Supporting Information for

Conditional cooperation with longer memory

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

Nikoleta E. Glynatsi.
E-mail: glynatsi@evolbio.mpg.de

This PDF file includes:

Supporting text
Figs. S1 to S2
Table S1
SI References

Supporting Information Text

This document provides further details on the previous literature, our methodology, our analytical findings, and our evolutionary simulations. Section 1 describes previous work on direct reciprocity with longer memory. For several important articles, we summarize which key question they explore, and how our article further contributes to this strand of work. Section 2 summarizes our model. Here, we introduce all relevant strategy spaces, and we show how to compute long-term payoffs for strategies with more than one-round memory. Section 3 contains our key analytical results. Here, we define partner strategies, we present an algorithm that allows us to verify whether a given reactive-n strategy is a partner, and we apply this algorithm to fully characterize the reactive-n partner strategies for $n=2$ and $n=3$. We perform a similar analysis for the stable defecting strategies. Finally, Section **??** presents some complementary evolutionary results. Finally, the Appendix in Section 4 contains the proofs of our mathematical results.

1. An overview of the related literature

By now, there is a vast literature on the evolution of reciprocal cooperation in repeated games. This literature has highlighted various strategies that either succeed in evolutionary simulations and computer tournaments, or that exhibit some interesting theoretical properties. This literature also describes how results depend on the payoffs of the game, or on implementation errors. For a general overview on these results, we refer to the reviews and text books in Refs. (1–5). In the following we discuss some key papers that discuss the role of memory on direct reciprocity.

Particular higher-memory strategies. There are several papers that highlight particular strategies with more than one-round memory. For example, Yi *et al* (6) describes a variant of Tit-for-Tat (`TFT`), referred to as `TFT-ATFT`. This strategy depends on the outcome of the last two rounds. It typically repeats the co-player's previous move, just as `TFT` does. But in case the focal player defected because of an error, the strategy prescribes to do the opposite of the co-player's previous move. The authors show that `TFT-ATFT` is robust with respect to implementation errors, and that it shows a good performance in evolutionary simulations.

Following up on this work, Baek and Murase (7) introduce a class of strategies called 'friendly rivals'. These strategies mutually cooperate with themselves. At the same time, a player with such a strategy always gets at least the payoff of its opponent, irrespective of the opponent's strategy. In particular, such strategies satisfy our notion of being a partner. While explicitly considering all pure strategies with three rounds of memory, they highlight one strategy that performs particularly well, referred to as `CAPRI`. In (8), the authors examine whether friendly rivals emerge in an evolutionary setting. They conclude that in well-mixed populations, friendly rivals play a minor role. Instead, partner strategies often seem to be sufficient to promote cooperation.

Hilbe et al (9) use an axiomatic approach to derive stable memory-n strategies of direct reciprocity. They impose three conditions on the strategies they look for: (*i*) *Mutual cooperation:* A strategy ought to fully cooperate with itself. (*ii*) *Retaliation:* Once a co-player defects, the strategy ought to defect for at least k rounds. (*iii*) *Error-correction:* When two players with that strategy interact, they return to mutual cooperation after at most k rounds, after any possible previous history. The authors find that there is a class of memory-n strategies that satisfies these properties. They refer to the elements of this class as All-or-None strategies, or $\texttt{AON}_k$. These $\texttt{AON}_k$ strategies satisfy the conditions for partner strategies. The well-known strategy Win-Stay Lose-Shift (10) is a special case of an $\texttt{AON}_k$-strategy with $k=1$.

Finally, Li *et al* (11) describe a strategy of *cumulative reciprocity*, or `CURE`. This strategy counts how often the two players cooperated in all previous rounds. A player with that strategy cooperates if the co-player's cooperation count is above a certain threshold (that depends on the player's own cooperation count). `CURE` satisfies the properties of a partner strategy. Moreover, the authors show that it performs well when competing with memory-1 strategies.

Compared to these previous articles, we take a more systematic approach. Instead of describing particular partner strategies, we aim to identify *all* partner strategies within an important strategy space, the space of (stochastic) reactive-n strategies. For the special case of reactive-n counting strategies, we provide explicit conditions for all memory lengths n. For the more general case of reactive-n strategies, we provide explicit conditions for $n=2$ and $n=3$. Moreover, we provide a general algorithm to test whether a given reactive-n strategy is a partner. This algorithm is valid for any n.

Particular classes of higher-memory strategies. In addition to the work on particular higher-memory strategies, there are also several articles that describe properties of certain strategy classes. One such class is the set of *zero-determinant* (ZD) strategies. This strategy class was first described by Press and Dyson (12). They showed that there are certain memory-1 strategies that can unilaterally enforce a linear relationship between the players' payoffs. Ueda (13, 14) extends this formalism to memory-n strategies. To this end, he derives a version of what we refer to as Akin's lemma (see Section 3). Based on this result, he describes a particular subset of these ZD strategies, which he calls variants of Tit-for-Tat.

Another strategy class that has been studied in more detail is the class of *reactive learning strategies*, as introduced by McAvoy and Nowak (15). These strategies gradually modify their cooperation propensity each round, based on the opponent's last action. McAvoy and Nowak focus on the feasible payoff space (i.e., the payoffs that can be achieved against a given player with fixed strategy). They define a map that assigns to any memory-one strategy an associated reactive learning strategy. The payoffs feasible against this reactive learning strategy are a strict subset of the feasible payoffs against the original memory-one strategy. McAvoy and Nowak interpret this result as an indication that players with reactive-learning strategies have more control over the resulting payoffs.

While both the papers by Ueda (13, 14) and by McAvoy and Nowak (15) offer some remarkable mathematical results, their scope is different from ours. They aim to derive general properties of the strategy classes they consider, independent of whether the resulting strategies form a Nash equilibrium. In particular, they do not explore which of their strategies can sustain cooperation (i.e., which of their strategies satisfies the notion of being a partner). These partner strategies are the main focus of our work.

Work on evolutionary robust cooperating strategies. Perhaps the paper most closely related to our is the work by Stewart and Plotkin (16). They consider memory-n strategies for public goods games in groups of size $m \geq 2$. They are particularly interested in those strategies that are 'evolutionary robust' – strategies that can resist selective invasion by any mutant. In the limit of large populations, evolutionary robust strategies are closely related to the Nash equilibria of the repeated game. Using an elegant coordinate transform, Stewart and Plotkin (16) give a rigorous characterization of those evolutionary robust strategies that sustain mutual cooperation. They use this characterization to compute the respective volume of evolutionary robust cooperating strategies. Analytically and with simulations, they show that this volume increases in the player's memory, relative to the respective volume of evolutionary robust strategies that lead to mutual defection.

Importantly, however, even in the case of $m = 2$ players, Stewart and Plotkin's characterization of evolutionary robust strategies is implicit. For example, for cooperating strategies, their Eq. (2) contains terms w^{l_0,l_p}. These terms in turn depend on the specific mutant strategy considered. Especially for larger values of n, these terms become increasingly expensive to compute. Moreover, in principle, one would have to check these conditions for all possible mutants. The authors argue that only four particular cases of w^{l_0,l_p} need to be studied (as described in detail in their SI Section 5.4). However, two of these cases correspond to an entire family of possible mutant strategies. In one case, one needs to check mutants that play in such a way that the two players do not defect for n consecutive rounds. In the other case, one needs to check mutants that make sure the two players do not cooperate for n consecutive rounds. Because there are many such mutants, these two cases can lead to more than two inequalities when characterizing the set of evolutionary robust strategies. In fact, for our example of finding the partners among the reactive-n counting strategies, we find that these two cases give rise to $n-1$ inequalities. Together with the condition that the strategy ought to be robust against `ALLD`, we obtain n conditions overall.

In contrast to Stewart and Plotkin's work, we explicitly characterize the partner strategies for reactive-2, reactive-3, and for reactive-n counting strategies. To this end, we also prove an interesting theoretical result: To test whether any given reactive-n strategy is a Nash equilibrium, it suffices to compare it to all pure self-reactive-n strategies.

2. Model and basic results

A. 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

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

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

$$T > R > P > S \quad \text{and} \quad 2R > T + S. \qquad [2]$$

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 parameters, the benefit b and the cost c of cooperation. The payoff matrix of the donation game takes the form

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

For this game to satisfy the conditions Eq. (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 do 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 repeated game, we define the players' payoffs as the expected payoff per round,

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

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 Eq. (4) is guaranteed, as we discuss further below.

B. 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 $\mathbf{h^i}=(a^i_{-n},\ldots,a^i_{-1})\in\{C,D\}^n$. 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.

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

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

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 $\mathcal{M}_1$, the one with $n=1$ is the most frequently studied. Memory-1 strategies take the form $\mathbf{m}=(m_{CC},m_{CD},m_{DC},m_{DD})$. 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. One well-known memory-1 strategy is Win-Stay Lose-Shift (10), $\mathbf{m}=(1,0,0,1)$. However, there are many others (4).

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*,

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

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

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

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 $\mathbf{p}=(p_C,p_D)$. Probably the best-known example of a reactive-1 strategy is Tit-for-Tat, `TFT` (17). `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` (18, 19). `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

$$p^*_D := \min\left\{1-(T-R)/(R-S),\ (R-P)/(T-P)\right\}. \qquad [8]$$

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,

$$\mathcal{S}_n := \left\{\mathbf{m}\in\mathcal{M}_n\;\middle|\; m_{(\mathbf{h^i},\mathbf{h^{-i}})}=m_{(\mathbf{h^i},\tilde{\mathbf{h}}^{\mathbf{-i}})} \text{ for all } \mathbf{h^i}\in H^i \text{ and } \mathbf{h^{-i}},\tilde{\mathbf{h}}^{\mathbf{-i}}\in H^{-i}\right\}. \qquad [9]$$

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,

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

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* $\tilde{\mathbf{p}}^{k-\mathrm{Rep}}$ choose the same action as they did k rounds ago. Formally, the entries of $\tilde{\mathbf{p}}^{k-\mathrm{Rep}}$ are defined by

$$\tilde{p}^{k-\mathrm{Rep}}_{\mathbf{h^i}} = \begin{cases} 1 & \text{if } a^i_{-k}=C \\ 0 & \text{if } a^i_{-k}=D. \end{cases} \qquad [11]$$

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 $\mathbf{m}^{k-\mathrm{Rep}}\in[0,1]^{2^{2n}}$ instead of $\tilde{\mathbf{p}}^{k-\mathrm{Rep}}\in[0,1]^{2^n}$.

C. Computing the payoffs of finite-memory strategies. A repeated game 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,

$$M_{\mathbf{h},\tilde{\mathbf{h}}} = x^1 \cdot x^2, \qquad [12]$$

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

$$x^i = \begin{cases} m^i_{(\mathbf{h^i},\mathbf{h^{-i}})} & \text{if } \tilde{a}^i_{-1} = C, \quad \text{and} \quad \tilde{a}^i_{-t} = a^i_{-t+1} \quad \text{for} \quad t \in \{2,\ldots,n\} \\ 1 - m^i_{(\mathbf{h^i},\mathbf{h^{-i}})} & \text{if } \tilde{a}^i_{-1} = D, \quad \text{and} \quad \tilde{a}^i_{-t} = a^i_{-t+1} \quad \text{for} \quad t \in \{2,\ldots,n\} \\ 0 & \text{if } \tilde{a}^i_{-t} \neq a^i_{-t+1} \quad \text{for some} \quad t \in \{2,\ldots,n\}. \end{cases} \qquad [13]$$

The resulting $2^{2n} \times 2^{2n}$ transition matrix $M = (M_{\mathbf{h},\tilde{\mathbf{h}}})$ fully describes the game dynamics 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 $m^i_{\mathbf{h}}$ 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 to $\mathbf{v}$,

$$\mathbf{v} = \lim_{\tau \to \infty} \frac{1}{\tau} \sum_{t=n}^{n+\tau-1} \mathbf{v}(t). \qquad [14]$$

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,

$$g^i_{\mathbf{h}}(k) = \begin{cases} R & \text{if } a^i_{-k} = C \text{ and } a^{-i}_{-k} = C \\ S & \text{if } a^i_{-k} = C \text{ and } a^{-i}_{-k} = D \\ T & \text{if } a^i_{-k} = D \text{ and } a^{-i}_{-k} = C \\ P & \text{if } a^i_{-k} = D \text{ and } a^{-i}_{-k} = D. \end{cases} \qquad [15]$$

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

$$\pi^i_{\mathbf{m^1},\mathbf{m^2}}(t) = \big\langle \mathbf{v}(t), \mathbf{g^i}(1) \big\rangle = \sum_{\mathbf{h} \in H} v_{\mathbf{h}}(t) \cdot g^i_{\mathbf{h}}(1). \qquad [16]$$

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

$$\begin{aligned} \pi^i(\mathbf{m^1},\mathbf{m^2}) &\overset{\text{Eq. (4)}}{=} \lim_{\tau \to \infty} \frac{1}{\tau} \sum_{t=1}^{\tau} \pi^i_{\mathbf{m^1},\mathbf{m^2}}(t) = \lim_{\tau \to \infty} \frac{1}{\tau} \sum_{t=n}^{\tau} \pi^i_{\mathbf{m^1},\mathbf{m^2}}(t) \\ &\overset{\text{Eq. (16)}}{=} \lim_{\tau \to \infty} \frac{1}{\tau} \sum_{t=n}^{\tau} \big\langle \mathbf{v}(t), \mathbf{g^i}(1) \big\rangle = \Big\langle \lim_{\tau \to \infty} \frac{1}{\tau} \sum_{t=n}^{\tau} \mathbf{v}(t), \mathbf{g^i}(1) \Big\rangle \\ &\overset{\text{Eq. (14)}}{=} \big\langle \mathbf{v}, \mathbf{g^i}(1) \big\rangle. \end{aligned} \qquad [17]$$

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. (17), one can also show

$$\big\langle \mathbf{v}, \mathbf{g^i}(1) \big\rangle = \big\langle \mathbf{v}, \mathbf{g^i}(2) \big\rangle = \ldots = \big\langle \mathbf{v}, \mathbf{g^i}(n) \big\rangle. \qquad [18]$$

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.

D. 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 (20): 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. (14) exists, then

$$\left\langle \mathbf{v}\,, \mathbf{m} - \mathbf{m}^{1-\mathrm{Rep}} \right\rangle = 0. \qquad [19]$$

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* (21). With similar methods as in Ref. (20), one can generalize this result to the context of memory-n strategies (13). Here, we provide the version that will be most relevant for us.

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. (14) *exists. Then for each k with $1 \leq k \leq n$, we obtain*

$$\left\langle \mathbf{v}\,, \mathbf{m} - \mathbf{m}^{k-Rep} \right\rangle = 0. \qquad [20]$$

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 to that n-history. That second interpretation leads to the expression $\langle \mathbf{v}, \mathbf{m}^{k-\mathrm{Rep}} \rangle$. According to Eq. (20), both interpretations are equivalent.

3. Characterizing the partner strategies among the reactive-n strategies

A. 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 (20), or as *partner strategies* by Hilbe *et al* (22). 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

$$\pi^1(\sigma, \sigma) \geq \pi^1(\sigma', \sigma) \;\text{ for all other strategies } \sigma' \text{ for which } \pi^1(\sigma', \sigma) \text{ exists.} \qquad [21]$$

(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 $\mathbf{h_C} = (\mathbf{h_C^i}, \mathbf{h_C^{-i}})$ 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, we note that our definition of Nash equilibria only requires that players cannot profitably deviate towards strategies *for which a payoff can be defined.* If the strategy σ is a memory-n strategy, in the following we make the slightly looser requirement that the strategy is a best response among all σ' for which the limit Eq. (14) exists. Fourth, 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. (21) 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 (12) 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.

B. 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. (13, 14). In the following, we say a memory-n strategy $\mathbf{m}$ is a *zero-determinant strategy* if there is a number γ and

vectors $\boldsymbol{\alpha}=(\alpha_1,\ldots,\alpha_n)$, $\boldsymbol{\beta}=(\beta_1,\ldots,\beta_n)$, $\boldsymbol{\delta}=(\delta_1,\ldots,\delta_n)$ with $\sum_{k=1}^n \delta_k = 1$ such that $\mathbf{m^i}$ can be written as

$$\mathbf{m^i} = \sum_{k=1}^n \alpha_k\, \mathbf{g}^i(k) + \sum_{k=1}^n \beta_k\, \mathbf{g}^{-i}(k) + \gamma\, \mathbf{1} + \sum_{k=1}^n \delta_k \mathbf{m}^{k-\mathrm{Rep}}. \qquad [22]$$

In this expression, $\mathbf{g}^i(k)$ is the vector that returns player i's payoff k rounds ago, as defined by Eq. (15), $\mathbf{m}^{k-\mathrm{Rep}}$ 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. If we define $\hat{\alpha}=\sum_{k=1}^n \alpha_i$ and $\hat{\beta}=\sum_{k=1}^n \beta_i$, we obtain based on Akin's lemma

$$\begin{aligned} 0 &\overset{\text{Eq. (20)}}{=} \sum_{k=1}^n \delta_k \big\langle \mathbf{v}, \mathbf{m^i} - \mathbf{m}^{k-\mathrm{Rep}} \big\rangle = \big\langle \mathbf{v}, \mathbf{m^i} - \sum_{k=1}^n \delta_k \mathbf{m}^{k-\mathrm{Rep}} \big\rangle \\ &\overset{\text{Eq. (22)}}{=} \big\langle \mathbf{v}, \sum_{k=1}^n \alpha_k\, \mathbf{g}^i(k) + \sum_{k=1}^n \beta_k\, \mathbf{g}^{-i}(k) + \gamma\, \mathbf{1} \big\rangle = \sum_{k=1}^n \alpha_k \langle \mathbf{v}, \mathbf{g}^i(k)\rangle + \sum_{k=1}^n \beta_k \langle \mathbf{v}, \mathbf{g}^{-i}(k)\rangle + \gamma \langle \mathbf{v}, \mathbf{1}\rangle \\ &\overset{(17),(18)}{=} \hat{\alpha}\, \pi^i(\mathbf{m^i}, \sigma^{-i}) + \hat{\beta}\, \pi^{-i}(\mathbf{m^i}, \sigma^{-i}) + \gamma. \end{aligned} \qquad [23]$$

That is, a player with a zero-determinant strategy enforces a linear relationship between the players' payoffs, irrespective of the co-player's strategy. Remarkably, the parameters $\hat{\alpha}$, $\hat{\beta}$, and γ of that linear relationship are entirely under player i's control.

Generalized versions of Tit-for-tat. Suppose for $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\delta}$ there is a k such that $\alpha_k = -\beta_k = 1/(T-S)$, $\delta_k = 1$ and $\alpha_j = \beta_j = \delta_j = 0$ for $j \neq k$. Moreover, set $\gamma = 0$. In that case, formula Eq. (22) yields the strategy with entries

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

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. (23), the strategy enforces equal payoffs against any co-player,

$$\pi^i(\mathbf{m^i}, \sigma^{-i}) = \pi^{-i}(\mathbf{m^i}, \sigma^{-i}). \qquad [24]$$

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 Eq. (24) 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 $\boldsymbol{\alpha} = \mathbf{0}$, $\gamma = 1 - c/b$, and there is a k such that $\beta_k = -1/b$, $\delta_k = 1$ and $\beta_j = \delta_j = 0$ for $j \neq k$. In that case Eq. (22) yields the strategy

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

That is, the generated strategy is `GTFT` (if $k=1$), or a delayed version thereof (for $k>1$). By Eq. (23), 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 Nash equilibria (independent of how they act during the first n rounds). If we additionally require them to cooperate during the first n rounds, they are also nice. Hence, they are partner strategies.

Equalizers. We can further generalize the previous insight that a strategy like Generous Tit-for-Tat (and delayed versions thereof) can unilaterally determine the co-player's payoff. To this end, consider the donation game and let $\boldsymbol{\delta}$ be a non-negative distribution such that $\delta_k \geq 0$ for all k, and $\sum_{k=1}^n \delta_i = 1$. Moreover, let $\kappa \in [0,1]$ be a parameter, and define

$$\boldsymbol{\alpha} = 0, \qquad \boldsymbol{\beta} = -\frac{1}{b}\boldsymbol{\delta}, \qquad \gamma = \kappa(1 - c/b). \qquad [25]$$

By Eq. (23), the enforced payoff relationship is now $\pi^{-i}(\mathbf{m^i}, \sigma^{-i}) = \kappa(b-c)$. That is, a player with a zero-determinant strategy that satisfies Eq. (25) sets the co-player's payoff to be a fixed number between 0 (for $\kappa = 0$) and $b - c$ (for $\kappa = 1$). In line with the previous literature, we refer to such a strategy as an *equalizer* (12, 32). The generalized versions of Generous Tit-for-Tat are a special case with $\kappa = 1$ and $\boldsymbol{\delta}$ being a distribution that puts full weight on a single round k. By keeping $\kappa = 1$ but varying the distribution $\boldsymbol{\delta}$, we obtain other equalizer strategies that fix the co-player's payoff to be $b - c$. They all represent linear combinations of the various delayed versions of `GTFT`; and they all are partner strategies.

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

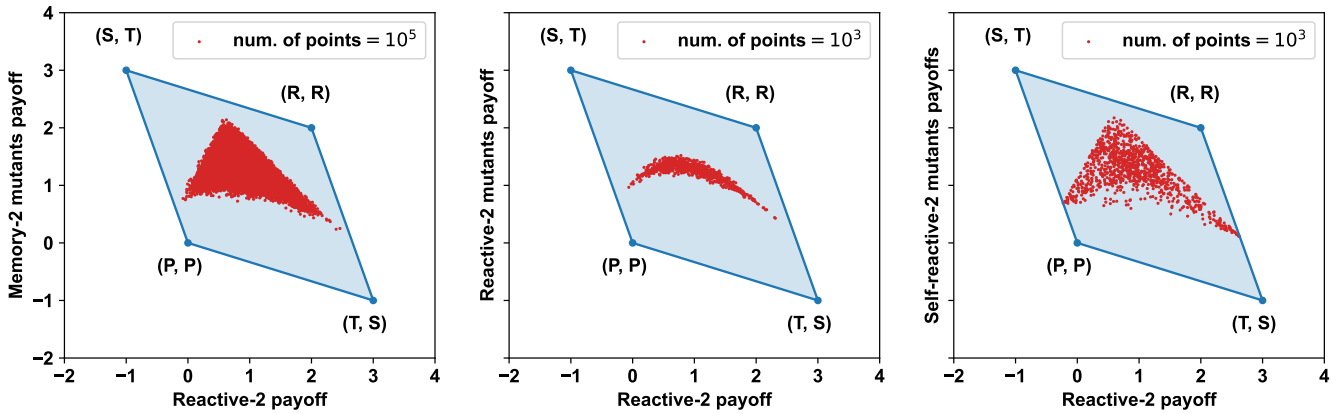

Fig. S1. Feasible payoffs for a reactive-2 strategy. We consider a player with reactive-2 strategy $\mathbf{p}=(0.37, 0.89, 0.95, 0.23)$. The player interacts with many other players (referred to as 'mutants') who adopt either some random memory-2 strategy (left), a random reactive-2 strategy (middle), or a random self-reactive-2 strategy (right panel). The panels show the resulting payoffs to the two players as red dots, with the x-axis showing the payoff of the focal player, and the y-axis showing the payoff of the mutants. We observe that when mutants use memory-2 strategies and self-reactive-2 strategies, we obtain the same region of feasible payoffs, in line with Lemma 2. In contrast, if mutants are restricted to reactive-2 strategies, the set of feasible payoffs is strictly smaller. Here, we consider a donation game with $b=3$ and $c=1$.

Sufficiency of checking 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 (12). 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 2. Moreover, suppose player 1 adopts some strategy σ such that for the resulting game, the time average $\mathbf{v}$ according to Eq. (14) *exists. Then there is a self-reactive-n strategy $\tilde{\mathbf{p}}\in\mathcal{S}_n$ such that $\pi^i(\sigma,\mathbf{p})=\pi^i(\tilde{\mathbf{p}},\mathbf{p})$ for $i\in\{1,2\}$.*

For an illustration of this result, see Figure S1. It shows that against a reactive-2 player, any payoff that can be achieved with a memory-2 strategy can already be achieved with a self-reactive-2 strategy.

If we are to verify that some given reactive-n strategy $\mathbf{p}$ is a Nash equilibrium, Lemma 2 simplifies our task considerably. Instead of checking condition Eq. (21) for all possible strategies σ', we only need to check it for all self-reactive strategies $\tilde{\mathbf{p}}\in\mathcal{S}_n$. 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

$$\pi^1(\tilde{\mathbf{p}},\mathbf{p})\geq\pi^1(\sigma',\mathbf{p}) \quad \textit{for all other strategies } \sigma' \textit{ for which the limit } \text{Eq. (14)} \textit{ exists.} \qquad [26]$$

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 $\tilde{\mathbf{p}}\in\mathcal{S}_n$.

Corollary 1 gives us a straightforward procedure to check whether a given reactive strategy $\mathbf{p}$ is a Nash equilibrium (for a depiction, see Algorithm 1). To verify that $\mathbf{p}$ is a Nash equilibrium, we merely need to compare its payoff against itself to the payoff of a deviation towards one of the 2^n pure self-reactive strategies.

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 $\tilde{\mathbf{p}}=(\tilde{p}_{\mathbf{h}^1})_{\mathbf{h}^1\in H^i}$, whereas player 2 uses a reactive-n strategy $\mathbf{p}=(p_{\mathbf{h}^{-1}})_{\mathbf{h}^{-1}\in H^{-i}}$. Our algorithm to compute payoffs for the two players in Section 2C would require us to interpret the two strategies as memory-n strategies. We would thus compute

Algorithm 1 An algorithm to verify whether a given reactive strategy $\mathbf{p}$ is a Nash equilibrium.

input : $\mathbf{p}, n$

pure_self_reactive_strategies $\leftarrow \left\{ \tilde{\mathbf{p}} \;\middle|\; \tilde{\mathbf{p}} \in \{0,1\}^{2^n} \right\}$

isNash $\leftarrow$ True

for $\tilde{\mathbf{p}} \in$ *pure_self_reactive_strategies* **do**

  if $\pi^1(\mathbf{p},\mathbf{p}) < \pi^1(\tilde{\mathbf{p}},\mathbf{p})$ **then**

    isNash $\leftarrow$ False

return ($\mathbf{p}$, isNash)

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 $h^1=(a^1_{-n},\dots,a^1_{-1})$ be the state in the current round. Then we obtain the following probability that the state after one round is $\tilde{h}^1=(\tilde{a}^1_{-n},\dots,\tilde{a}^1_{-1})$,

$$\tilde{M}_{\mathbf{h^1},\tilde{\mathbf{h}}^1} = \begin{cases} \tilde{p}_{\mathbf{h^1}} & \text{if } \tilde{a}^1_{-1}=C, \text{ and } \tilde{a}^1_{-t}=a^1_{-t+1} \text{ for all } t\in\{2,\dots,n\} \\ 1-\tilde{p}_{\mathbf{h^1}} & \text{if } \tilde{a}^1_{-1}=D, \text{ and } \tilde{a}^1_{-t}=a^1_{-t+1} \text{ for all } t\in\{2,\dots,n\} \\ 0 & \text{if } \tilde{a}^1_{-t}\neq a^1_{-t+1} \text{ for some } t\in\{2,\dots,n\}. \end{cases} \quad [27]$$

Similar to the vector $\mathbf{v}$ for matrix M, let $\tilde{\mathbf{v}}=(\tilde{v}_{\mathbf{h^1}})_{\mathbf{h^1}\in H^1}$ be the limiting distribution of the dynamics defined by $\tilde{M}$ (which only in exceptional cases depends on player 1's behavior during the first n rounds). Then the players' payoffs are given by

$$\begin{aligned} \pi^1(\tilde{\mathbf{p}},\mathbf{p}) &= \sum_{\mathbf{h^1}\in H^1} \tilde{v}_{\mathbf{h^1}} \Big(\tilde{\mathbf{p}}_{\mathbf{h^1}}\mathbf{p}_{\mathbf{h^1}}\cdot R + \tilde{\mathbf{p}}_{\mathbf{h^1}}(1-\mathbf{p}_{\mathbf{h^1}})\cdot S + (1-\tilde{\mathbf{p}}_{\mathbf{h^1}})\mathbf{p}_{\mathbf{h^1}}\cdot T + (1-\tilde{\mathbf{p}}_{\mathbf{h^1}})(1-\mathbf{p}_{\mathbf{h^1}})\cdot P \Big), \\ \pi^2(\tilde{\mathbf{p}},\mathbf{p}) &= \sum_{\mathbf{h^1}\in H^1} \tilde{v}_{\mathbf{h^1}} \Big(\tilde{\mathbf{p}}_{\mathbf{h^1}}\mathbf{p}_{\mathbf{h^1}}\cdot R + \tilde{\mathbf{p}}_{\mathbf{h^1}}(1-\mathbf{p}_{\mathbf{h^1}})\cdot T + (1-\tilde{\mathbf{p}}_{\mathbf{h^1}})\mathbf{p}_{\mathbf{h^1}}\cdot S + (1-\tilde{\mathbf{p}}_{\mathbf{h^1}})(1-\mathbf{p}_{\mathbf{h^1}})\cdot P \Big). \end{aligned} \quad [28]$$

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 $\mathbf{p^2}=(p^2_C,p^2_D)$. If we use the algorithm in Section 2C, we first formally represent these strategies as memory-1 strategies, $\mathbf{m^1}=(\tilde{p}^1_C,\tilde{p}^1_C,\tilde{p}^1_D,\tilde{p}^1_D)$ and $\mathbf{m^2}=(p^2_C,p^2_D,p^2_C,p^2_D)$. The respective transition matrix according to Eq. (12) is

$$M = \begin{pmatrix} \tilde{p}^1_C p^2_C & \tilde{p}^1_C(1-p^2_C) & (1-\tilde{p}^1_C)p^2_C & (1-\tilde{p}^1_C)(1-p^2_C) \\ \tilde{p}^1_C p^2_C & \tilde{p}^1_C(1-p^2_C) & (1-\tilde{p}^1_C)p^2_C & (1-\tilde{p}^1_C)(1-p^2_C) \\ \tilde{p}^1_D p^2_D & \tilde{p}^1_D(1-p^2_D) & (1-\tilde{p}^1_D)p^2_D & (1-\tilde{p}^1_D)(1-p^2_D) \\ \tilde{p}^1_D p^2_D & \tilde{p}^1_D(1-p^2_D) & (1-\tilde{p}^1_D)p^2_D & (1-\tilde{p}^1_D)(1-p^2_D) \end{pmatrix}. \quad [29]$$

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. (16), Player 1's payoff is the scalar product

$$\pi^1(\tilde{\mathbf{p}}^{\mathbf{1}},\mathbf{p^2}) = \big\langle \mathbf{v}\,,(R,S,T,P)\big\rangle. \quad [30]$$

Following Corollary 1, we can use these observations to characterize under which conditions a nice reactive strategy $\mathbf{p^2}=(1,p^2_D)$ 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 $\texttt{ALLC}=(1,1)$, $\texttt{ALLD}=(0,0)$, and $\texttt{Alternator}=(0,1)$; 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

$$\begin{aligned} \pi^1(\texttt{ALLC},\mathbf{p^2}) &= R, \\ \pi^1(\texttt{ALLD},\mathbf{p^2}) &= p^2_D\cdot T + (1-p^2_D)\cdot P \\ \pi^1(\texttt{Alternator},\mathbf{p^2}) &= p^2_D/2\cdot R + (1-p^2_D)/2\cdot S + 1/2\cdot T. \end{aligned} \quad [31]$$

We conclude that player 2's reactive strategy $\mathbf{p^2}$ 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

$$p^2_D \leq \min\left\{1-(T-R)/(R-S),\ (R-P)/(T-P)\right\}. \quad [32]$$

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

Instead of computing the 4×4 matrix M in Eq. (29), we could also consider the simplified 2×2 transition matrix Eq. (27). Here, the two possible states are $\mathbf{h^1} \in \{C, D\}$, and hence the matrix is

$$\tilde{M} = \begin{pmatrix} \tilde{p}_C^1 & 1-\tilde{p}_C^1 \\ \tilde{p}_D^1 & 1-\tilde{p}_D^1 \end{pmatrix}. \qquad [33]$$

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

$$\tilde{\mathbf{v}} = (\tilde{v}_C, \tilde{v}_D) = \left(\frac{\tilde{p}_D^1}{1-(\tilde{p}_C^1 - \tilde{p}_D^1)}, \frac{1-\tilde{p}_C^1}{1-(\tilde{p}_C^1 - \tilde{p}_D^1)} \right). \qquad [34]$$

If we take this invariant distribution and compute player 1's payoff according to Eq. (28), we recover the same expression as in Eq. (30), as expected.

D. 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 in the donation game)
A nice reactive-2 strategy $\mathbf{p}$*, is a partner strategy if and only if its entries satisfy the conditions*

$$p_{CC}=1, \qquad \frac{p_{CD}+p_{DC}}{2} \leq 1-\frac{1}{2}\cdot\frac{c}{b}, \qquad p_{DD} \leq 1-\frac{c}{b}. \qquad [35]$$

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 Eq. (35) 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.

Reactive-3 Partner Strategies. Next, we focus on the case $n=3$. 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 $p_{\mathbf{h}^{-\mathbf{i}}}$ refers to the player's cooperation probability, depending on the co-player's previous three actions, $\mathbf{h}^{-\mathbf{i}} = (a_{-3}^{-i}, a_{-2}^{-i}, a_{-1}^{-i})$. For the respective partner strategies, we obtain the following characterization.

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

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

Compared to the case $n=2$, the respective conditions are now somewhat more difficult to interpret. However, the general pattern reflected in each inequality remains the same. On average, a player's cooperation probability needs to be diminished in proportion to the number of defections in memory. More specifically, in a given inequality, let x be the proportion of D's in the histories on the left-hand's side. Then the threshold on the right-hand side is given by $1 - x \cdot c/b$. For example, in case of the second inequality in Eq. (36), the left-hand side features the histories CCD, CDC, DCC. Because the proportion of D's across all three histories is $1/3$, the threshold on the right-hand side is $1 - 1/3 \cdot c/b$.

Partners among the reactive-n counting strategies.. Using the same methods as before, one can in principle also characterize the partners among the reactive-4 and the reactive-5 strategies. However, the respective conditions quickly become unwieldy. However, in the special case of *reactive-n counting strategies*, we can provide a simple condition for any memory length. To formally 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

$$|\mathbf{h}^{-i}|=|\tilde{\mathbf{h}}^{-\mathbf{i}}| \quad \Rightarrow \quad p_{\mathbf{h}^{-\mathbf{i}}}=p_{\tilde{\mathbf{h}}^{-\mathbf{i}}}. \qquad [37]$$

That is, the reactive player's cooperation probability only depends on the number of cooperative acts during the past n rounds and not on their timing. Such reactive-n counting strategies can be written as n+1-dimensional vectors $\mathbf{r}=(r_k)_{k\in\{n,\ldots,0\}}$. Here, r_k is the player's cooperation probability, given the co-player cooperated k times during the past n rounds.

For example, for $n=2$, counting strategies are those reactive-2 strategies for which $p_{CD}=p_{DC}$. In that case, we associate $r_2=p_{CC}$, $r_1=p_{CD}=p_{DC}$, and $r_0=p_{DD}$. Within the set of reactive-n counting strategies, we get the following simple characterization of partners.

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

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

Here, the interpretation is again straightforward. For every defection in memory, the player's cooperation probability needs to be diminished by $1/n\cdot c/b$.

Note that for $n\geq 2$, the space of reactive-n counting strategies is strictly smaller than the space of reactive-n strategies. For example, the space of reactive-2 strategies is 4-dimensional, whereas the respective space of counting strategies is 3-dimensional. However, in terms of memory required to implement the respective strategies, the two strategy spaces are equivalent. In each case, the player needs to memorize the exact sequence $\mathbf{h}^{-\mathbf{i}}=(a^{-i}_{-n},\ldots,a^{-i}_{-1})$ of the co-player's last n actions. In the case of reactive-n strategies this kind of memory is already required to choose the appropriate cooperation probability $p_{\mathbf{h}^{-\mathbf{i}}}$ in the next round. However, also for reactive-n counting strategies merely memorizing the number $k:=|\mathbf{h}^{-i}|\in\{0,\ldots,n\}$ is not sufficient. This number would suffice to choose the correct cooperation probability r_k in the respective round. However, to correctly update the co-player's cooperation count $|\mathbf{h}^{-i}|$, the player still needs to remember which action the co-player chose n rounds ago (because that is the part of the n-history that will be dropped after the current round). The memory requirements in the next round will be similar; a player will then need to remember exactly what happened (what is currently) $n-1$ rounds ago. By this logic, it follows that even a memory-n counting strategy requires a memory of the exact sequence of the co-player's last n actions, not a mere memory of how often the co-player cooperated during the past n rounds.

E. 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 3C to analyze the general prisoner's dilemma. For the case of reactive-2 strategies, we obtain the following characterization.

Theorem 5 (Reactive-2 partner 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

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

Compared to the donation game, there are now more conditions, and these conditions are more difficult to interpret. Reassuringly, however, the conditions simplify to the conditions Eq. (35) 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 (Reactive-3 partner strategies in the prisoner's dilemma)
A nice reactive-3 strategy $\mathbf{p}$ is a partner strategy if and only if its entries satisfy the conditions in Table S1.

Given the large number of conditions in Table S1, we do not pursue a similar characterization for $n>3$, even though the same methods remain applicable.

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

Table S1. Necessary and sufficient conditions for a nice reactive-3 strategy to be a partner in the prisoner's dilemma.

F. Games with errors. In the previous section, we have studied equilibrium behavior when actions are executed perfectly. Individuals who wished to cooperate always had the means to do so; individuals who wished to defect always did. In the following, we allow decisions to be subject to errors. There are three motivations to do so. First, a model with errors seems to be the more realistic approximation to the games played in real life, where decisions tend to be noisy. Second, empirical evidence suggests that individuals take error rates into account, and adapt their behavior accordingly (23, 24). Finally, the study of errors is of considerable theoretical interest. For example, it has been noted early on that errors can have a drastic effect on the effectiveness of classical strategies such as `TFT` (19, 25, 26). Similarly, errors have a considerable impact on whether cooperation can be evolutionarily stable at all (27–29). These reasons make an extension of our results to games with errors desirable.

At first, such an extension may appear straightforward. Indeed, many of our key results naturally carry over. For example, in order to explore whether a reactive-n strategy is a Nash equilibrium, it still suffices to compare it to all self-reactive-n strategies. Nevertheless, errors give rise to additional complications. In particular, they make it more difficult to obtain closed-form expressions for the payoff of a partner strategy against itself. Without errors, this payoff is simply $b-c$ (in the donation game), or R (in the general prisoner's dilemma). In games with errors, the expressions quickly become more complex. Already for relatively simple strategy spaces, like reactive-2 or memory-1 strategies, these additional complications render an analytical treatment difficult. For example, as of now the set of all Nash equilibria among the memory-1 strategies is only known for games without errors (20, 30). More recently, some progress has been possible in games with an alternating move structure (31). But for games in which players decide simultaneously (as considered herein), an analytical treatment remains challenging.

In the following, we extend our results into two directions. First, we characterize equilibrium strategies when error rates are positive, but vanishingly small. In a second step, we then characterize a subset of reactive-n strategies, the so-called equalizers (12, 32). These strategies form a Nash equilibrium for any error rate.

Implementation errors. Herein, we consider implementation errors. An individual who intends to cooperate instead defects with some probability ε (conversely, an individual who intends to defect instead cooperates with the same probability). It is straightforward to include such errors into our model (see also 31). Once a player i with some memory-n strategy $\mathbf{m^i}=(m^i_{\mathbf{h}})_{\mathbf{h}\in H}$ is subject to errors, the player acts as if it used a strategy

$$\mathbf{m}^{\mathbf{i},\varepsilon}=\varphi^{\varepsilon}(\mathbf{m^i}):=(1-\varepsilon)\mathbf{m^i}+\varepsilon(\mathbf{1}-\mathbf{m^i}). \qquad [40]$$

The transformation φ^ε is a bijection that maps any *nominal strategy* $\mathbf{m} \in [0,1]^{2^{2n}}$ to an effective strategy $\mathbf{m}^\varepsilon \in [\varepsilon, 1-\varepsilon]^{2^{2n}}$. The inverse mapping is given by

$$\left(\varphi^\varepsilon\right)^{-1}(\mathbf{m}^\varepsilon) = \frac{\mathbf{m}^\varepsilon - \varepsilon \cdot \mathbf{1}}{1-2\varepsilon}. \qquad [41]$$

In addition to $\mathbf{m}^\varepsilon$, we use an analogous notation $\mathbf{p}^\varepsilon$ and $\tilde{\mathbf{p}}^\varepsilon$ for reactive and self-reactive strategies, respectively. For example, in games with errors, the strategy ALLD is mapped to $\texttt{ALLD}^\varepsilon = (\varepsilon, \ldots, \varepsilon)$. Similarly, ALLC is mapped onto $\texttt{ALLC}^\varepsilon = (1-\varepsilon, \ldots, 1-\varepsilon)$. Finally, a strategy like Tit-for-two-Tat is mapped from $\mathbf{p} = (1,1,1,0)$ to $\mathbf{p}^\varepsilon = (1-\varepsilon, 1-\varepsilon, 1-\varepsilon, \varepsilon)$.

Games with errors can be analyzed the same way games without errors. We only need to replace the players' nominal strategies with their effective strategies. For example, suppose we wish to compute payoffs when a player with nominal strategy $\mathbf{m^1}$ interacts with a co-player with strategy $\mathbf{m^2}$, in a game with error rate ε. Then we can use Eq. (4) to obtain

$$\pi^{i,\varepsilon}(\mathbf{m^1}, \mathbf{m^2}) = \pi^i(\mathbf{m^{1,\varepsilon}}, \mathbf{m^{2,\varepsilon}}). \qquad [42]$$

A few remarks are in order.

Remark 1 (*i*) *On payoffs being well-defined.* Because all entries of $m^{1,\varepsilon}$ and $m^{2,\varepsilon}$ are strictly between 0 and 1 for $\varepsilon > 0$, the payoffs $\pi^{i,\varepsilon}(\mathbf{m^1}, \mathbf{m^2})$ are independent of the players' actions during the first n rounds. They only depend on m^1 and m^2, and on the error rate ε.

(*ii*) *On the definition of partners.* For the case with errors, we define a memory-n strategy $\mathbf{m}$ to be self-cooperative if it yields the mutual cooperation payoff against itself as $\varepsilon \to 0$, i.e., when errors are rare. We say the strategy is a Nash equilibrium for the error rate ε if $\pi^{1,\varepsilon}(\mathbf{m}, \mathbf{m}) \geq \pi^{1,\varepsilon}(\mathbf{m'}, \mathbf{m})$ for all other memory-n strategies $\mathbf{m'}$. We note that because of the results of Press and Dyson (12) and Levinsky et al (33) the restriction that $\mathbf{m'}$ has the same memory length n imposes no restriction on the generality of this definition. We say a self-cooperative strategy is a partner for the error rate ε if it is a Nash equilibrium for that error rate.

(*iii*) *On the continuity of payoffs with respect to the error rate.* We note that in general, the payoffs $\pi^{i,\varepsilon}(\mathbf{m^1}, \mathbf{m^2})$ may be discontinuous as a function of ε at $\varepsilon = 0$. Such discontinuities can arise when the transition matrix M according to Eq. (12) has multiple absorbing states in the game without errors. In that case, the players' payoffs depends on their actions during the initial n rounds. The most prominent example of such a case occurs when two TFT players interact. Under the assumption that these players cooperate in the first round, we obtain for the donation game $\pi^1(\texttt{TFT}, \texttt{TFT}) = b - c$. That is, players achieve the mutual cooperation payoff. Once we allow for errors, however, we obtain $\pi^{1,\varepsilon}(\texttt{TFT}, \texttt{TFT}) = (b-c)/2$ even when error rates are vanishingly small, $\varepsilon \to 0$. Importantly, however, these discontinuities are rare. If the players' nominal strategies are such that the resulting transition matrix M has a unique absorbing state, the resulting payoff function $\pi^{i,\varepsilon}(\mathbf{m^1}, \mathbf{m^2})$ is continuous in ε at $\varepsilon = 0$.

(*iv*) *On the complexity of computing payoffs.* In general, it can be difficult to derive algebraic expressions for the players' payoffs when both use general reactive-n strategies. Already for $n = 2$, such a payoff computation requires us to derive an eigenvector of a 16×16 matrix. In the previous section without errors, we could bypass some of these complexities. For example, when a self-cooperative strategy $\mathbf{p} = (1, p_{CD}, p_{DC}, p_{DD})$ interacts with another, they cooperate indefinitely. Hence, the relevant eigenvector follows immediately. It is the 16-dimensional vector $\mathbf{v} = (v_\mathbf{h})$ with an entry of $v_{CC,CC} = 1$ and $v_\mathbf{h} = 0$ for all other histories $\mathbf{h}$. Unfortunately, this bypass is no longer available in games with errors. In that case, all of the 16 possible 2-histories $\mathbf{h}$ will be played with positive probability (in fact, that probability is bounded from below by ε^4, the probability that both players experience that history by a sequence of consecutive errors). It follows that for games with errors, we can no longer hope for a simple characterization of all Nash equilibria. To still make progress, in the following we use two distinct approaches. According to the first approach, we weaken the equilibrium notion we use, by considering approximate Nash equilibria. According to the second approach we consider a subset of Nash equilibria for which rigorous results are feasible.

Approximate Nash equilibria in games with implementation errors. Instead of Nash equilibria, in the following we consider a slightly looser equilibrium concept. Specifically, we say a memory-n strategy $\mathbf{m}$ is an approximate Nash equilibrium for a threshold of $\Delta > 0$ and an error rate ε if

$$\pi^{1,\varepsilon}(\mathbf{m'}, \mathbf{m}) - \pi^{1,\varepsilon}(\mathbf{m}, \mathbf{m}) \leq \Delta \quad \text{for all } m' \in \mathcal{M}_n. \qquad [43]$$

According to this definition, any payoff advantage that any alternative strategy $\mathbf{m'}$ might yield is bounded from above by Δ. For $\Delta = 0$, this notion recovers the definition of a Nash equilibrium. For positive Δ, the set of approximate equilibria is a proper superset of the set of Nash equilibria. If $\mathbf{m}$ is self-cooperative and a proximate Nash equilibrium, we call it an *approximate partner strategy*.

Especially in the context of stochastic evolutionary dynamics (as in our paper), the concept of an approximate Nash equilibrium is still informative. To see why, consider a finite resident population that uses the strategy $\mathbf{m}$. Furthermore, suppose this resident population is occasionally challenged by introducing a novel mutant strategy $\mathbf{m'}$. If $\mathbf{m}$ is a conventional Nash equilibrium, no mutant strategy has a selective advantage. Nevertheless, if the mutant strategy yields the same payoff as the resident, it may invade due to neutral drift (29). For approximate Nash equilibria, the situation is similar. Even if an

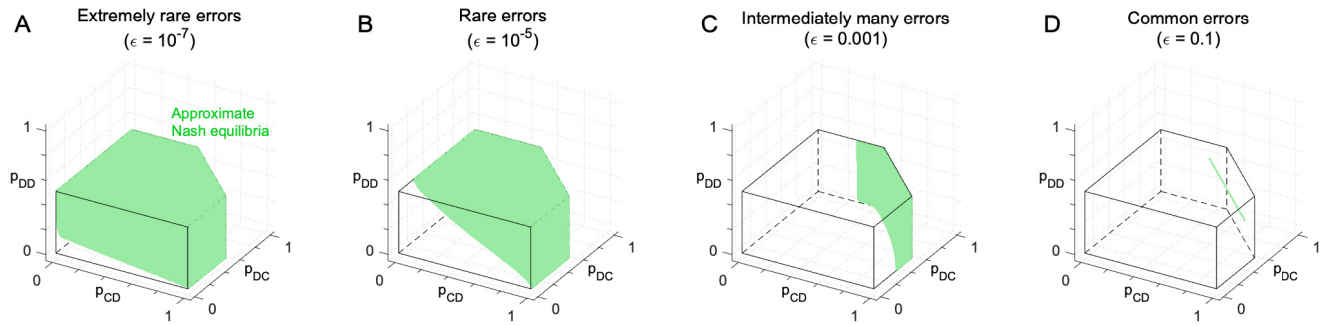

Fig. S2. The set of approximate Nash equilibria for varying error rates. We numerically compute the set of approximate Nash equilibria among the reactive-2 strategies for the donation game (using the same parameters as in Fig. 2C of the main text). To construct this figure, we use a finite grid of self-cooperative strategies $\mathbf{p} = (1, p_{CD}, p_{DC}, p_{DD})$. The entries p_{CD}, p_{DC}, p_{DD} are taken from the set $\{0, 0.01, 0.02, \ldots, 1\}$. For each resulting strategy $\mathbf{p}$, we explore whether any pure self-reactive-2 strategy $\tilde{\mathbf{p}}$ satisfies $\pi^{1,\varepsilon}(\tilde{\mathbf{p}}, \mathbf{p}) - \pi^{1,\varepsilon}(\mathbf{p}, \mathbf{p}) > \Delta$. We use a comparably small value of $\Delta = 0.001$, and we vary the error rate between $\varepsilon = 10^{-7}$ and $\varepsilon = 10^{-1}$. For sufficiently small error rates, the set of approximate partners for the game with errors approaches the set of partners for the game without errors (panel A). As we increase the error rate, the set of approximate partners approaches a line (panel D).

advantageous mutant strategy $\mathbf{m}'$ appears, the initial dynamics will be well-approximated by neutral drift – provided the threshold Δ and the selection strength β are sufficiently small. In this sense, we may still expect approximate Nash equilibria to be played for a substantial amount of time during the evolutionary process.

For analytical purposes, the notion of approximate Nash equilibria is useful because they allow us to exploit the properties of continuous functions. If a strategy $\mathbf{m}$ is a Nash equilibrium of the game without errors, and payoffs are continuous in ε, then $\mathbf{m}$ remains an approximate Nash equilibrium for sufficiently small error rates. By this argument, and by Remark 1(*iii*), we immediately obtain the following result.

Theorem 7 (Approximate partners for games with errors)
Suppose $\mathbf{m}$ *is a partner strategy and suppose all entries of* $\mathbf{m}$ *are strictly positive. Then for any threshold* Δ *one can find a threshold* $\bar{\varepsilon}$ *such that* $\mathbf{m}$ *is an approximate partner for sufficiently small error rates,* $\varepsilon \leq \bar{\varepsilon}$.

For sufficiently small error rates, this theorem guarantees that almost all partner strategies identified in the previous section remain approximate partners once we allow for errors. In particular, the theorem captures all of the reactive-2 strategies in the interior of the polyhedron displayed in Figure 2C,D. Only some boundary strategies such as `TFT` are excluded. At the same time, Theorem 7 does not quantify how small error rates need to be for a strategy $\mathbf{m}$ to be an approximate partner. To explore this issue in more detail, Figure S2 illustrates the set of approximate partners among the reactive-2 strategies for the donation game. As expected, the respective set approaches the polyhedron displayed in Figure 2C as errors become increasingly rare (Fig. S2A). However, as errors become more abundant, the set shrinks considerably. It approaches a line once $\varepsilon = 0$. As we describe in the next section, this line corresponds to the line of so-called 'equalizer strategies'. In particular, we show that all strategies on that line are not merely approximate Nash equilibria; they are conventional Nash equilibria.

Equalizers for games with implementation errors.

4. Appendix: Proofs

A. Proof of Lemma 1: Akin's lemma.

Proof. The proof is based on a similar argument as the proof of Eq. (18), showing that different ways of calculating payoffs are equivalent. Let us first introduce some notation. Let $\mathbf{m^1}$ be the memory-n strategy of player 1. For $t \geq n$ and the given strategy of player 2, let $\mathbf{v}(t) = (v_{\mathbf{h}})_{\mathbf{h} \in H}$ be the probability that player 1 observes the n-history $\mathbf{h}$ after players have made their t-th decision. By assumption, we can compute the limiting distribution

$$\mathbf{v} = \lim_{\tau \to \infty} \frac{1}{\tau} \sum_{t=n}^{n+\tau-1} \mathbf{v}(t). \qquad [44]$$

Moreover, let $\rho^i(t)$ be player i's cooperation probability in round t. For $t \geq n+1$, we obtain

$$\rho^1(t) = \left\langle \mathbf{v}(t-1), \mathbf{m^1} \right\rangle = \left\langle \mathbf{v}(t+k-1), \mathbf{m^{k-\mathrm{Rep}}} \right\rangle. \qquad [45]$$

That is, we either need to know how likely each n-history occurred at time $t-1$, and then we compute how likely player 1 is to cooperate in the next round, based on player 1's strategy. Or, we need to know how likely each n-history occurred after round $t+k-1$; and then we compute the correct probability by assuming player 1 cooperates in the next round if and only if the player cooperated k rounds before. Eq. (45) gives us two different ways to compute player 1's average payoff across all rounds,

$$\rho^1 := \lim_{\tau \to \infty} \frac{1}{\tau} \sum_{t=1}^{\tau} \rho^1(t). \qquad [46]$$

The first way is to take

$$\begin{aligned} \rho^1 &= \lim_{\tau \to \infty} \frac{1}{\tau} \sum_{t=1}^{\tau} \rho^i(t) = \lim_{\tau \to \infty} \frac{1}{\tau} \sum_{t=n+1}^{n+\tau} \rho^i(t) \\ &= \lim_{\tau \to \infty} \frac{1}{\tau} \sum_{t=n+1}^{n+\tau} \left\langle \mathbf{v}(t-1), \mathbf{m^1} \right\rangle = \left\langle \lim_{\tau \to \infty} \frac{1}{\tau} \sum_{t=n+1}^{n+\tau} \mathbf{v}(t-1), \mathbf{m^1} \right\rangle = \left\langle \mathbf{v}, \mathbf{m^1} \right\rangle. \end{aligned}$$

In particular, because $\left\langle \mathbf{v}, \mathbf{m^1} \right\rangle$ is well-defined, so is the limiting time average ρ^1. The second way is to take

$$\begin{aligned} \rho^1 &= \lim_{\tau \to \infty} \frac{1}{\tau} \sum_{t=1}^{\tau} \rho^i(t) = \lim_{\tau \to \infty} \frac{1}{\tau} \sum_{t=n+1}^{n+\tau} \rho^i(t) \\ &= \lim_{\tau \to \infty} \frac{1}{\tau} \sum_{t=n+1}^{n+\tau} \left\langle \mathbf{v}(t+k-1), \mathbf{m^{k-\mathrm{Rep}}} \right\rangle = \left\langle \lim_{\tau \to \infty} \frac{1}{\tau} \sum_{t=n+1}^{n+\tau} \mathbf{v}(t+k-1), \mathbf{m^{k-\mathrm{Rep}}} \right\rangle = \left\langle \mathbf{v}, \mathbf{m^{k-\mathrm{Rep}}} \right\rangle. \end{aligned}$$

We conclude $0 = \rho^1 - \rho^1 = \left\langle \mathbf{v}, \mathbf{m^1} \right\rangle - \left\langle \mathbf{v}, \mathbf{m^{k-\mathrm{Rep}}} \right\rangle = \left\langle \mathbf{v}, \mathbf{m^1} - \mathbf{m^{k-\mathrm{Rep}}} \right\rangle$. □

B. Proof of Lemma 2: Sufficiency of testing self-reactive strategies.

Proof. The proof uses similar arguments as in a study by Park on alternating games *et al* (31). For the given game between player 1 (with arbitrary strategy σ^1) and player 2 (with reactive-n strategy $\mathbf{p^2}$), let $v_{\mathbf{h}}(t)$ denote the probability to observe an n-history $\mathbf{h}$ at time $t \geq n$. By assumption, the following time averages are well-defined,

$$v_{\mathbf{h}} := \lim_{\tau \to \infty} \frac{1}{\tau} \sum_{t=n}^{n+\tau-1} v_{\mathbf{h}}(t) \qquad [47]$$

Moreover, for any $t \geq n$ and $\mathbf{h} \in H$, let $\sigma^1_{\mathbf{h}}(t)$ denote the conditional probability that player 1 cooperates at time $t+1$, given the n-history after round t is $\mathbf{h}$. Depending on $\left(\sigma^1_{\mathbf{h}}(t)\right)$ and $\mathbf{v}$, we define an associated self-reactive strategy $\tilde{\mathbf{p}}^{\mathbf{1}}$ for player 1. For any given history $\mathbf{h^1} \in H^1$, the corresponding probability $\tilde{p}^1_{\mathbf{h^1}}$ is defined as an implicit solution of the equation

$$\left(\sum_{\mathbf{h^2} \in H^2} v_{(\mathbf{h^1}, \mathbf{h^2})} \right) \tilde{p}^1_{\mathbf{h^1}} = \sum_{\mathbf{h^2} \in H^2} \left(\lim_{\tau \to \infty} \frac{1}{\tau} \sum_{t=n}^{n+\tau-1} v_{(\mathbf{h^1}, \mathbf{h^2})}(t) \cdot \sigma^1_{(\mathbf{h^1}, \mathbf{h^2})}(t) \right). \qquad [48]$$

Note that for each $\mathbf{h} \in H$, the limit in the bracket on the right hand side exists, for otherwise the limits $v_{\mathbf{h}}$ according to Eq. (47) would not exist. Also note that if the bracket on the left hand's side is zero, the right hand side must be zero, and $\tilde{p}^1_{\mathbf{h^1}}$ can be chosen arbitrarily. Only if the bracket on the left hand side is positive, $\tilde{p}^1_{\mathbf{h^1}}$ is uniquely defined.

We are going to show: If player 1 uses $\tilde{\mathbf{p}}^{\mathbf{1}}$ instead of σ^1, then $\mathbf{v}$ defined by Eq. (47) is an invariant distribution of the corresponding transition matrix M defined by Eq. (12) (hence it is also the limiting distribution of the resulting game if the first n moves are chosen accordingly). For simplicity, we show the required relationship $\mathbf{v}=\mathbf{v}M$ for one of the 2^{2^n} equations. For the one equation we show, we consider the history according to which everyone fully cooperates, $\mathbf{h_C}=(\mathbf{h^1_C},\mathbf{h^2_C})$. For an arbitrary n-history $\mathbf{h}^i=(a^i_{-n},\ldots,a^i_{-i})$, we say the n-history $\tilde{\mathbf{h}}^{\mathbf{i}}=(\tilde{a}^i_{-n},\ldots,\tilde{a}^i_{-1})$ *is a possible successor of* $\mathbf{h}$ if $\tilde{a}^i_{-t}=a^i_{-t+1}$ for $t\in\{2,\ldots,n\}$. To indicate successorship, we define a function $e_{\mathbf{h},\tilde{\mathbf{h}}}$ that is one if $\tilde{\mathbf{h}}$ is a possible successor of $\mathbf{h}$, and zero otherwise. By definition of $v_{\mathbf{h}}(t)$, $\sigma^1_{\mathbf{h}}(t)$, and $p^2_{\mathbf{h}}(t)$, we obtain for $t\geq n$

$$v_{(\mathbf{h^1_C},\mathbf{h^2_C})}(t+1)=\sum_{\mathbf{h^1}\in H^1}\sum_{\mathbf{h^2}\in H^2} v_{(\mathbf{h^1},\mathbf{h^2})}(t)\cdot\sigma^1_{(\mathbf{h^1},\mathbf{h^2})}(t)\cdot p^2_{\mathbf{h^1}}\cdot e_{\mathbf{h^1},\mathbf{h^1_C}}\cdot e_{\mathbf{h^2},\mathbf{h^2_C}}. \quad [49]$$

If we sum up this equation from time $t=n$ to $t=n+\tau-1$, divide by τ, and rearrange the terms, we obtain

$$\frac{1}{\tau}\sum_{t=n}^{n+\tau-1} v_{(\mathbf{h^1_C},\mathbf{h^2_C})}(t+1)=\sum_{\mathbf{h^1}\in H^1}\sum_{\mathbf{h^2}\in H^2}\left(\frac{1}{\tau}\sum_{t=n}^{n+\tau-1} v_{(\mathbf{h^1},\mathbf{h^2})}(t)\cdot\sigma^1_{(\mathbf{h^1},\mathbf{h^2})}(t)\right)\cdot p^2_{\mathbf{h^1}}\cdot e_{\mathbf{h^1},\mathbf{h^1_C}}\cdot e_{\mathbf{h^2},\mathbf{h^2_C}}. \quad [50]$$

Taking the limit $\tau\to\infty$, and taking into account the relationships Eq. (47) and Eq. (48), this simplifies to

$$v_{(\mathbf{h^1_C},\mathbf{h^2_C})}=\sum_{\mathbf{h^1}\in H^1}\sum_{\mathbf{h^2}\in H^2} v_{(\mathbf{h^1},\mathbf{h^2})}\cdot\left(\tilde{p}^1_{\mathbf{h^1}}\,e_{\mathbf{h^1},\mathbf{h^1_C}}\right)\cdot\left(p^2_{\mathbf{h^1}}\,e_{\mathbf{h^2},\mathbf{h^2_C}}\right). \quad [51]$$

By using the definition of transition probabilities in Eq. (12), this expression further simplifies to

$$v_{\mathbf{h_C}}=\sum_{\mathbf{h}} v_{\mathbf{h}}\cdot M_{\mathbf{h},\mathbf{h_C}} \quad [52]$$

That is, out of the 2^{2n} individual equations in the linear system $\mathbf{v}=\mathbf{v}M$, we have verified the equation for the probability to observe full cooperation $\mathbf{h_C}$ after one round. All other equations follow analogously. □

C. Proof of Theorem 1: Sufficiency of pure self-reactive strategies. By Lemma 2, there exists a best response to $\mathbf{p}$ within the self-reactive n strategies. It 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. The first such result uses an insight by Press & Dyson (12). They showed that given the transition matrix of a game among two memory-1 players, one can compute the players' payoffs by considering determinants of certain associated matrices. Herein, we apply their method to the transition matrix $\tilde{M}=(\tilde{M}_{\mathbf{h},\mathbf{h}'})$ according to Eq. Eq. (27) for a given self-reactive strategy $\tilde{\mathbf{p}}\in\mathcal{S}_n$. For some fixed n-history $\mathbf{h}'$, we define an associated matrix $\tilde{M}_{\mathbf{h}'}$ that one obtains from $\tilde{M}$ with the following two steps:

1. Subtract the $2^n\times 2^n$ identity matrix I from $\tilde{M}$.
2. In the resulting matrix, replace the last column by a column that only contains zeros, except for the row corresponding to the history $\mathbf{h}'$, for which the entry is one.

These matrices $\tilde{M}_{\mathbf{h}'}$ can be used to compute the invariant distribution of the original matrix $\tilde{M}$ as follows.

Auxiliary result 1: Let $\tilde{\mathbf{p}}\in\mathcal{S}_n$ be such that its transition matrix $\tilde{M}$ according to Eq. Eq. (27) has a unique invariant distribution $\tilde{\mathbf{v}}=(\tilde{v}_{\mathbf{h^1}})_{\mathbf{h^1}\in H^1}$. Then for all $\mathbf{h}'\in H^1$ we have

$$\tilde{v}_{\mathbf{h}'}=\frac{\det(\tilde{M}_{\mathbf{h}'})}{\sum_{\mathbf{h^1}\in H^1}\det(\tilde{M}_{\mathbf{h^1}})}. \quad [53]$$

Proof of Auxiliary result 1. The result follows from Press & Dyson's formula for the dot product of the invariant distribution $\tilde{\mathbf{v}}$ with an arbitrary vector $\mathbf{f}$, by taking the vector $\mathbf{f}$ to be the unit vector with only the entry for history $\mathbf{h}'$ being one. □

Based on this first auxiliary result, we have an explicit representation of the payoff function $\pi^1(\tilde{\mathbf{p}},\mathbf{p})$ that describes the payoff of a self-reactive player with strategy $\tilde{\mathbf{p}}$ against a reactive player with strategy $\mathbf{p}$. Specifically, by plugging Eq. Eq. (53) into Eq. (28), we obtain

$$\pi^1(\tilde{\mathbf{p}},\mathbf{p})=\frac{\sum_{\mathbf{h^1}\in H^1}\det(\tilde{M}_{\mathbf{h^1}})\Big(\tilde{\mathbf{p}}_{\mathbf{h^1}}\mathbf{p}_{\mathbf{h^1}}\cdot R+\tilde{\mathbf{p}}_{\mathbf{h^1}}(1-\mathbf{p}_{\mathbf{h^1}})\cdot S+(1-\tilde{\mathbf{p}}_{\mathbf{h^1}})\mathbf{p}_{\mathbf{h^1}}\cdot T+(1-\tilde{\mathbf{p}}_{\mathbf{h^1}})(1-\mathbf{p}_{\mathbf{h^1}})\cdot P\Big)}{\sum_{\mathbf{h^1}\in H^1}\det(\tilde{M}_{\mathbf{h^1}})}. \quad [54]$$

For our purposes, the following properties of this payoff function will be important.

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

Proof of Auxiliary Result 2. By its definition, each $\det(\tilde{M}_{\mathbf{h^i}})$ is a polynomial. Moreover, because for each history $\mathbf{h}'$, the cooperation probability $\tilde{p}_{\mathbf{h}'}$ only appears in a single row of $\tilde{M}_{\mathbf{h^i}}$ (and there it appears linearly), it also appears linearly in $\det(\tilde{M}_{\mathbf{h^i}})$. Finally, we note that $\det(\tilde{M}_{\mathbf{h^i}})$ does not depend on $\tilde{p}_{\mathbf{h^i}}$. To see this, we can compute $\det(\tilde{M}_{\mathbf{h^i}})$ using Laplace expansion along the last column. As a result, we obtain that this determinant is up to its sign equal to the determinant of the matrix one obtains from $\tilde{M}_{\mathbf{h^i}}$ by deleting the last column, and the row $\mathbf{h^i}$ (which is the only row of $\tilde{M}_{\mathbf{h^i}}$ that contains $\tilde{p}_{\mathbf{h^i}}$).

Finally, we note that the payoff function is bounded, because as an average payoff per round, payoffs need to be between T and S. Taken together, these observations imply the result for $\pi^1(\tilde{\mathbf{p}}, \mathbf{p})$. □

The following result describes a useful property of bounded linear rational functions.

Auxiliary Result 3: Suppose $g, h : [0,1]^k \to \mathbb{R}$ and suppose both $g(\mathbf{x})$ and $h(\mathbf{x})$ are linear in each component of $\mathbf{x} = (x_1, \ldots, x_k)$. Moreover, suppose $f := g/h$ is bounded on $[0,1]^k$. For a given $\mathbf{x}$ and $j \in \{1, \ldots, k\}$, we define an associated function $f_{\mathbf{x},j} : [-x_j, 1-x_j] \to \mathbb{R}$ by only varying the j-th component, $f_{\mathbf{x},j}(t) = f(x_1, \ldots, x_j + t, \ldots, x_k)$. Then for all $\mathbf{x} \in [0,1]^k$ and j, the function $f_{\mathbf{x},j}(t)$ is either monotonically increasing, monotonically decreasing, or constant.

Proof of Auxiliary Result 3. Let $g(\mathbf{x}) := a_0 + a_1 x_1 + \ldots + a_k x_k$ and $h(\mathbf{x}) := b_0 + b_1 x_1 + \ldots + b_k x_k$, and consider some arbitrary but fixed $\mathbf{x} \in [0,1]^k$ and j. We compute

$$f'_{\mathbf{x},j}(t) = \frac{\partial}{\partial t} f(x_1, \ldots, x_j + t, \ldots, x_k) = \frac{a_j \left(\sum_{i \neq j} b_i x_i\right) - b_j \left(\sum_{i \neq j} a_i x_i\right)}{\left(b_0 + b_1 x_1 + \ldots + b_j (x_j + t) + \ldots + b_k x_k\right)^2}. \quad [55]$$

Because f is bounded on the entire domain, the denominator in this expression for $f'_{\mathbf{x},j}(t)$ is strictly positive. Moreover, we note that the numerator is independent of t. Thus, depending on the sign of the numerator, $f'_{\mathbf{x},j}(t)$ is either monotonically increasing, monotonically decreasing, or constant. □

After these preparations, we are ready to prove the main result.

Proof of Theorem 1. For a given reactive strategy $\mathbf{p} \in \mathcal{R}_n$, let the self-reactive $\tilde{\mathbf{p}} \in \mathcal{S}_n$ be a best response. Suppose there is some history $\mathbf{h}'$ such that $0 < \tilde{p}_{\mathbf{h}} < 1$. It follows from the Auxiliary Results 2 and 3 that $\pi^1(\tilde{\mathbf{p}}, \mathbf{p})$ is either monotonically increasing, monotonically decreasing, or constant in $\tilde{p}_{\mathbf{h}'}$. If it was increasing or decreasing, we end up with a contradiction, because no local improvement should be possible for a best response. Therefore, $\pi^1(\tilde{\mathbf{p}}, \mathbf{p})$ must be independent of $\tilde{p}_{\mathbf{h}'}$, and hence we can set $\tilde{p}_{\mathbf{h}'} = 0$ or $\tilde{p}_{\mathbf{h}'} = 1$ without changing $\pi^1(\tilde{\mathbf{p}}, \mathbf{p})$. By iteratively applying this reasoning to all histories $\mathbf{h}$ for which $0 < \tilde{p}_{\mathbf{h}} < 1$, we obtain the desired result. □

D. Proof of Theorem 2: Reactive-2 partner strategies in the donation game.

Proof. Given that player 1 uses a nice reactive-2 strategy $\mathbf{p} = (1, p_{CD}, p_{DC}, p_{DD})$, the claim is true if and only if it is true for all deviation towards the sixteen pure self-reactive-2 strategies $\tilde{\mathbf{p}} \in \{0,1\}^{16}$. In the following, we enumerate these sixteen strategies, $\{\tilde{\mathbf{p}}_{\mathbf{0}}, \ldots, \tilde{\mathbf{p}}_{\mathbf{15}}\}$, by interpreting them as binary numbers,

$$\tilde{\mathbf{p}} = (\tilde{p}_{CC}, \tilde{p}_{CD}, \tilde{p}_{DC}, \tilde{p}_{DD}) \;\mapsto\; \tilde{p}_{CC} \cdot 2^3 + \tilde{p}_{CD} \cdot 2^2 + \tilde{p}_{DC} \cdot 2^1 + p_{DD} \cdot 2^0. \quad [56]$$

In particular, $\texttt{ALLD} = (0,0,0,0)$ is mapped to the number $j = 0$, and $\texttt{ALLC} = (1,1,1,1)$ is mapped to $j = 15$. The possible payoffs against the reactive strategy $\mathbf{p}$ can be computed by Eq. (28), which yields

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

In this list, some strategy indices j appear multiple times. Those instances correspond to strategies that have multiple invariant distributions (such as the strategy 1-round repeat, with $j = 10$). For those strategies, we have computed the payoffs for all possible initial n-histories. Requiring the payoffs in this list to be at most the mutual cooperation payoff $b - c$, we get the following unique conditions,

$$p_{DD} \leq 1 - \frac{c}{b}, \qquad \frac{p_{CD} + p_{DC}}{2} \leq 1 - \frac{1}{2}\frac{c}{b}, \qquad \frac{p_{CD} + p_{DC} + p_{DD}}{3} \leq 1 - \frac{2}{3}\frac{c}{b}.$$

Because the last condition is implied by the first two, we end up with the conditions in Eq. (35). □

E. Proof of Theorem 3: Reactive-3 partner strategies in the donation game.

Proof. The proof is similar to the previous one. Again, enumerating the 256 pure self-reactive 3 strategies $\tilde{\mathbf{p}}$ by interpreting the strategy as a binary number, we obtain the following payoffs.

$$\pi^1(\tilde{\mathbf{p}}_\mathbf{j}, \mathbf{p}) = b\, p_{DDD} \qquad \text{for } j \in \{0, 2, 4, 6, \ldots, 250, 252, 254\}$$

$$\pi^1(\tilde{\mathbf{p}}_\mathbf{j}, \mathbf{p}) = \frac{p_{CDD}+p_{DCD}+p_{DDC}+p_{DDD}}{4}\, b - \frac{1}{4}\, c \qquad \text{for } j \in \{1, 9, 33, 41, 65, 73, 97, 105, 129, 137, 161, 169, 193, 201, 225, 233\}$$

$$\pi^1(\tilde{\mathbf{p}}_\mathbf{j}, \mathbf{p}) = \frac{p_{CCD}+p_{CDD}+p_{DCC}+p_{DDC}+p_{DDD}}{5}\, b - \frac{2}{5}\, c \qquad \text{for } j \in \{3, 7, 35, 39, 131, 135, 163, 167\}$$

$$\pi^1(\tilde{\mathbf{p}}_\mathbf{j}, \mathbf{p}) = \frac{p_{CDC}+p_{DCD}}{2}\, b - \frac{1}{2}\, c \qquad \text{for } j \in \{4-7, 12-15, 20-23, 28-31, 68-71, 76-79, 84-87, 92-95, 132-135, 140-143, 148-151, 156-159, 196-199, 204-207, 212-215, 220-223\}$$

$$\pi^1(\tilde{\mathbf{p}}_\mathbf{j}, \mathbf{p}) = \frac{1+p_{CCD}+p_{CDD}+p_{DCC}+p_{DDC}+p_{DDD}}{6}\, b - \frac{1}{2}\, c \qquad \text{for } j \in \{11, 15, 43, 47\}$$

$$\pi^1(\tilde{\mathbf{p}}_\mathbf{j}, \mathbf{p}) = \frac{p_{CDD}+p_{DCD}+p_{DDC}}{3}\, b - \frac{1}{3}\, c \qquad \text{for } j \in \{16, 17, 24, 25, 48, 49, 56, 57, 80, 81, 88, 89, 112, 113, 120, 121, 144, 145, 152, 153, 176, 177, 184, 185, 208, 209, 216, 217, 240, 241, 248, 249\}$$

$$\pi^1(\tilde{\mathbf{p}}_\mathbf{j}, \mathbf{p}) = \frac{p_{CCD}+p_{CDD}+p_{DCC}+p_{DDC}}{4}\, b - \frac{1}{2}\, c \qquad \text{for } j \in \{18, 19, 22, 23, 50, 51, 54, 55, 146, 147, 150, 151, 178, 179, 182, 183\}$$

$$\pi^1(\tilde{\mathbf{p}}_\mathbf{j}, \mathbf{p}) = \frac{1+p_{CCD}+p_{CDD}+p_{DCC}+p_{DDC}}{5}\, b - \frac{3}{5}\, c \qquad \text{for } j \in \{26, 27, 30, 31, 58, 59, 62, 63\}$$

$$\pi^1(\tilde{\mathbf{p}}_\mathbf{j}, \mathbf{p}) = \frac{p_{CCD}+p_{CDC}+p_{CDD}+p_{DCC}+p_{DCD}+p_{DDC}+p_{DDD}}{7}\, b - \frac{3}{7}\, c \qquad \text{for } j \in \{37, 67, 165, 195\}$$

$$\pi^1(\tilde{\mathbf{p}}_\mathbf{j}, \mathbf{p}) = \frac{1+p_{CCD}+p_{CDC}+p_{CDD}+p_{DCC}+p_{DCD}+p_{DDC}+p_{DDD}}{8}\, b - \frac{1}{2}\, c \qquad \text{for } j \in \{45, 75\}$$

$$\pi^1(\tilde{\mathbf{p}}_\mathbf{j}, \mathbf{p}) = \frac{p_{CCD}+p_{CDC}+p_{CDD}+p_{DCC}+p_{DCD}+p_{DDC}}{6}\, b - \frac{1}{2}\, c \qquad \text{for } j \in \{52, 53, 82, 83, 180, 181, 210, 211\}$$

$$\pi^1(\tilde{\mathbf{p}}_\mathbf{j}, \mathbf{p}) = \frac{1+p_{CCD}+p_{CDC}+p_{CDD}+p_{DCC}+p_{DCD}+p_{DDC}}{7}\, b - \frac{4}{7}\, c \qquad \text{for } j \in \{60, 61, 90, 91\}$$

$$\pi^1(\tilde{\mathbf{p}}_\mathbf{j}, \mathbf{p}) = \frac{p_{CCD}+p_{CDC}+p_{DCC}}{3}\, b - \frac{2}{3}\, c \qquad \text{for } j \in \{96-103, 112-119, 224-231, 240-247\}$$

$$\pi^1(\tilde{\mathbf{p}}_\mathbf{j}, \mathbf{p}) = \frac{1+p_{CCD}+p_{CDC}+p_{DCC}}{4}\, b - \frac{3}{4}\, c \qquad \text{for } j \in \{104-111, 120-127\}$$

$$\pi^1(\tilde{\mathbf{p}}_\mathbf{j}, \mathbf{p}) = b - c \qquad \text{for } j \in \{128, 129, 130, \ldots, 255\}$$

Requiring these payoffs to be at most equal to the mutual cooperation payoff $b-c$ gives

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

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

$$\frac{p_{CDD}+p_{DCD}+p_{DDC}+p_{DDD}}{4} \leq 1-\frac{3}{4}\cdot\frac{c}{b}, \qquad \frac{p_{CCD}+p_{CDC}+p_{CDD}+p_{DCC}+p_{DCD}+p_{DDC}+p_{DDD}}{7} \leq 1-\frac{4}{7}\cdot\frac{c}{b},$$

$$\frac{p_{CCD}+p_{CDD}+p_{DCC}+p_{DDC}+p_{DDD}}{5} \leq 1-\frac{3}{5}\cdot\frac{c}{b}, \qquad \frac{p_{CCD}+p_{CDC}+p_{CDD}+p_{DCC}+p_{DCD}+p_{DDC}}{6} \leq 1-\frac{1}{2}\cdot\frac{c}{b}.$$

The statement follows by noting that the five conditions in the first two rows imply the four other conditions. □

F. Proof of Theorem 4: Reactive-n counting strategies in the donation game. Before we go into the details of the proof, we first start with two useful observations.

1. Assume player 1 adopts a given self-reactive strategy $\tilde{\mathbf{p}}$ and player 2 adopts the reactive-n strategy $\mathbf{r} = (r_k)_{k\in\{n,\ldots,0\}}$. For the resulting game, suppose $\mathbf{v}$ is the limiting distribution according to Eq. (14). Then it is useful to express $\mathbf{v}$ in terms of what the counting player can remember. To this end, let H^1_k be the set of n-histories according to which player 1 has cooperated exactly k times,

$$H^1_k = \left\{ \mathbf{h^1} \in H^1 \;\middle|\; |\mathbf{h^1}| = k \right\}. \qquad [57]$$

Accordingly, let $\mathbf{u}=(u_k)_{k\in\{0,\dots,n\}}$ be the distribution that summarizes how often, on average, player 1 cooperates j times during n consecutive rounds,

$$u_k^1 = \sum_{\mathbf{h^1}\in H_k^1} v_{\mathbf{h^1}}. \qquad [58]$$

In particular, the entries of $\mathbf{u}$ are normalized,

$$\sum_{k=0}^{n} u_k^1 = 1. \qquad [59]$$

Moreover, the average cooperation rate of the two players can be written as

$$\rho^1 = \sum_{k=0}^{n} \frac{k}{n}\, u_k^1 \qquad \text{and} \qquad \rho^2 = \sum_{k=0}^{n} r_k\, u_k^1. \qquad [60]$$

Because payoffs in the donation game only depend on the players' average cooperation rates (but not on the timing of cooperation), we conclude that player 1's payoff is

$$\pi^1(\tilde{\mathbf{p}}, \mathbf{r}) = \sum_{k=0}^{n} (r_k\, b - \frac{k}{n}\, c)\, u_k^1. \qquad [61]$$

2. There is a set of strategies for which payoffs are particularly easy to compute. We refer to them as simple periodic strategies, σ_k with $k\in\{0,\dots,n\}$. A player with strategy σ_k cooperates in round t if and only if

$$t-1 \mod n < k. \qquad [62]$$

That is, such a player cooperates in the first k rounds, then defects for $n-k$ rounds, then cooperates for another k rounds, only to defect in the $n-k$ subsequent rounds, etc. Such strategies are interesting for two reasons. First, they all can be interpreted as a round-n repeat strategy $\tilde{\mathbf{p}}^{\mathbf{n-Rep}}$, as defined by Eq. (11). During the initial n rounds, they cooperate according to Eq. (62); thereafter, they simply repeat whatever they have done n rounds ago. Second, players with strategy σ_k always act in such a way that according to any resulting n-history, they have cooperated exactly k times during the last n rounds. As a result, if player 1 adopts such a strategy in a donation game against a player with a reactive-n counting strategy $\mathbf{r}$, then player 1's average payoff is

$$\pi^1(\sigma_k, \mathbf{r}) = r_k\, b - \frac{k}{n}\, c. \qquad [63]$$

After these observations, we are ready for the actual proof.

Proof of Theorem 4.

($\Rightarrow$) Suppose the reactive-n counting strategy $\mathbf{r}$ is a partner. Because it is nice, it cooperates against an unconditional cooperator, and hence $r_n=1$. Because it is a Nash equilibrium, player 1 must not have an incentive to deviate towards any of the simple periodic strategies σ_k. By Eq. (63), this means that for all $k\in\{0,\dots,n\}$ we have

$$r_k\, b - \frac{k}{n}\, c \le b-c. \qquad [64]$$

These conditions are equivalent to $r_{n-k} \le 1-\frac{k}{n}\,\frac{c}{b}$, the inequalities in Eq. (38).

($\Leftarrow$) Because $\mathbf{r}$ is nice, $r_n=1$. The proof is now by contradiction; suppose the conditions in Eq. (38) hold, but $\mathbf{r}$ is not a Nash equilibrium. Then there needs to be some self-reactive $\tilde{\mathbf{p}}$ such that $\pi^1(\tilde{\mathbf{p}}, \mathbf{r}) > b-c$. It follows that

$$\begin{aligned} 0 \quad &< \quad \pi^1(\tilde{\mathbf{p}}, \mathbf{r}) - (b-c) \\ &\overset{\text{Eq. (59),Eq. (61)}}{=} \sum_{k=0}^{n} (r_k\, b - \frac{k}{n}\, c)\, u_k^1 \;-\; \sum_{k=0}^{n} (b-c) u_k^1 \\ &= \quad (r_n-1) b\, u_n + \sum_{k=0}^{n-1} \Big((r_k-1)b + \frac{n-k}{n} c \Big)\, u_k^1 \\ &= \quad b \cdot \sum_{k=1}^{n} \underbrace{\Big(r_{n-k} - \big(1-\frac{k}{n}\frac{c}{b}\big)\Big)}_{\le 0 \text{ by Eq. (38)}} u_{n-k}^1 \le 0. \end{aligned} \qquad [65]$$

We end up with $0<0$, a contradiction.

□

G. Proof of Theorem 5: Reactive-2 partner strategies in the prisoner's dilemma.

Proof. The proof is analogous to the proof of Theorem 2 for the donation game. For the general prisoner's dilemma, the payoffs of the 16 pure self-reactive-2 strategies are

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

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

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

$$\pi^1(\tilde{\mathbf{p}}_\mathbf{j}, \mathbf{p}) = \frac{Rp_{CD} + S(1-p_{CD}) + Tp_{DC} + P(1-p_{DC})}{2} \qquad \text{for } i \in \{4,5,12,13\}$$

$$\pi^1(\tilde{\mathbf{p}}_\mathbf{j}, \mathbf{p}) = \frac{R(p_{CD}+p_{DC}) + S(2-p_{CD}-p_{DC}) + T}{3} \qquad \text{for } i \in \{6,7\}$$

$$\pi^1(\tilde{\mathbf{p}}_\mathbf{j}, \mathbf{p}) = R \qquad \text{for } i \in \{8,9,10,11,12,13,14,15\}$$

By requiring these expressions to be at most equal to R, we obtain

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

□

H. Proof of Theorem 6: Reactive-3 partner strategies in the prisoner's dilemma. Again, we compute payoffs for all 256 self-reactive-3 strategies. The expressions are given below,

$$\pi^1(\tilde{\mathbf{p}}_{\mathbf{j}}, \mathbf{p}) = \frac{(T-P)\left(p_{CDD}+p_{DCD}+p_{DDC}\right)+3P+(R-S)\,p_{DDD}+S}{4} \quad \text{for } j \in \{1, 9, 33, 41, 65, 73, 97, 105, 129, 137, 161, 169, 193, 201, 225, 233\}$$

$$\pi^1(\tilde{\mathbf{p}}_{\mathbf{j}}, \mathbf{p}) = \frac{(T-P)\,p_{CDC}+P+(R-S)\,p_{DCD}+S}{2} \quad \text{for } j \in \{4-7, 12-15, 20-23, 28-31, 68-71, 76-79, 84-87, 92-95, 132-135, 140-143, 148-151, 156-159, 196-199, 204-207, 212-215, 220-223\}$$

$$\pi^1(\tilde{\mathbf{p}}_{\mathbf{j}}, \mathbf{p}) = -P\left(p_{DDD}-1\right)+T\,p_{DDD} \quad \text{for } j \in \{0, 2, 4, \ldots, 252, 254\}$$

$$\pi^1(\tilde{\mathbf{p}}_{\mathbf{j}}, \mathbf{p}) = \frac{(T-P)(p_{CCD}+p_{CDD}+p_{DDC})+3P+(R-S)(p_{CDC}+p_{DCC}+p_{DCD}+p_{DDD})+4S+T}{8} \quad \text{for } j \in \{45\}$$

$$\pi^1(\tilde{\mathbf{p}}_{\mathbf{j}}, \mathbf{p}) = \frac{(T-P)\,p_{DCC}+P+(R-S)(p_{CDC}+p_{CCD})+2S}{3} \quad \text{for } j \in \{96-103, 112-119, 224-231, 240-247\}$$

$$\pi^1(\tilde{\mathbf{p}}_{\mathbf{j}}, \mathbf{p}) = \frac{(T-P)(p_{CCD}+p_{DCC}+p_{DDC})+3P+(R-S)(p_{CDC}+p_{CDD}+p_{DCD})+3S}{6} \quad \text{for } j \in \{52, 53, 180, 181\}$$

$$\pi^1(\tilde{\mathbf{p}}_{\mathbf{j}}, \mathbf{p}) = \frac{(T-P)(p_{CCD}+p_{DDC})+2P+T+(R-S)(p_{CDC}+p_{CDD}+p_{DCC}+p_{DCD})+4S}{7} \quad \text{for } j \in \{60, 61\}$$

$$\pi^1(\tilde{\mathbf{p}}_{\mathbf{j}}, \mathbf{p}) = \frac{(T-P)(p_{CCD}+p_{CDD}+p_{DCC})+3P+(R-S)(p_{DDC}+p_{DDD})+2S}{5} \quad \text{for } j \in \{3, 7, 35, 39, 131, 135, 163, 167\}$$

$$\pi^1(\tilde{\mathbf{p}}_{\mathbf{j}}, \mathbf{p}) = \frac{(T-P)(p_{DCD}+p_{DDC})+2P+(R-S)\,p_{CDD}+S}{3} \quad \text{for } j \in \{16, 17, 24, 25, 48, 49, 56, 57, 80, 81, 88, 89, 112, 113, 120, 121, 144, 145, 152, 153, 176, 177, 184, 185, 208, 209, 216, 217, 240, 241, 248, 249\}$$

$$\pi^1(\tilde{\mathbf{p}}_{\mathbf{j}}, \mathbf{p}) = R \quad \text{for } j \in \{128, 129, \ldots, 255\}$$

$$\pi^1(\tilde{\mathbf{p}}_{\mathbf{j}}, \mathbf{p}) = \frac{(T-P)\,p_{CCD}+P+T+(R-S)(p_{CDD}+p_{DCC}+p_{DDC})+3S}{5} \quad \text{for } j \in \{26, 27, 30, 31, 58, 59, 62, 63\}$$

$$\pi^1(\tilde{\mathbf{p}}_{\mathbf{j}}, \mathbf{p}) = \frac{(T-P)(p_{CCD}+p_{DCC})+2P+(R-S)(p_{CDD}+p_{DDC})+2S}{4} \quad \text{for } j \in \{18, 19, 22, 23, 50, 51, 54, 55, 146, 147, 150, 151, 178, 179, 182, 183\}$$

$$\pi^1(\tilde{\mathbf{p}}_{\mathbf{j}}, \mathbf{p}) = \frac{(T-P)(p_{CDC}+p_{DCD})+2P+T+(R-S)(p_{CCD}+p_{CDD}+p_{DCC}+p_{DDC})+4S}{7} \quad \text{for } j \in \{90, 91\}$$

$$\pi^1(\tilde{\mathbf{p}}_{\mathbf{j}}, \mathbf{p}) = \frac{(T-P)(p_{CDC}+p_{CDD}+p_{DCD})+3P+T+(R-S)(p_{CCD}+p_{DCC}+p_{DDC}+p_{DDD})+4S}{8} \quad \text{for } j \in \{75\}$$

$$\pi^1(\tilde{\mathbf{p}}_{\mathbf{j}}, \mathbf{p}) = \frac{(T-P)(p_{CDC}+p_{DCC}+p_{DCD})+3P+(R-S)(p_{CCD}+p_{CDD}+p_{DDC})+3S}{6} \quad \text{for } j \in \{82, 83, 210, 211\}$$

$$\pi^1(\tilde{\mathbf{p}}_{\mathbf{j}}, \mathbf{p}) = \frac{(T-P)(p_{CCD}+p_{CDD}+p_{DCC}+p_{DDC})+4P+(R-S)(p_{CDC}+p_{DCD}+p_{DDD})+3S}{7} \quad \text{for } j \in \{37, 165\}$$

$$\pi^1(\tilde{\mathbf{p}}_{\mathbf{j}}, \mathbf{p}) = \frac{T+(R-S)(p_{CCD}+p_{CDC}+p_{DCC})+3S}{4} \quad \text{for } j \in \{104-111, 120-127\}$$

$$\pi^1(\tilde{\mathbf{p}}_{\mathbf{j}}, \mathbf{p}) = \frac{(T-P)(p_{CCD}+p_{CDD})+2P+T+(R-S)(p_{DCC}+p_{DDC}+p_{DDD})+3S}{6} \quad \text{for } j \in \{11, 15, 43, 47\}$$

$$\pi^1(\tilde{\mathbf{p}}_{\mathbf{j}}, \mathbf{p}) = \frac{(T-p)(p_{CDC}+p_{CDD}+p_{DCC}+p_{DCD})+4P+(R-S)(p_{CCD}+p_{DDC}+p_{DDD})+3S}{7} \quad \text{for } j \in \{67, 195\}$$

By requiring the above expressions to be smaller than or equal to R, we obtain the inequalities in Table S1.

5. Supplementary References

1. R Axelrod, *The evolution of cooperation.* (Basic Books, New York, NY), (1984).
2. MA Nowak, *Evolutionary dynamics.* (Harvard University Press, Cambridge MA), (2006).
3. K Sigmund, *The calculus of selfishness.* (Princeton University Press), (2010).
4. C Hilbe, K Chatterjee, MA Nowak, Partners and rivals in direct reciprocity. *Nat. Hum. Behav.* **2**, 469–477 (2018).
5. N Glynatsi, V Knight, A bibliometric study of research topics, collaboration and centrality in the field of the Iterated Prisoner's Dilemma. *Humanit. Soc. Sci. Commun.* **8**, 45 (2021).
6. S Do Yi, SK Baek, JK Choi, Combination with anti-tit-for-tat remedies problems of tit-for-tat. *J. Theor. Biol.* **412**, 1–7 (2017).
7. Y Murase, SK Baek, Five rules for friendly rivalry in direct reciprocity. *Sci. reports* **10**, 16904 (2020).
8. Y Murase, SK Baek, Grouping promotes both partnership and rivalry with long memory in direct reciprocity. *PLoS Comput. Biol.* **19**, e1011228 (2023).
9. C Hilbe, LA Martinez-Vaquero, K Chatterjee, MA Nowak, Memory-n strategies of direct reciprocity. *Proc. Natl. Acad. Sci.* **114**, 4715–4720 (2017).
10. M Nowak, K Sigmund, A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game. *Nature* **364**, 56–58 (1993).
11. J Li, et al., Evolution of cooperation through cumulative reciprocity. *Nat. Comput. Sci.* **2**, 677–686 (2022).
12. WH Press, FJ Dyson, Iterated prisoner's dilemma contains strategies that dominate any evolutionary opponent. *Proc. Natl. Acad. Sci.* **109**, 10409–10413 (2012).
13. M Ueda, Memory-two zero-determinant strategies in repeated games. *Royal Soc. open science* **8**, 202186 (2021).
14. M Ueda, Controlling conditional expectations by zero-determinant strategies. *Oper. Res. Forum* **3**, 48 (2022).
15. A McAvoy, MA Nowak, Reactive learning strategies for iterated games. *Proc. Royal Soc. A* **475**, 20180819 (2019).
16. AJ Stewart, JB Plotkin, Small groups and long memories promote cooperation. *Sci. reports* **6**, 1–11 (2016).
17. R Axelrod, WD Hamilton, The evolution of cooperation. *Science* **211**, 1390–1396 (1981).
18. MA Nowak, K Sigmund, Tit for tat in heterogeneous populations. *Nature* **355**, 250–253 (1992).
19. P Molander, The optimal level of generosity in a selfish, uncertain environment. *J. Confl. Resolut.* **29**, 611–618 (1985).
20. E Akin, The iterated prisoner's dilemma: good strategies and their dynamics in *Ergodic Theory, Advances in Dynamical Systems.* (de Gruyter Berlin), pp. 77–107 (2016).
21. C Hilbe, B Wu, A Traulsen, MA Nowak, Cooperation and control in multiplayer social dilemmas. *Proc. Natl. Acad. Sci. USA* **111**, 16425–16430 (2014).
22. C Hilbe, A Traulsen, K Sigmund, Partners or rivals? strategies for the iterated prisoner's dilemma. *Games economic behavior* **92**, 41–52 (2015).
23. D Fudenberg, DG Rand, A Dreber, Slow to anger and fast to forgive: Cooperation in an uncertain world. *Am. Econ. Rev.* **102**, 720–749 (2012).
24. AA Arechar, A Dreber, D Fudenberg, DG Rand, "I'm just a soul whose intentions are good": The role of communication in noisy repeated games. *Games Econ. Behav.* **104**, 726–743 (2017).
25. R Axelrod, D Dion, The further evolution of cooperation. *Science* **242**, 1385–1390 (1988).
26. J Wu, R Axelrod, How to cope with noise in the iterated prisoner's dilemma. *J. Confl. Resolut.* **39**, 183–189 (1995).
27. R Boyd, J Lorberbaum, No pure strategy is evolutionary stable in the iterated prisoner's dilemma game. *Nature* **327**, 58–59 (1987).
28. R Boyd, Mistakes allow evolutionary stability in the repeated Prisoner's Dilemma game. *J. Theor. Biol.* **136**, 47–56 (1989).
29. J García, M van Veelen, In and out of equilibrium I: Evolution of strategies in repeated games with discounting. *J. Econ. Theory* **161**, 161–189 (2016).
30. AJ Stewart, JB Plotkin, Collapse of cooperation in evolving games. *Proc. Natl. Acad. Sci. USA* **111**, 17558 – 17563 (2014).
31. PS Park, MA Nowak, C Hilbe, Cooperation in alternating interactions with memory constraints. *Nat. Commun.* **13**, 737 (2022).
32. MC Boerlijst, MA Nowak, K Sigmund, Equal pay for all prisoners. *Am. Math. Mon.* **104**, 303–307 (1997).
33. R Levinsky, A Neyman, M Zeleny, Should i remember more than you? On the best response to factor-based strategies. *Int. J. Game Theory* **49**, 1105–1124 (2020).