\tau \to \infty} \frac{1}{\tau} \sum_{t=n}^{n+\tau-1} \mathbf{v}(t), \mathbf{g^i}(t) \Big\rangle \tag{13}$$

$$\overset{(10)}{=} \big\langle \mathbf{v}, \mathbf{g^i}(1) \big\rangle.$$

That is, given we know the invariant distribution $\mathbf{v}$ that captures the game's long-run dynamics, it is straight-forward to compute payoffs by taking the scalar product with the vector $\mathbf{g^i}(1)$. With a similar approach as in Eq. (13), one can also show

$$\left\langle \mathbf{v}, \mathbf{g^i}(1) \right\rangle = \left\langle \mathbf{v}, \mathbf{g^i}(2) \right\rangle = \ldots = \left\langle \mathbf{v}, \mathbf{g^i}(n) \right\rangle. \tag{14}$$

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.

An Extension of Akin's Lemma. The Markov chain approach allows us to analyze games when both players adopt memory-n strategies. But even if there is only one player who 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 [1]: if player 1 adopts a memory-1 strategy $\mathbf{m}$ against an arbitrary opponent, and if the time average $\mathbf{v}$ of the resulting game as defined by Eq. (10) exists, then

$$\left\langle \mathbf{v}, \mathbf{m} - \mathbf{m}^{1-\mathrm{Rep}} \right\rangle = 0. \tag{15}$$

That is, the limiting distribution $\mathbf{v}$ needs to be orthogonal to the vector $\mathbf{m} - \mathbf{m}^{1-\mathrm{Rep}}$. This result has been termed *Akin's lemma* [3]. With similar methods as in Ref. [1], one can generalize it to the context of memory-n strategies.

Lemma 1.1 (A generalized version of Akin's Lemma)
Let player 1 *use a memory-n strategy, and let player* 2 *use any arbitrary strategy. For the resulting game, let* $\mathbf{v}(t) = \left(v_\mathbf{h}(t) \right)_{\mathbf{h} \in H}$ *denote the probability distribution of observing each possible n-history* $\mathbf{h} \in H$ *after some given round $t \geq n$. Moreover, suppose the respective time average* $\mathbf{v}$ *according to Eq.* (10) *exists. Then for each k with $1 \leq k \leq n$, we obtain*

$$\left\langle \mathbf{v}, \mathbf{m} - \mathbf{m}^{k-Rep} \right\rangle = 0. \tag{16}$$

All proofs are presented in the Appendix. Here we provide an intuition: The expression $\langle \mathbf{v}, \mathbf{m} \rangle = \sum_\mathbf{h} v_\mathbf{h} m_\mathbf{h}$ may be interpreted as player 1's average cooperation rate across 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 $\langle \mathbf{v}, \mathbf{m}^{k-\mathrm{Rep}} \rangle$. According to Eq. (16), both interpretations are equivalent.

2 Characterizing the partner strategies among the reactive-n strategies

2.1 Partner strategies

Let's provide definitions for some additional terms that will be used in this manuscript.

Definition 2.1 (Nash Strategies)
A strategy $\mathbf{m^i}$, is a Nash strategy if,

$$s_{\mathbf{m^{-i}},\mathbf{m}} \leq s_{\mathbf{m^i},\mathbf{m^i}} \ \forall \ \mathbf{m^{-i}}. \tag{17}$$

Note that for a memory-n strategy, we only need to check the above condition for all mutant strategies of memory n. This is a result from the work of [7].

Definition 2.2 (Nice Strategies)
A player's strategy is nice, if the player is never the first to defect. A nice strategy against itself receives the mutual cooperation payoff, $(b - c)$.

Definition 2.3 (Partner Strategies)
A partner strategy is a strategy which is both nice and Nash.

Partners strategies are of interest because they are strategies that strive to achieve the mutual cooperation payoff of $(b - c)$ with their co-player. However, if the co-player doesn't cooperate, they are prepared to penalize them with lower payoffs. Partner strategies, by definition, are best responses to themselves [4]. All partner strategies are Nash strategies, but not all Nash strategies are partner strategies.

3 Tit For Tat and Generous Tit For Tat across All Memory Length

Building upon Lemma 1.1, we can develop a theory of zero-determinant strategies within the class of memory-n 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 $\mathbf{m^i}$ can be written as

$$\mathbf{m^i} = \alpha \mathbf{S}_{k_1}^i + \beta \mathbf{S}_{k_2}^{-i} + \gamma \mathbf{1} + \mathbf{m}^{k_3 - \text{Rep}}, \tag{18}$$

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

$$0 = \mathbf{v} \cdot (\mathbf{m^i} - \mathbf{m}^{k_3 - \text{Rep}}) = \mathbf{v} \cdot (\alpha \mathbf{S}_{k_1}^i + \beta \mathbf{S}_{k_2}^{-i} + \gamma \mathbf{1}) = \alpha s_{\mathbf{m^i},\mathbf{m^{-i}}} + \beta s_{\mathbf{m^{-i}},\mathbf{m^i}} + \gamma. \tag{19}$$

That is, payoffs satisfy a linear relationship. Thus, $\mathbf{m^i}$ is a zero-determinant strategy.

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 (18) yields the strategy

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

That is, this strategy implements Tit-for-Tat (for $k = 1$) or delayed versions thereof (for $k > 1$). These strategies are partner strategies that also satisfy a stronger relationship. According to Eq. (19), they enforce the payoff relationship $s_{\mathbf{m^i},\mathbf{m^{-i}}} = s_{\mathbf{m^{-i}},\mathbf{m^i}}$.

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. (18) yields the strategy

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

That is, the generated strategy is GTFT (if $k = 1$), or delayed versions thereof (for $k > 1$). By Eq. (19), the enforced payoff relationship is $s_{\mathbf{m^{-i}},\mathbf{m^i}} = b - c$. In particular, these strategies are partner strategies.

4 Sufficiency of Self Reactive Strategies

[7] 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 history that one player does not observe, then the co-player gains nothing by considering the history not shared with the reactive player.

Lemma 4.1

Let $\mathbf{p}$ be a reactive$-n$ strategy for player 1. Then, for any memory$-n$ strategy $\mathbf{m}$ used by player 2, player 1's score is exactly the same as if 2 had played a specific self-reactive memory-n strategy $\tilde{\mathbf{p}}$.

Proof. ... □

NG: pip install christian_hilbe

NG: Could you please try to come up with the proof?

Several results arise from Lemma 4.1. For example, assume that player 1 is using a reactive-n strategy, and player 2 is using a memory-n strategy. If we wanted to verify that player 1's strategy is Nash, we would have to show that condition (17) holds for all memory-n mutant strategies. However, now we can only consider self-reactive-n mutant strategies for player 2.

From here on, when we are considering the case where player 1 is using a reactive-n strategy, we will assume that player 2 is using a self-reactive-n strategy.

A more efficient way to calculate payoffs. Assume that player 1 is playing a reactive-n strategy $\mathbf{p}$ and the co-player is playing an arbitrary strategy $\tilde{\mathbf{p}}$. To calculate the long-term payoffs of the two play-

7

ers, we need to compute the stationary distribution of the Markov chain M and then calculate the payoffs using Eq. (**??**). However, since the co-player is playing a self-reactive strategy, their actions only rely on their own actions. This can be modeled as a Markov process with H^2 states and a transition matrix $\tilde{M}$. Let $h^2 = ((a^2_{-n}, \ldots, a^2_{-1}))$ be the state in the current round. The probability that in the next turn $\tilde{h}^2 = ((\tilde{a}^2_{-n}, \ldots, \tilde{a}^2_{-1}))$ is observed is given by,

$$\tilde{M}_{h^2,\tilde{h}^2} = \begin{cases} \tilde{p}_{h^2} & \text{if } \tilde{\alpha}^2_{-1} = C \text{ and } \tilde{\alpha}^2_{-t} = \alpha^2_{-t+1} \text{ for all other } \tilde{\alpha}^2_{-t} \\ 1 - \tilde{p}_{h^2} & \text{if } \tilde{\alpha}^2_{-1} = D \text{ and } \tilde{\alpha}^2_{-t} = \alpha^2_{-t+1} \text{ for all other } \tilde{\alpha}^2_{-t} \\ 0 & \text{if } \tilde{\alpha}^2_{-t} \neq \alpha^2_{-t+1} \text{ for some } 2 \leq t \leq n, \end{cases} \tag{20}$$

and let $\tilde{\mathbf{v}}$ be a stationary distribution of the Markov chain $\tilde{M}$. Now player 1's payoff ($s_{\tilde{\mathbf{p}},\mathbf{p}}$) for the general Prisoner's Dilemma is given by,

$$s_{\tilde{\mathbf{p}},\mathbf{p}} = a_R \cdot R + a_S \cdot S + a_T \cdot T + a_P \cdot P, \quad where$$

$$\begin{aligned} a_R &= \sum_{h^2 \in H^2} \tilde{u}_{h^2} \cdot p_{h^2} \cdot \tilde{p}_{h^2}, \\ a_S &= \sum_{h^2 \in H^2} \tilde{u}_{h^2} \cdot (1 - p_{h^2}) \cdot \tilde{p}_{h^2}, \\ a_T &= \sum_{h^2 \in H^2} \tilde{u}_{h^2} \cdot p_{h^2} \cdot (1 - \tilde{p}_{h^2}), \\ a_P &= \sum_{h^2 \in H^2} \tilde{u}_{h^2} \cdot (1 - p_{h^2}) \cdot (1 - \tilde{p}_{h^2}). \end{aligned}$$

Player 2's payoff is calculated in a similar way. In the case of the simple donation games, it is sufficient to define the payoffs of the two players based on their cooperation rates. More specifically, we can define the payoffs of the two players as,

$$\begin{aligned} s_{\mathbf{p},\tilde{\mathbf{p}}} &= b\,\rho_{\tilde{\mathbf{p}}} - c\,\rho_{\mathbf{p}} \\ s_{\tilde{\mathbf{p}},\mathbf{q}} &= b\,\rho_{\mathbf{p}} - c\,\rho_{\tilde{\mathbf{p}}}. \end{aligned} \tag{21}$$

where,

$$\begin{aligned} \rho_{\tilde{\mathbf{p}}} &= \tilde{\mathbf{v}} \cdot \tilde{\mathbf{p}}^{1-\text{Rep}}, \quad and \\ \rho_{\mathbf{p}} &= \sum_{h^2 \in H^2} \tilde{v}^{h^2} \cdot p_{h^2}. \end{aligned}$$

Example. Let's consider an example in the case of $n = 1$ to demonstrate the result. Assume that player 1 is playing the reactive-1 strategy $\mathbf{p} = \left(\frac{1}{2}, \frac{1}{3}, \frac{1}{2}, \frac{1}{3}\right)$, and player 2 is playing the self-reactive-1 strategy $\tilde{\mathbf{p}} = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{2}, \frac{1}{2}\right)$. We construct the transition matrix M based on Eq. (8) as follows:

$$M = \begin{bmatrix} \frac{1}{8} & \frac{3}{8} & \frac{1}{8} & \frac{3}{8} \\[4pt] \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} \\[4pt] \frac{1}{8} & \frac{3}{8} & \frac{1}{8} & \frac{3}{8} \\[4pt] \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$$

The stationary distribution is given by $\mathbf{v} = \left(\frac{4}{25}, \frac{6}{25}, \frac{6}{25}, \frac{9}{25} \right)$. The payoffs for players 1 and 2, based on Eq. (**??**), are equal and are equal to $s_{\mathbf{p},\tilde{\mathbf{p}}} = s_{\tilde{\mathbf{p}},\mathbf{q}} = \frac{2}{5}(b - c)$.

Now, we repeat the exercise, but this time we construct the transition matrix $\tilde{M}$ based on Eq. (20):

$$\tilde{M} = \begin{bmatrix} \frac{1}{4} & \frac{3}{4} \\[4pt] \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

The stationary distribution is given by $\tilde{\mathbf{v}} = \left(\frac{2}{5}, \frac{3}{5} \right)$, and the cooperation rates $\rho_{\tilde{\mathbf{p}}} = \frac{2}{5}$ and $\rho_{\mathbf{p}} = \frac{2}{5} \times \frac{1}{2} + \frac{3}{5} \times \frac{1}{3} = \frac{2}{5}$. Thus, based on Equation (**??**), the payoffs are $s_{\mathbf{p},\tilde{\mathbf{p}}} = s_{\tilde{\mathbf{p}},\mathbf{q}} = \frac{2}{5}(b - c)$.

5 An algorithm for Nash Equilibria amongst Reactive Strategies

To predict which reactive-n strategies are partner strategies, we must characterize which nice reactive-n strategies are Nash equilibria. Determining whether a given strategy, $\mathbf{p}$, is a Nash equilibrium is not straightforward. In principle, this would involve comparing the payoff of $\mathbf{p}$ to the payoff of all possible self-reactive strategies, and there are infinitely many self-reactive strategies. However, here we demonstrate that we can reduce the space of mutant strategies to a finite set.

Initially, we will demonstrate that the payoff of a self-reactive strategy against a reactive strategy is linear in the self-reactive strategy entries. Secondly, we will show that because the payoff is linear, the best response self-reactive strategy to the reactive strategy is a pure self-reactive strategy.

Lemma 5.1

A reactive-n strategy $\mathbf{p}$ for player 1, is a Nash strategy if, and only if, no pure self-reactive-n strategy can achieve a higher payoff against itself.

Lemma 5.1 implies that the space of strategies we need to check against is even more constrained in the case of reactive strategies. This has a huge implication on the computational complexity of finding Nash strategies.

NG: Christian could you please take over this section? We have already discussed the proofs.

6 Reactive Partner Strategies

In this section, we will characterize the reactive-n partner strategies. We will present a series of results, mainly proven based on Lemma 5.1. Interestingly, several of our results can also be demonstrated through

an independent proof relying on the generalization of Akin (Lemma 1.1). We will provide both proofs in Appendices A and B.

6.1 Reactive-2 Partner Strategies

In this section, we focus on the case of $n = 2$. reactive-2 strategies are denoted as a vector $\mathbf{p} = (p_{CC}, p_{CD}, p_{DC}, p_{DD})$ where p_{CC} is the probability of cooperating in this turn when the co-player cooperated in the last 2 turns, p_{CD} is the probability of cooperating given that the co-player cooperated in the second to last turn and defected in the last, and so forth. A nice reactive-2 strategy is represented by the vector $\mathbf{p} = (1, p_{CD}, p_{DC}, p_{DD})$.

Theorem 6.1 ("Reactive-2 Partner Strategies")
A reactive-2 strategy $\mathbf{p}$, is a partner strategy if and only if, the strategy entries satisfy the conditions:

$$p_{CC} = 1, \qquad \frac{p_{CD} + p_{DC}}{2} < 1 - \frac{1}{2} \cdot \frac{c}{b} \quad and \quad p_{DD} \leq 1 - \frac{c}{b}. \tag{22}$$

For proofs see Appendix sections A.1 and B.2.

6.2 Reactive-3 Partner Strategies

In this section, we focus on the case of $n = 3$. reactive-3 strategies are denoted as a vector

$$\mathbf{p} = (p_{CCC}, p_{CCD}, p_{CDC}, p_{CDD}, p_{DCC}, p_{DCD}, p_{DDC}, p_{DDD})$$

where p_{CCC} is the probability of cooperating in round t when the co-player cooperates in the last 3 rounds, 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, and so forth. A nice reactive-3 strategy has $p_{CCC} = 1$.

Theorem 6.2 ("Reactive-3 Partner Strategies")
A reactive-3 strategy $\mathbf{p}$, is a partner strategy if and only if, the strategy entries satisfy the conditions:

$$
\begin{aligned}
p_{CCC} &= 1 \\
\frac{p_{CDC} + p_{DCD}}{2} &\leq 1 - \frac{1}{2} \cdot \frac{c}{b} \\
\frac{p_{CCD} + p_{CDC} + p_{DCC}}{3} &\leq 1 - \frac{1}{3} \cdot \frac{c}{b} \\
\frac{p_{CDD} + p_{DCD} + p_{DDC}}{3} &\leq 1 - \frac{2}{3} \cdot \frac{c}{b} \\
\frac{p_{CCD} + p_{CDD} + p_{DCC} + p_{DDC}}{4} &\leq 1 - \frac{1}{2} \cdot \frac{c}{b} \\
p_{DDD} &\leq 1 - \frac{c}{b}
\end{aligned}
\tag{23}
$$

For proofs see Appendix sections A.2 and B.3.

6.3 Reactive Counting Strategies

A special case of reactive strategies is reactive counting strategies. These are strategies that respond to the co-player's actions, but they do not distinguish between when cooperations/defections occurred; they solely consider the count of cooperations in the last n turns. A reactive-n counting strategy is represented by a vector $\mathbf{r} = (r_i)_{i \in \{n, n-1, \ldots, 0\}}$, where the entry r_i indicates the probability of cooperating given that the co-player cooperated i times in the last n turns.

Reactive-2 counting strategies are denoted by the vector $\mathbf{r} = (r_2, r_1, r_0)$. We can characterise partner strategies among the reactive-2 counting strategies by setting $r_2 = 1$, and $p_{CD} = p_{DC} = r_1$ and $p_{DD} = r_0$ in conditions (22). This gives us the following result.

Corollary 6.2.1
A reactive-2 counting strategy $\mathbf{r} = (r_2, r_1, r_0)$ is a partner strategy if and only if,

$$r_2 = 1, \qquad r_1 < 1 - \frac{1}{2} \cdot \frac{c}{b} \qquad and \qquad r_0 < 1 - \frac{c}{b}. \tag{24}$$

Reactive-3 counting strategies are denoted by the vector $\mathbf{r} = (r_3, r_2, r_1, r_0)$. We can characterise partner strategies among reactive-3 counting strategies by setting $r_3 = 1$, and $p_{CCD} = p_{CDC} = p_{DCC} = r_2$, $p_{DCD} = p_{DDC} = p_{CDD} = r_1$ and $p_{DDD} = r_0$ in conditions (23). This gives us the following result.

Corollary 6.2.2
A reactive-3 counting strategy $\mathbf{r} = (r_3, r_2, r_1, r_0)$ is a partner strategy if and only if,

$$r_3 = 1 \qquad r_2 < 1 - \frac{1}{3} \cdot \frac{c}{b}, \quad r_1 < 1 - \frac{2}{3} \cdot \frac{c}{b} \qquad and \qquad r_0 < 1 - \frac{c}{b}. \tag{25}$$

In the case of counting reactive strategies, we generalize to the case of n.

Theorem 6.3 ("Reactive-Counting Partner Strategies")
A reactive-n counting strategy $\mathbf{r} = (r_i)_{i \in \{n, n-1, \ldots, 0\}}$, is a partner strategy if and only if:

$$r_n = 1 \qquad and \qquad r_{n-k} < 1 - \frac{k}{n} \cdot \frac{c}{b}, \text{ for } k \in \{1, 2, \ldots, n\}. \tag{26}$$

For proof see Appendix section C.

6.4 General Prisoner's Dilemma

So far we have focused on a special case of the Prisoner's Dilemma, the donation game. In this section we show that the results of Sections 6.1 and 6.2 can be generalized for the iterated Prisoner's Dilemma. For the case of reactive-2 strategies.

Theorem 6.4
A reactive-2 strategy $\mathbf{p}$, is a partner strategy if and only if, the strategy entries satisfy the conditions:

$$p_{CC} = 1,$$
$$(T - P)\, p_{DD} < R - P,$$
$$(R - S)\, (p_{CD} + p_{DC}) < 3R - 2S - T,$$
$$(T - P)\, p_{DC} + (R - S)\, p_{CD} < 2R - S - P,$$
$$(T - P)\, (p_{CD} + p_{DC}) + (R - S)\, p_{DD} < 3R + S - 2\, P,$$
$$(T - P)\, p_{CD} + (R - S)\, (p_{CD} + p_{DD}) < 4\, R - 2\, S - P - T.$$

For proof see Appendix section A.3.

For the case of reactive-3 strategies.

Theorem 6.5

A reactive-3 strategy **p***, is a partner strategy if and only if, the strategy entries satisfy the conditions:*

$$p_{CCC} = 1,$$

$$(T-P)\left(p_{CDD}+p_{DCD}+p_{DDC}\right)+(R-S)\,p_{DDD} < 4R-3P-S$$

$$(T-P)\,p_{CDC}+(R-S)\,p_{DCD} < 2R-P-S$$

$$(T-P)\,p_{DDD} < R-P$$

$$(T-P)\left(p_{CCD}+p_{CDD}+p_{DDC}\right)+(R-S)\left(p_{CDC}+p_{DCC}+p_{DCD}+p_{DDD}\right) < 8R-3P-4S-T$$

$$(T-P)\,p_{DCC}+(R-S)\left(p_{CCD}+p_{CDC}\right) < 3R-P-2S$$

$$(T-P)\left(p_{CCD}+p_{DCC}+p_{DDC}\right)+(R-S)\left(p_{CDC}+p_{CDD}+p_{DCD}\right) < 6R-3P-3S$$

$$(T-P)\left(p_{CCD}+p_{DDC}\right)+(R-S)\left(p_{CDC}+p_{CDD}+p_{DCC}+p_{DCD}\right) < 7R-2P-4S-T$$

$$(T-P)\left(p_{CCD}+p_{CDD}+p_{DCC}\right)+(R-S)\left(p_{DDC}+p_{DDD}\right) < 5R-3P-2S$$

$$(T-P)\left(p_{DCD}+p_{DDC}\right)+(R-S)\,p_{CDD} < 3R-2P-S$$

$$(T-P)\,p_{CCD}+(R-S)\left(p_{CDD}+p_{DCC}+p_{DDC}\right) < 5R-P-3S-T$$

$$(T-P)\left(p_{CCD}+p_{DCC}\right)+(R-S)\left(p_{CDD}+p_{DDC}\right) < 4R-2P-2S$$

$$(T-P)\left(p_{CDC}+p_{DCD}\right)+(R-S)\left(p_{CCD}+p_{CDD}+p_{DCC}+p_{DDC}\right) < 7R-2P-4S-T$$

$$(T-P)\left(p_{CDC}+p_{CDD}+p_{DCD}\right)+(R-S)\left(p_{CCD}+p_{DCC}+p_{DDC}+p_{DDD}\right) < 8R-3P-4S-T$$

$$(T-P)\left(p_{CDC}+p_{DCC}+p_{DCD}\right)+(R-S)\left(p_{CCD}+p_{CDD}+p_{DDC}\right) < 6R-3P-3S$$

$$(T-P)\left(p_{CCD}+p_{CDD}+p_{DCC}+p_{DDC}\right)+(R-S)\left(p_{CDC}+p_{DCD}+p_{DDD}\right) < 7R-4P-3S$$

$$(R-S)\left(p_{CCD}+p_{CDC}+p_{DCC}\right) < 4R-3S-T$$

$$(T-P)\left(p_{CCD}+p_{CDD}\right)+(R-S)\left(p_{DCC}+p_{DDC}+p_{DDD}\right) < 6R-2P-3S-T$$

$$(T-P)\left(p_{CDC}+p_{CDD}+p_{DCC}+p_{DCD}\right)+(R-S)\left(p_{CCD}+p_{DDC}+p_{DDD}\right) < 7R-4P-3S$$

For proof see Appendix section A.4.

13

Proof of Lemma 1.1. 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 nth 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,

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

Equivalently,

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

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

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

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

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

This implies,

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

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 $1 < k \leq n$.

$\square$

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

A.1 Proof of Theorem 6.1

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:

$$s_{\tilde{\mathbf{p}}^i, \mathbf{p}} = b \cdot p_{DD} \qquad for \ i \in \quad \{0, 2, 4, 6, 8, 10, 12, 14\}$$

$$s_{\tilde{\mathbf{p}}^i, \mathbf{p}} = \frac{b \cdot (p_{CD} + p_{DC} + p_{DD})}{3} - \frac{1}{3} \cdot c \qquad for \ i \in \quad \{1, 9\}$$

$$s_{\tilde{\mathbf{p}}^i, \mathbf{p}} = \frac{b \cdot (p_{CD} + p_{DC} + p_{DD} + 1)}{4} - \frac{1}{2} \cdot c \qquad for \ i \in \quad \{3\}$$

$$s_{\tilde{\mathbf{p}}^i, \mathbf{p}} = \frac{b \cdot (p_{CD} + p_{DC})}{2} - \frac{1}{2} \cdot c \qquad for \ i \in \quad \{4, 5, 12, 13\}$$

$$s_{\tilde{\mathbf{p}}^i, \mathbf{p}} = \frac{b \cdot (p_{CD} + p_{DC} + 1)}{3} - \frac{2}{3} \cdot c \qquad for \ i \in \quad \{6, 7\}$$

$$s_{\tilde{\mathbf{p}}^i, \mathbf{p}} = b - c \qquad for \ i \in \quad \{8, 9, 10, 11, 12, 13, 14, 15\}$$

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,

$$p_{DD} \leq 1 - \frac{c}{b} \tag{27}$$

$$\frac{p_{CD} + p_{DC}}{2} \leq 1 - \frac{1}{2} \cdot \frac{c}{b} \tag{28}$$

$$\frac{p_{CD} + p_{DC} + p_{DD}}{3} \leq 1 - \frac{2}{3} \cdot \frac{c}{b} \tag{29}$$

Notice that only conditions (27) and (28) are necessary.

$\square$

A.2 Proof of Theorem 6.2

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,\dots,250,252,254\} \\[4pt]
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= \frac{b\cdot(p_{CDD}+p_{DCD}+p_{DDC}+p_{DDD})}{4} - \frac{1}{4}\cdot c & for\ i \in\ & \{1,9,33,41,65,73,97,105,129,137,161, \\ & & & \ \ 169,193,201,225,233\} \\[4pt]
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= \frac{b\cdot(p_{CCD}+p_{CDD}+p_{DCC}+p_{DDC}+p_{DDD})}{5} - \frac{2}{5}\cdot c & for\ i \in\ & \{3,7,35,39,131,135,163,167\} \\[4pt]
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= \frac{b\cdot(p_{CDC}+p_{DCD})}{2} - \frac{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\} \\[4pt]
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= \frac{b\cdot(p_{CCD}+p_{CDD}+p_{DCC}+p_{DDC}+p_{DDD}+1)}{6} - \frac{1}{2}\cdot c & for\ i \in\ & \{11,15,43,47\} \\[4pt]
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= \frac{b\cdot(p_{CDD}+p_{DCD}+p_{DDC})}{3} - \frac{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\} \\[4pt]
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= \frac{b\cdot(p_{CCD}+p_{CDD}+p_{DCC}+p_{DDC})}{4} - \frac{1}{2}\cdot c & for\ i \in\ & \{18,19,22,23,50,51,54,55,146,147, \\ & & & \ \ 150,151,178,179,182,183\} \\[4pt]
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= \frac{b\cdot(p_{CCD}+p_{CDD}+p_{DCC}+p_{DDC}+1)}{5} - \frac{3}{5}\cdot c & for\ i \in\ & \{26,27,30,31,58,59,62,63\} \\[4pt]
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= \frac{b\cdot(p_{CCD}+p_{CDC}+p_{CDD}+p_{DCC}+p_{DCD}+p_{DDC}+p_{DDD})}{7} - \frac{3}{7}\cdot c & for\ i \in\ & \{37,67,165,195\} \\[4pt]
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= \frac{b\cdot(p_{CCD}+p_{CDC}+p_{CDD}+p_{DCC}+p_{DCD}+p_{DDC}+p_{DDD}+1)}{8} - \frac{1}{2}\cdot c & for\ i \in\ & \{45,75\} \\[4pt]
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= \frac{b\cdot(p_{CCD}+p_{CDC}+p_{CDD}+p_{DCC}+p_{DCD}+p_{DDC})}{6} - \frac{1}{2}\cdot c & for\ i \in\ & \{52,53,82,83,180,181,210,211\} \\[4pt]
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= \frac{b\cdot(p_{CCD}+p_{CDC}+p_{CDD}+p_{DCC}+p_{DCD}+p_{DDC}+1)}{7} - \frac{4}{7}\cdot c & for\ i \in\ & \{60,61,90,91\} \\[4pt]
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= \frac{b\cdot(p_{CCD}+p_{CDC}+p_{DCC})}{3} - \frac{2}{3}\cdot c & for\ i \in\ & \{96-103,112-119,224-231,240-247\} \\[4pt]
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= \frac{b\cdot(p_{CCD}+p_{CDC}+p_{DCC}+1)}{4} - \frac{3}{4}\cdot c & for\ i \in\ & \{104-111,120-127\} \\[4pt]
s_{\tilde{\mathbf{p}}^i,\mathbf{p}} &= (b-c) & for\ i \in\ & \{128,129,130,\dots,255\}
\end{aligned}
\tag{30}
$$

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

$$p_{DDD} \leq 1 - \frac{c}{b}, \qquad \frac{p_{CDC} + p_{DCD}}{2} \leq 1 - \frac{1}{2} \cdot \frac{c}{b}, \qquad \frac{p_{CDD} + p_{DCD} + p_{DDC}}{3} \leq 1 - \frac{2}{3} \cdot \frac{c}{b}, \qquad (31)$$

$$\frac{p_{CCD} + p_{CDC} + p_{DCC}}{3} \leq 1 - \frac{1}{3} \cdot \frac{c}{b}, \qquad \frac{p_{CCD} + p_{CDD} + p_{DCC} + p_{DDC}}{4} \leq 1 - \frac{1}{2} \cdot \frac{c}{b} \qquad (32)$$

$$\frac{p_{CDD} + p_{DCD} + p_{DDC} + p_{DDD}}{4} \leq 1 - \frac{3}{4} \cdot \frac{c}{b}, \qquad (33)$$

$$\frac{p_{CCD} + p_{CDC} + p_{CDD} + p_{DCC} + p_{DCD} + p_{DDC} + p_{DDD}}{7} \leq 1 - \frac{4}{7} \cdot \frac{c}{b}, \qquad (34)$$

$$\frac{p_{CCD} + p_{CDD} + p_{DCC} + p_{DDC} + p_{DDD}}{5} \leq 1 - \frac{3}{5} \cdot \frac{c}{b}, \qquad (35)$$

$$\frac{p_{CCD} + p_{CDC} + p_{CDD} + p_{DCC} + p_{DCD} + p_{DDC}}{6} \leq 1 - \frac{1}{2} \cdot \frac{c}{b} \qquad (36)$$

Notice that only the conditions of Eq. (31) and (32) are necessary. The remaining conditions can be derived from the sums of conditions in Eq. (31) and (32). □

A.3 Proof of Theorem 6.4

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. (4). This gives the following payoff expressions:

$$s_{\tilde{\mathbf{p}}^i, \mathbf{p}} = P(1 - p_{DD}) + T p_{DD} \qquad for \ i \in \{0, 2, 4, 6, 8, 10, 12, 14\}$$

$$s_{\tilde{\mathbf{p}}^i, \mathbf{p}} = \frac{-P(p_{CD} + p_{DC} - 2) + R p_{DD} - S(p_{DD} - 1) + T(p_{CD} + p_{DC})}{3} \qquad for \ i \in \{1, 9\}$$

$$s_{\tilde{\mathbf{p}}^i, \mathbf{p}} = \frac{P(1 - p_{CD}) + R(p_{DC} + p_{DD}) - S(p_{DC} + p_{DD} - 2) + T(p_{CD} + 1)}{4} \qquad for \ i \in \{3\}$$

$$s_{\tilde{\mathbf{p}}^i, \mathbf{p}} = \frac{P(1 - p_{DC}) + R p_{CD} - S(p_{CD} - 1) + T p_{DC}}{2} \qquad for \ i \in \{4, 5, 12, 13\}$$

$$s_{\tilde{\mathbf{p}}^i, \mathbf{p}} = \frac{R(p_{CD} + p_{DC}) - S(p_{CD} + p_{DC} - 2) + T}{3} \qquad for \ i \in \{6, 7\}$$

$$s_{\tilde{\mathbf{p}}^i, \mathbf{p}} = R \qquad for \ i \in \{8, 9, 10, 11, 12, 13, 14, 15\}$$

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

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

A.4 Proof of Theorem 6.5

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,

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

$$129, 137, 161, 169, 193, 20\ldots$$
$$225, 233\}$$

$$s_{\tilde{\mathbf{p}}^i, \mathbf{p}} = \frac{(T-P)\,p_{CDC}+P+(R-S)\,p_{DCD}+S}{2} \qquad for \ i \in \ \{4-7, 12-15, 20-23,$$

$$28-31, 68-71, 76-79,$$
$$84-87, 92-95, 132-135\ldots$$
$$140-143, 148-151, 156\ldots$$
$$196-199, 204-207, 212-\ldots$$
$$220-223\}$$

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

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

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

$$224-231, 240-247\}$$

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

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

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

$$s_{\tilde{\mathbf{p}}^i, \mathbf{p}} = \frac{(T-P)(p_{DCD}+p_{DDC})+2\,P+(R-S)\,p_{CDD}+S}{3} \qquad for \ i \in \ \{16, 17, 24, 25, 48, 49, 56,$$

$$57, 80, 81, 88, 89, 112, 113\ldots$$
$$120, 121, 144, 145, 152, 15\ldots$$
$$176, 177, 184, 185, 208, 20\ldots$$
$$216, 217, 240, 241, 248, 24\ldots$$

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

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

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

$$146, 147, 150, 151, 178, 17\ldots$$
$$182, 183\}$$

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

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

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

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

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

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

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

19

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:

$$
v_{h^2}^2 = \sum_{h^1 \in H^1} v_{(h^1, h^2)}. \tag{37}
$$

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,

$$\sum_{h \in H^2} v_h^2 = 1. \tag{38}$$

Let $\mathbf{p^{k-Rep}}$ be a reactive round-k-repeat strategy. Then the cooperation rate of player 2, denoted as $\rho_{\tilde{\mathbf{p}}}$, and based on Lemma 1.1 is given by,

$$\rho_{\tilde{\mathbf{p}}} = \sum_{h^2 \in H^2} v_{h^2}^2 \cdot p_h^{1-\text{Rep}} \;=\; \sum_{h^2 \in H^2} v_{h^2}^2 \cdot p_h^{2-\text{Rep}} \;=\; \ldots \;=\; \sum_{h^2 \in H^2} v_{h^2}^2 \cdot p_h^{n-\text{Rep}}. \tag{39}$$

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

$$\rho_{\mathbf{p}} \;=\; \sum_{h^2 \in H^2} v_{h^2}^2 \cdot p_{h^2}. \tag{40}$$

B.2 Proof of Theorem 6.1

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 39 in the case of $n = 2$ is given by,

$$\rho_{\mathbf{m}} := v_{CC}^2 + v_{CD}^2 = v_{CC}^2 + v_{DC}^2. \tag{41}$$

We can use this equality to conclude that

$$v_{CD}^2 = v_{DC}^2. \tag{42}$$

Moreover the cooperation rate of player 1 based on Eq. 40 is given by,

$$\begin{aligned}
\rho_{\mathbf{p}} \;&=\; v_{CC}^2 \, p_{CC} + v_{CD}^2 \, p_{CD} + v_{DC}^2 \, p_{DC} + v_{DD}^2 \, p_{DD} \\
&=\; v_{CC}^2 \, p_{CC} + v_{CD}^2 \, (p_{CD} + p_{DC}) + v_{DD}^2 \, p_{DD}.
\end{aligned} \tag{43}$$

Here, the second equality is due to Eq. (42).

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

AllD $= (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$ and Alternator $= (0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1)$.

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

$$p_{DD} \leq 1 - \frac{c}{b} \quad \text{and} \quad p_{CD} + p_{DC} \leq 1 + \frac{b-c}{c}. \tag{44}$$

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

$$
\begin{array}{rl}
0 \quad < & s_{\mathbf{m},\mathbf{p}} - (b-c) \\[4pt]
\overset{Eq.\ (21)}{=} & b\rho_{\mathbf{p}} - c\rho_{\mathbf{m}} - (b-c) \\[4pt]
\overset{Eqs.\ (41),(43),(38)}{=} & b\left(v_{CC}^2 p_{CC} + v_{CD}^2(p_{CD}+p_{DC}) + v_{DD}^2 p_{DD}\right) - c\left(v_{CC}^2 + v_{CD}^2\right) - (b-c)\left(v_{CC}^2 + 2v_{CD}^2 + v_{DD}^2\right) \\[4pt]
= & v_{CC}^2\, b\,(p_{CC} - 1) + v_{CD}^2\left(b(p_{CD}+p_{DC}) + c - 2b\right) + v_{DD}^2\left(b p_{DD} - (b-c)\right).
\end{array}
\tag{45}
$$

Condition (45) can hold only if,

$$b(p_{CD}+p_{DC}) + c - 2b > 0, \quad b\,p_{DD} - (b-c) > 0. \tag{46}$$

Thus, Eq. (44) reassures that $\mathbf{p}$ is Nash strategy, and given that $p_{CC} = 1$, it is a partner strategy. $\quad\square$

B.3 Proof of Theorem 6.2

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. 39) for $n = 3$ is given by,

$$\rho_{\mathbf{m}} := v_{CCC}^2 + v_{CCD}^2 + v_{DCC}^2 + v_{DCD}^2 = v_{CCC}^2 + v_{DCC}^2 + v_{CDC}^2 + v_{DDC}^2 = v_{CCC}^2 + v_{CCD}^2 + v_{CDC}^2 + v_{CDD}^2. \tag{47}$$

We can use this equality to conclude that

$$v_{CCD}^2 = v_{DCC}^2 \tag{48}$$

$$v_{DDC}^2 = v_{CDD}^2 \tag{49}$$

$$v_{CCD}^2 + v_{DCD}^2 = v_{CDC}^2 + v_{DDC}^2 \Rightarrow v_{CCD}^2 = v_{CDC}^2 + v_{CDD}^2 - v_{DCD}^2 \tag{50}$$

22

The average cooperation rate of 1's (Eq. (40)) for $n = 3$ is given by,

$$
\begin{aligned}
\rho_{\mathbf{p}} \quad &= \quad v_{CCC}^2\, p_{CCC} + v_{CCD}^2\, p_{CCD} + v_{CDC}^2\, p_{CDC} + v_{CDD}^2\, p_{CDD} + v_{DCD}^2\, p_{DCD} + \\
&\qquad + v_{DDC}^2\, p_{DDC} + v_{DDD}^2\, p_{DDD} \\
&\overset{\text{Eqs. (48),(49)}}{=}\quad v_{CCC}^2\, p_{CCC} + v_{CCD}^2\, (p_{CCD} + p_{DCC}) + v_{CDC}^2\, p_{CDC} + v_{CDD}^2\, (p_{CDD} + p_{DDC}) + \\
&\qquad + v_{DCD}^2\, p_{DCD} + v_{DDD}^2\, p_{DDD}
\end{aligned}
$$

$$(51)$$

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,

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

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

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

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

or equivalently, if

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

$$(52)$$

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

$$
\begin{aligned}
0 \quad &\leq \quad s_{\mathbf{m},\mathbf{p}} - (b-c) \\
&\overset{Eq.\ (21)}{=} \quad b\rho_{\mathbf{p}} - c\rho_{\mathbf{m}} - (b-c) \\
&\overset{Eqs.\ (51),(38)}{=} \quad b\left(v^2_{CCC}p_{CCC}+v^2_{CCD}(p_{CCD}+p_{DCC})+v^2_{CDC}p_{CDC}+v^2_{DDC}(p_{CDD}+p_{DDC})+v^2_{DCD}p_{DCD}+v^2_{DDD}p_{DD}\right. \\
&\qquad -c\left(v^2_{CCC}+2v^2_{CCD}+v^2_{DCD}\right) - (b-c)\left(v^2_{CCC}+2v^2_{CCD}+v^2_{CDC}+2v^2_{DDC}+v^2_{DCD}+v^2_{DDD}\right) \\
&= \quad b\,v^2_{CCC}(p_{CCC}-1)+v^2_{CCD}(b\,(p_{CCD}+p_{DCC}-2))+v^2_{CDC}(b\,(p_{CDC}-1)+c)+ \\
&\qquad v^2_{CDD}(b\,(p_{CDD}+p_{DDC}-2)+2\,c)+v^2_{DCD}(b\,(p_{DCD}-1))+v^2_{DDD}(b\,(p_{DDD}-1)+c) \\
&\overset{Eq.\ (50)}{=} \quad b\,v^2_{CCC}(p_{CCC}-1)+v^2_{DDD}(b\,(p_{DDD}-1)+c)+v^2_{CDC}(b\,(p_{CCD}+p_{DCC}+p_{CDC}-3)+c)+ \\
&\qquad v^2_{CDD}(b\,(p_{CDD}+p_{DDC}+p_{CCD}+p_{DCC}-4)+2\,c)+v^2_{DCD}(b\,(p_{DCD}-1)-b\,(p_{CCD}+p_{DCC})-2)
\end{aligned}
$$
$$(53)$$

Condition (53) 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. (52) reassure that $\mathbf{p}$ is Nash strategy, and given that $p_{CC} = 1$, it is a partner strategy. $\qquad\square$

C Proof of Theorem 6.3

To prove Theorem 6.3, we need to introduce some additional notation. We introduce the vector $\mathbf{w} = (w_i)_{i\in\{0,1,\dots,n\}}$, 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

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

24

Then we define w_i as

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

Please note that

$$\sum_{i=0}^{r} w_i = 1. \tag{54}$$

The cooperation rate of the reactive player is given by

$$\rho_{\mathbf{p}} = \sum_{i=0}^{n} r_i \cdot w_i. \tag{55}$$

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

$$\rho_{\tilde{\mathbf{p}}} = \sum_{i=0}^{n} \frac{i}{n} \cdot w_i. \tag{56}$$

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 $A = \{\mathbf{A^0}, \mathbf{A^1}, \ldots, \mathbf{A^n}\}$.

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

$$s_{\mathbf{A^i}, \mathbf{r}} = b \cdot r_i - \frac{i}{n} \cdot c \ \ for \ \ i \in [0, n]. \tag{57}$$

The intuition behind Eq. (57) 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 $i \in [0, n]$),

$$s_{\mathbf{A^i},\mathbf{r}} \leq b - c$$

$$b \cdot r_i - \frac{i}{n} \cdot c \leq b - c$$

$$r_i \leq 1 - \frac{i}{n} \cdot \frac{c}{b}.$$

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

$$
\begin{aligned}
0 \quad &\leq \quad s_{\mathbf{m},\mathbf{p}} - (b-c) \\
&\overset{Eq.\ (21)}{=} \quad b\rho_{\mathbf{p}} - c\rho_{\mathbf{m}} - (b-c) \\
&\overset{Eqs.\ (54),(55),(56)}{=} \quad b \sum_{k=0}^{n} r_{n-k} \cdot u_{n-k} - c \sum_{k=0}^{n} \frac{n-k}{n} \cdot u_{n-k} - (b-c) \sum_{k=0}^{n} u_{n-k} \\
&\quad\quad u_n\Big(b\,(r_n - 1)\Big) + \sum_{k=1}^{n} u_{n-k}\Big(b \sum_{k=1}^{n} r_{n-k} - c \sum_{k=0}^{n-1} \frac{n-k}{n} - (b-c)\Big)
\end{aligned}
$$
(58)

This condition holds only if

$$\Big(b\,r_{n-k} - c\,\frac{n-k}{n} - (b-c)\Big) < 0 \Rightarrow$$

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

$$r_{n-k} < 1 - \frac{n}{k}$$

$\square$

References

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

[2] R. Axelrod and W. D. Hamilton. The evolution of cooperation. *science*, 211(4489):1390–1396, 1981.

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

[4] 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.

[5] M. Nowak and K. Sigmund. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game. *Nature*, 364(6432):56–58, 1993.

[6] M. A. Nowak and K. Sigmund. Tit for tat in heterogeneous populations. *Nature*, 355(6357):250–253, 1992.

[7] 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.

[8] K. Sigmund. *The calculus of selfishness*. Princeton University Press, 2010.