

Conditional cooperation with longer memory

Nikoleta E. Glynatsi^{a, 1}, Ethan Akin^b, Martin A. Nowak^c, and Christian Hilbe^a

This manuscript was compiled on November 11, 2024

Direct reciprocity is a wide-spread mechanism for the evolution of cooperation. In repeated interactions, players can condition their behavior on previous outcomes. A well known approach is given by reactive strategies, which respond to the co-player's previous move. Here we extend reactive strategies to longer memories. A reactive- n strategy takes into account the sequence of the last n moves of the co-player. A reactive- n counting strategy responds to how often the co-player cooperated during the last n rounds. We derive an algorithm to identify the partner strategies within these strategy sets. Partner strategies are those that ensure mutual cooperation without exploitation. We give explicit conditions for all partner strategies among reactive-2, reactive-3 strategies, and reactive- n counting strategies. To further explore the role of memory, we perform evolutionary simulations. We vary several key parameters, such as the cost-to-benefit ratio of cooperation, the error rate, and the strength of selection. Within the strategy sets we consider, we find that longer memory tends to promote cooperation. This positive effect of memory is particularly pronounced when individuals take into account the precise sequence of moves.

evolutionary game theory | direct reciprocity | evolution of cooperation | prisoner's dilemma

To a considerable extent, human cooperative behavior is governed by direct reciprocity (1, 2). This mechanism for cooperation can explain why people return favors (3), why they show more effort in group tasks when others do (4), or why they stop cooperating when they feel exploited (5, 6). The main theoretical framework to describe reciprocity is the repeated prisoner's dilemma (7–12). This game considers two individuals, referred to as players, who repeatedly decide whether to cooperate or to defect with one another (Fig. 1A). Both players prefer mutual cooperation to mutual defection. Yet given the co-player's action, each player has an incentive to defect. One common implementation of the prisoner's dilemma is the donation game. Here, cooperation simply means to pay a cost $c > 0$ for the co-player to get a benefit $b > c$. Despite the simplicity of these games, they can give rise to remarkable dynamical patterns. These patterns have been explored in numerous studies (13–32). Some of this literature describes how the evolution of cooperation depends on the game parameters, such as the benefit of cooperation, or the frequency with which errors occur (33–36). Others describe the effect of different learning dynamics (37, 38), of population structure (39–42), or of the strategies that players are permitted to use (43).

Strategies of the repeated prisoner's dilemma can vary in their complexity. While some are straightforward to implement, like always defect, many others are more sophisticated (44, 45). One way to quantify a strategy's complexity is to resort to the number of past rounds that the player needs to remember. Unconditional strategies like 'always defect' or 'always cooperate' are said to be memory-0. Strategies that only depend on the previous round, such as 'Tit-for-Tat' (7, 46) or 'Win-Stay Lose-Shift' (20, 21), are memory-1 (Fig. 1B). Similarly, one can distinguish strategies that require more than one round of memory, or strategies that cannot be implemented with finite memory (10).

Traditionally, most theoretical research on the evolution of reciprocity focuses on memory-1 strategies (21–31). Although one-round memory can explain some of the empirical regularities in human behavior (47–51), people often take into account more than the last round (52, 53). In experiments, longer memory seems particularly relevant for noisy games, where people occasionally defect because of unintended errors (54). However, a formal analysis of strategies with more than one-round memory is non-trivial, for two reasons. First, as the memory length n increases, strategies become harder to interpret. For example, because two consecutive rounds of the prisoner's dilemma allow for 16 possible outcomes, memory-2 strategies need to specify 16 conditional cooperation probabilities (55, 56). Although some of the resulting strategies have an intuitive interpretation, such as 'Tit-for-Two-Tat' (7), many others are difficult to make sense of. Second, the number

Significance Statement

In repeated interactions, people tend to cooperate conditionally. They are influenced by whether others cooperate, and react accordingly. Direct reciprocity is based on repeated interactions between two players. Nice strategies are those that are never the first to defect. Consequently, they never seek to exploit the other. Partner strategies are nice strategies which can sustain full cooperation as a Nash equilibrium. If you interact with such a partner then you maximize your own payoff by full cooperation. Therefore, partners resolve social dilemmas. Here we characterize partners among those strategies that react to an opponent's behavior during the last n interactions. Such players can sustain cooperation in an equilibrium, even if their opponent uses a longer memory strategy.

Author affiliations: ^aMax Planck Research Group on the Dynamics of Social Behavior, Max Planck Institute for Evolutionary Biology, Plön, Germany; ^bMathematics Department, The City College of New York, New York City, USA; ^cDepartment of Mathematics, Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, USA

N.G., E.A., M.N., and C.H. designed research; N.G. and C.H. performed research; N.G., analyzed data; and N.G., E.A., M.A. and C.H. wrote the paper.

The authors declare no competing interest.

¹To whom correspondence should be addressed. E-mail: glynatsi@evolbio.mpg.de

of strategies, and the time it takes to compute their payoffs, increases dramatically in n . For example, for memory-1, there are $2^4 = 16$ deterministic strategies (strategies that do not randomize between different actions). When both players adopt memory-1 strategies, computing their payoffs requires the inversion of a 4×4 matrix (9). After increasing the memory length to memory-2, there are $2^{16} = 64,536$ deterministic strategies, and payoffs now require the inverse of a 16×16 matrix.

There have been various approaches to tackle this problem. Some studies describe the strengths of particular strategies with more than one-round memory (57–60). Others explore the properties of entire strategy classes, such as ‘zero-determinant strategies’ (61, 62) or ‘reactive learning strategies’ (19). Stewart and Plotkin (63) characterize a set of memory- n strategies that is evolutionary robust. They show that for larger n , the volume of robust cooperative strategies exceeds the volume of strategies that lead to mutual defection. However, they do not provide an explicit description of the memory- n Nash equilibria. We give a more detailed account of these approaches in our SI.

To make further progress, we focus on an easy-to-interpret subset of memory- n strategies, the *reactive- n* strategies. Capturing the basic premise of conditional cooperation, they only depend on the co-player’s actions during the last n rounds (Fig. 1C,E). We show that within the reactive- n strategies, an explicit characterization of all Nash equilibria becomes feasible. Our results rely on a central insight, motivated by previous work of Press and Dyson (25): if one player adopts a reactive- n strategy, the other player can always find a best response among the deterministic *self-reactive- n* strategies. Self-reactive- n strategies are remarkably simple. They only depend on the player’s own previous n moves (Fig. 1D,F). Based on this insight, we study all reactive- n strategies that sustain full cooperation in a Nash equilibrium (the so-called *partner strategies*). We provide a full characterization for $n = 2$ and $n = 3$. Even stronger results are feasible when we restrict attention to so-called *counting strategies*. Such strategies only react to how often the co-player has cooperated in the last n rounds (irrespective of the exact timing of cooperation). For the donation game, we characterize the partners among the counting strategies for arbitrary n . The resulting conditions are straightforward to interpret: For every defection of the co-player in memory, the focal player’s cooperation rate needs to drop by $c/(nb)$. To further assess the relevance of partner strategies for the evolution of cooperation, we conduct extensive simulations for $n \in \{1, 2, 3\}$. Our findings indicate that the evolutionary process strongly favors partner strategies, and that these strategies are crucial for cooperation.

Overall, our results provide important insights into the logic of conditional cooperation when players have more than one-round memory. We show that partner strategies exist for all repeated prisoner’s dilemmas and for all memory lengths. These findings also allow us to reinterpret existing results on strategies with shorter memory. For example, we find that the well-known strategy Generous Tit-for-Tat (GTFT, see Ref. 64, 65) is just one instance of a more general strategy class. The same principles that make GTFT sustain cooperation within the reactive-1 strategies, allow us to construct partners within the reactive- n strategies.

Results

Model and notation. We consider a repeated game between two players, player 1 and player 2. Each round, players can choose to cooperate (C) or to defect (D). If both players cooperate, they receive the reward R , which exceeds the (punishment) payoff P for mutual defection. If only one player defects, the defecting player receives the temptation T , whereas the cooperator ends up with the sucker’s payoff S . We assume payoffs satisfy the typical relationships of a prisoner’s dilemma, $T > R > P > S$ and $2R > T + S$. Therefore, in each round, mutual cooperation is the best outcome for the pair, but players have some incentive to defect. The players’ aim is to maximize their average payoff per round, across infinitely many rounds. To make results easier to interpret, it is sometimes instructive to look at a particular variant of the prisoner’s dilemma, the donation game. Here, cooperation means to pay a cost $c > 0$ for the co-player to get a benefit $b > c$. The resulting payoffs are $R = b - c$, $S = -c$, $T = b$, $P = 0$. For simplicity, we focus on the donation game in the following. However, most of our findings are straightforward to extend to the general prisoner’s dilemma (or to other repeated 2×2 games, see SI).

We consider players who use strategies with finite memory. To describe such strategies formally, we introduce some notation. The last n actions of each player $i \in \{1, 2\}$ are referred to as the player’s *n -history*. We write this n -history as a tuple $\mathbf{h}^i = (a_{-n}^i, \dots, a_{-1}^i) \in \{C, D\}^n$. Each entry a_{-k}^i corresponds to player i ’s action k rounds ago. We use H^i for the set of all n -histories. This set contains $|H^i| = 2^n$ elements. Based on this notation, we can define a *reactive- n strategy* for player 1 as a vector $\mathbf{p} = (p_{\mathbf{h}})_{\mathbf{h} \in H^2} \in [0, 1]^{2^n}$. The entries $p_{\mathbf{h}}$ correspond to player 1’s cooperation probability in any given round, contingent on player 2’s actions during the last n rounds. The strategy is called pure or deterministic if any entry is either zero or one. We note that the above definition leaves player 1’s moves during the first n rounds unspecified. However, in infinitely repeated games without discounting, these initial moves tend to be inconsequential. Hence, we neglect them in the following.

For $n = 1$, we recover the classical format of reactive-1 strategies (9), $\mathbf{p} = (p_C, p_D)$. Here, p_C and p_D are the player’s cooperation probability given that the co-player cooperated or defected in the previous round, respectively. This set contains, for example, the strategies of unconditional defection, ALLD = (0, 0), and Tit-for-Tat, TFT = (1, 0). The next complexity class is the set of reactive-2 strategies, $\mathbf{p} = (p_{CC}, p_{CD}, p_{DC}, p_{DD})$. In addition to ALLD and TFT, this set contains, for instance, the strategies Tit-for-Two-Tat, TF2T = (1, 1, 1, 0) and Two-Tit-for-Tat, 2TFT = (1, 0, 0, 0). Similar examples exist for $n > 2$. When both players adopt reactive- n strategies (or more generally, memory- n strategies), it is straightforward to compute their expected payoffs, by representing the game as a Markov chain. The respective procedure is described in the SI.

Herein, we are particularly interested in those reactive- n strategies that sustain full cooperation. Such strategies ought to have two properties. First, they ought to be *nice*, meaning that they are never the first to defect (7). This property ensures that two players with nice strategies fully cooperate. In particular, if \mathbf{h}_C is a co-player’s n -history that consists of n bits of cooperation, a nice strategy needs to respond by

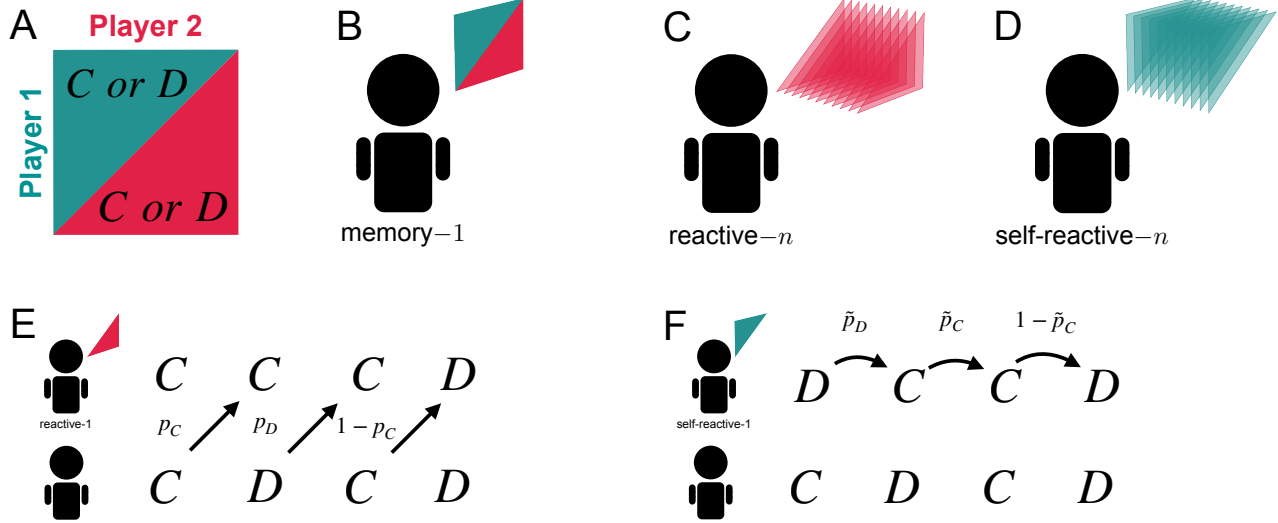


Fig. 1. The repeated prisoner's dilemma among players with finite memory. **A**, In the repeated prisoner's dilemma, in each round two players independently decide whether to cooperate (C) or to defect (D). **B**, When players adopt memory-1 strategies, their decisions depend on the entire outcome of the previous round. That is, they consider both their own and the co-player's previous action. **C**, When players adopt a reactive- n strategy, they make their decisions based on the co-player's actions during the past n rounds. **D**, A self-reactive- n strategy is contingent on the player's own actions during the past n rounds. **E**, To illustrate these concepts, we show a game between a player with a reactive-1 strategy (top) and an arbitrary player (bottom). Reactive-1 strategies can be represented as a vector $\mathbf{p} = (p_C, p_D)$. The entry p_C is the probability of cooperating given the co-player cooperated in the previous round. The entry p_D is the cooperation probability after the co-player defected. **F**, Now, the top player adopts a self-reactive-1 strategy, $\tilde{\mathbf{p}} = (\tilde{p}_C, \tilde{p}_D)$. Here, the player's cooperation probability depends on its own previous action.

cooperating with certainty, $p_{hC} = 1$. Second, the strategy ought to form a *Nash equilibrium*, such that no co-player has an incentive to deviate. Strategies that have both properties are called *partner strategies* (66) or *partners*. The partners among the reactive-1 strategies are well known. For the donation game, partners are those strategies with $p_C = 1$ and $p_D \leq 1 - c/b$ (29). However, a general theory of partners for $n \geq 2$ is lacking. This is what we aim to derive in the following. In the main text, we provide the main intuition for our results; all proofs are in the SI.

An algorithm to identify partners among reactive- n strategies. It is comparably easy to verify whether a reactive- n strategy \mathbf{p} is nice. Demonstrating that the strategy is also a Nash equilibrium, however, is far less trivial. In principle, this requires uncountably many payoff comparisons. We would have to show that if player 2's strategy is fixed to \mathbf{p} , no other strategy σ for player 1 can result in a higher payoff. That is, player 1's payoff needs to satisfy $\pi^1(\sigma, \mathbf{p}) \leq \pi^1(\mathbf{p}, \mathbf{p})$ for all σ . Fortunately, this task can be simplified considerably. Already Press & Dyson (25) showed that it is sufficient to test only those σ with at most n rounds of memory. Based on two insights, we can even further restrict the search space of strategies σ that need to be tested.

First, suppose player 1 uses some arbitrary strategy σ against player 2 with reactive- n strategy $\mathbf{p} = (p_h)_{h \in H^1}$. Then we prove that instead of σ , player 1 may switch to a *self-reactive- n* strategy $\tilde{\mathbf{p}}$ without changing either player's payoffs. When adopting a self-reactive strategy, player 1 only takes into account her own actions during the last n rounds, $\tilde{\mathbf{p}} = (\tilde{p}_h)_{h \in H^1}$. In particular, if σ is a best response to \mathbf{p} , then there is an associated self-reactive strategy $\tilde{\mathbf{p}}$ that is also a best response. This result follows the same intuition as a similar result of Press & Dyson (25): if there is a part of the joint history that player 2 does not take into account, player 1

gains nothing by considering that part of the history. In our case, because player 2 only considers the last n actions of player 1, it is sufficient for player 1 to do the same. **Fig. 2A,B** provides an illustration. There, we depict a game in which player 1 adopts a memory-1 strategy against a reactive-1 opponent. Due to the above result, we can find an equivalent self-reactive-1 strategy for player 1. While that self-reactive strategy is simpler, on average it induces the same game dynamics. Hence, it results in identical payoffs.

The above result guarantees that for any reactive- n strategy, there is always a best response among the self-reactive- n strategies. In a second step, we prove that such a best response can always be found among the *deterministic* self-reactive- n strategies. This further reduces the search space for best responses, from an uncountable set to a finite set of size 2^{2^n} . For $n=2$, this leaves us with 16 self-reactive strategies to test. For $n=3$, we end up with (at most) 256 strategies. While this may still appear to be a large number, many of the different strategies impose redundant constraints on partner strategies. This redundancy further reduces the number of conditions a partner needs to satisfy.

Partners among the reactive-2 and the reactive-3 strategies.

To illustrate the above algorithm, we first characterize the partners among the reactive-2 strategies. To this end, we note that it is straightforward to compute the payoff of a specific self-reactive-2 strategy against a general reactive-2 strategy \mathbf{p} (see SI). By computing the payoffs of all 16 pure self-deterministic strategies $\tilde{\mathbf{p}}$, and by requiring $\pi^1(\tilde{\mathbf{p}}, \mathbf{p}) \leq \pi^1(\mathbf{p}, \mathbf{p})$ for all of them, we end up with only three conditions. Specifically, we prove that \mathbf{p} is a partner if and only if

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

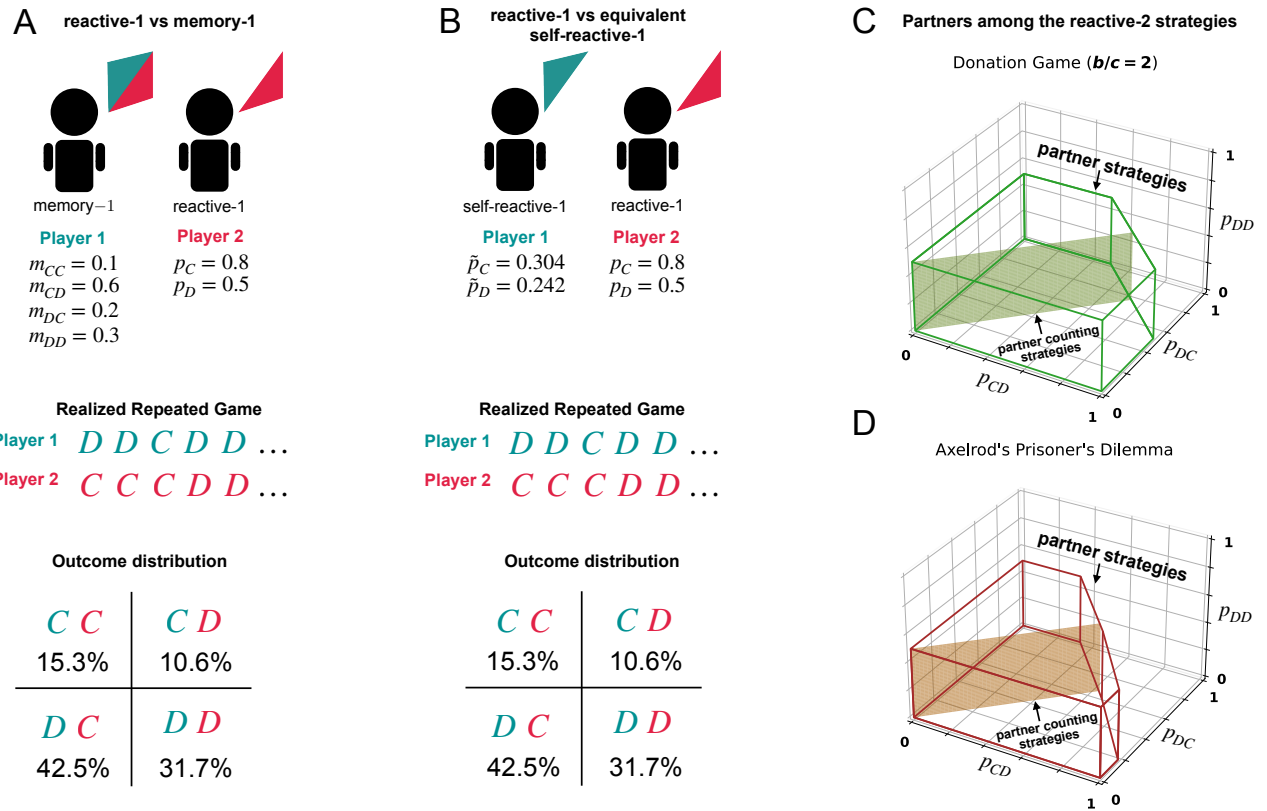


Fig. 2. Characterizing partners among the reactive- n strategies. A,B, To characterize the reactive- n partner strategies, we prove the following result. Suppose the focal player adopts a reactive- n strategy. Then, for any strategy of the opponent (with arbitrary memory), one can find an associated self-reactive- n strategy that yields the same payoffs. Here, we show an example. Player 1 uses a reactive-1 strategy against player 2 with a memory-1 strategy. Our result implies that player 2 can switch to a well-defined self-reactive-1 strategy. This switch leaves the outcome distribution unchanged. In both cases, players are equally likely to experience mutual cooperation, unilateral cooperation, or mutual defection in the long run. C, Based on this insight, we can explicitly characterize the reactive-2 partner strategies (with $p_{CC} = 1$). Here, we represent the corresponding conditions in Eq. (1) for a donation game with $b/c = 2$. Among the reactive-2 strategies, the counting strategies correspond to the subset with $p_{CD} = p_{DC}$. Counting strategies only depend on how often the co-player cooperated in the past, not on the timing of cooperation. D, Similarly, we can also characterize the reactive-2 partner strategies for the general prisoner's dilemma. Here, we use the payoff matrix of Axelrod (7).

The above conditions define a three-dimensional polyhedron within the space of all nice reactive-2 strategies (Fig. 2C). The condition $p_{CC} = 1$ follows from the requirement that the strategy ought to be nice. As long as the co-player cooperates, the reactive- n player goes along. The other two conditions imply that for each defection in memory, the player's cooperation rate decreases proportionally. Interestingly, in cases with a mixed 2-history (one cooperation, one defection), the above conditions suggest that the exact timing of cooperation does not matter. It is only required that the two cooperation probabilities p_{CD} and p_{DC} are sufficiently small *on average*. Notably, the above conditions also imply that to check whether a given reactive-2 strategy is a partner, it suffices to check two deviations. These deviations are the strategy that strictly alternates between cooperation and defection (yielding the first inequality), and ALLD (yielding the second inequality) (Fig. 3). We note that this last implication is specific to the donation game. For the general prisoner's dilemma (depicted in Fig. 2D), there are more than two inequalities that need to be satisfied (see SI).

Analogously, we can also characterize the partners among the reactive-3 strategies. A reactive-3 strategy can be represented by a vector $\mathbf{p} = (p_{CCC}, p_{CCD}, p_{CDC}, p_{CDD}, p_{DCC}, p_{DCD}, p_{DDC}, p_{DDD})$. It

is a partner strategy if and only if

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

These conditions follow a similar logic as in the previous case with $n = 2$. For every co-player's defection in memory, the respective cooperation probability needs to be diminished proportionally. As an example, the second inequality in Eq. (2) considers three sequences *CCD*, *CDC*, *DCC*. The proportion of *D*'s across these three sequences is $1/3$. Hence, the threshold on the right hand side is $1 - 1/3 \cdot c/b$. Because reactive-2 strategies are a subset of reactive-3 strategies, we

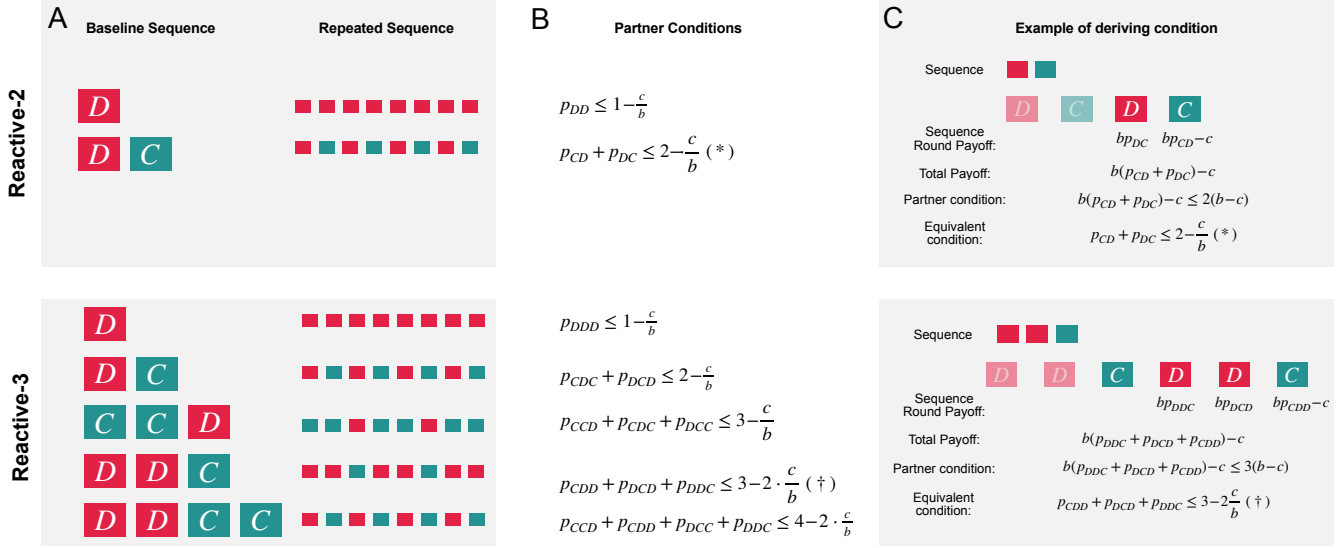


Fig. 3. Conditions for partners among reactive-2 and reactive-3 strategies. **A**, Pure self-reactive strategies generate simple repetitive sequences of actions that are independent of the co-player. For example, in the case of $n = 2$, the pure self-reactive strategy $\tilde{\mathbf{p}} = (0, 1)$ generates the indefinitely repeated alternating sequence DC . **B**, For a nice reactive strategy \mathbf{p} to be a partner, all of these self-reactive strategies need to achieve at most the mutual cooperation payoff against \mathbf{p} . This leads to necessary conditions for \mathbf{p} to be a partner, which we show here for $n = 2$, and $n = 3$. Interestingly, we prove that these necessary conditions are also sufficient, see **SI**. **C**, To derive the conditions, we consider the average payoff of each repetitive sequence. In the top panel, we illustrate an example for $n = 2$. Here, the repetitive sequence DC plays against the reactive strategy $\mathbf{p} = (1, p_{CD}, p_{DC}, p_{DD})$. In odd rounds, the sequence player receives a benefit b with probability p_{DC} , without paying any cost. In even rounds, the player receives the benefit b with probability p_{CD} , while paying a cost c . Over the course of two consecutive rounds, the player thus receives $(p_{DC} + p_{CD})b - c$. This payoff needs to be smaller or equal than what a partner strategy achieves against itself, which is $2(b - c)$. This leads to condition $(*)$. In the bottom panel, we illustrate a similar example for $n = 3$, explaining condition (\dagger) .

can also derive the conditions in Eq. (1) as a special case of Eq. (2) (see **SI**).

Moreover, the above conditions imply that to check whether a given reactive-3 strategy is a partner, it suffices to check five deviations. Similarly to the previous case, two of these deviations include the strategy that strictly alternates between cooperation and defection, and ALLD. The remaining conditions arise from deviations towards self-reactive strategies that repeat certain sequences, where the sequences are CCD , CDD , and $CCDD$ (**Fig. 3**).

For $n = 3$, there are now more conditions to consider than in the previous case. These conditions become even more complex for the general prisoner's dilemma. Given these complexities, we do not present conditions for reactive- n partner strategies beyond $n = 3$, even though the algorithm presented in the previous section still applies.

Partners among the reactive- n counting strategies. We can more easily generalize these formulas to arbitrary n if we further restrict the strategy space. In the following, we consider reactive- n counting strategies. These strategies only depend on how often the co-player cooperated during the past n rounds; they do not take into account in which of the past n rounds the co-player cooperated. We represent such strategies as a vector $\mathbf{r} = (r_i)_{i \in \{n, n-1, \dots, 0\}}$. Each entry r_i indicates the player's cooperation probability if the co-player cooperated i times during the last n rounds. We note that although reactive- n counting strategies have fewer entries (bits) than reactive- n strategies, they are equally complex in terms of their memory requirements. Even a player with a reactive- n counting strategy needs to keep a record of the exact sequence of the opponent's last n actions. Only by doing so, the player can update its opponent's cooperation

count each round, by discarding the opponent's oldest action in memory (see also **SI**). Any reactive-1 strategy $\mathbf{p} = (p_C, p_D)$ is a counting strategy by definition. However, for larger n , the set of counting strategies is a strict subset of the reactive- n strategies. For example, for $n = 2$, counting strategies are those strategies that satisfy $p_{CD} = p_{DC} =: r_1$. As a result, the partners among the counting strategies form a 2-dimensional plane within the 3-dimensional polyhedron of reactive-2 partner strategies (**Fig. 2C,D**).

For the donation game, it is possible to characterize the set of partner strategies for arbitrary n . We find that a counting strategy \mathbf{r} is a partner if and only if

$$r_n = 1 \quad \text{and} \quad r_{n-k} \leq 1 - \frac{k}{n} \cdot \frac{c}{b} \quad \text{for } k \in \{1, 2, \dots, n\}. \quad [3]$$

That is, for every defection of the opponent in memory, the maximum cooperation probability needs to be reduced by $c/(nb)$. It is worth to highlight that this result is general. These strategies are Nash equilibria even if players are allowed to deviate towards strategies that do not merely count the co-player's cooperative acts, or towards strategies that take into account more than the last n rounds.

Further analytical results. In the **SI**, we use our framework to derive a number of additional results. Here we provide a brief summary. First, instead of partners we can equally use our formalism to characterize 'defector strategies' – reactive- n strategies that lead to stable mutual defection. Following Stewart and Plotkin (63), we can use this characterization to compare the relative volume of partners and defectors. We find that for sufficiently small cost-to-benefit ratios, the set of partners has the larger volume. Moreover, the relative volume of partners increases in n , both for reactive- n

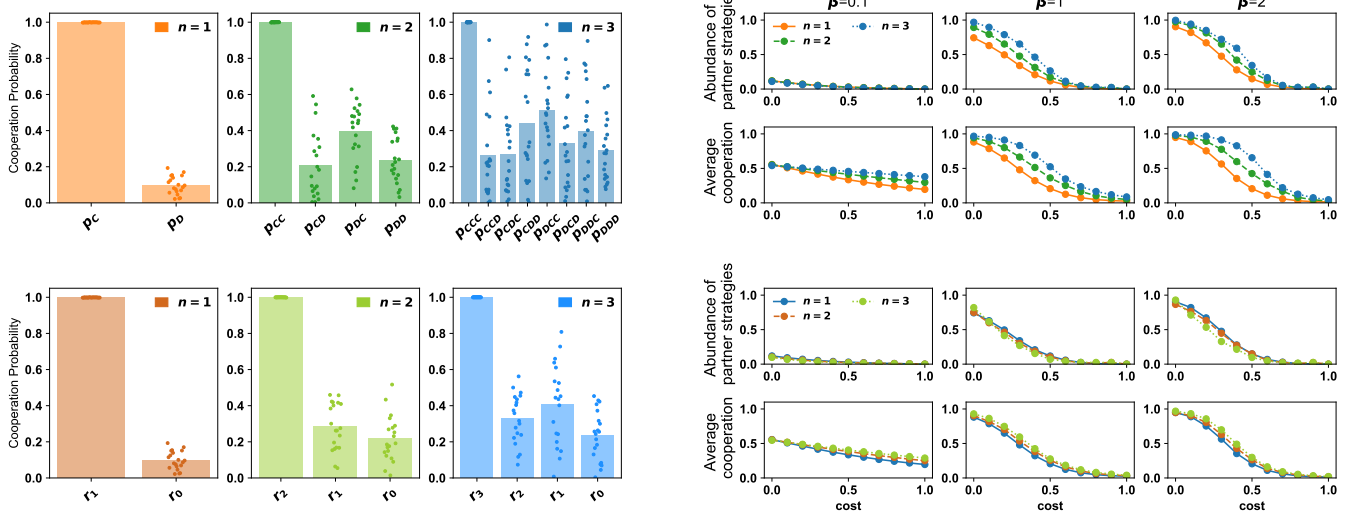


Fig. 4. Evolutionary dynamics of reactive- n strategies. To explore the evolutionary dynamics among reactive- n strategies, we run simulations based on the method of Imhof and Nowak (68). This method assumes rare mutations. Every time a mutant strategy appears, it goes extinct or fixes before the arrival of the next mutant strategy. **A,B**, We run twenty independent simulations for reactive- n strategies and for reactive- n counting strategies. For each simulation, we record the most abundant strategy (the strategy that resisted most mutants). The respective average cooperation probabilities are in line with the conditions for partner strategies. **C,D**, With additional simulations, we explore the average abundance of partner strategies and the population's average cooperation rate. For a given resident strategy to be classified as a partner by our simulation, it needs to satisfy all inequalities in the respective characterization. In addition, it needs to cooperate after full cooperation with a probability of at least 95%. For all considered parameter values, we only observe high cooperation rates when partner strategies evolve. Simulations are based on a donation game with $b = 1$, $c = 0.5$, a selection strength $\beta = 1$ and a population size $N = 100$, unless noted otherwise. For n equal to 1 and 2, simulations are run for 10^7 time steps. For $n = 3$ we use $2 \cdot 10^7$ time steps.

and for reactive- n counting strategies (Fig. S3, Tab. S2). This finding has interesting implications for the evolution of cooperation (28, 63). If evolutionary processes generate mutant strategies at random, larger memory lengths make it increasingly likely that mutants adopt strategies in the vicinity of partner strategies, compared to defectors.

Second, we use our formalism to explore the effects of implementation errors (33). When such an error occurs (with some exogenous probability ε), players implement the opposite action of what they intend to do. For this scenario, we derive two sets of results. First, we assume errors to be vanishingly rare. In that case, we find that almost all our previously-described partner strategies remain approximate Nash equilibria. That is, even if there are profitable deviations, the respective payoff advantage is guaranteed to be arbitrarily small, see Fig. S2. Second, we derive a result for donation games when the error rate is strictly positive. In that case, we describe a subset of so-called 'equalizer' strategies (67), which can sustain cooperation in equilibrium. Among the reactive-1 strategies, this set includes a single strategy, Generous Tit-for-Tat (GTFT). For $n \geq 2$, we derive additional variants of GTFT, which punish defection with some delay. While we characterize the set of reactive- n equalizers explicitly, it remains an open question whether they are the only partner strategies for positive error rates.

Evolutionary Dynamics. With our previous equilibrium analysis we have identified the strategies that can sustain cooperation in principle. In a next step, we determine whether these strategies evolve in the first place. Here, we no longer presume that individuals would play equilibrium strategies. Rather they initially implement some random behavior. Over time, they adapt their strategies based on social learning. To model this learning process, we consider a population of

individuals who update their strategies based on pairwise comparisons. The efficacy of the resulting learning process is determined by a strength of selection parameter β . The larger β , the more likely individuals imitate strategies with a higher payoff. In addition, mutations occasionally introduce new strategies. We describe the exact setup of this learning process in the **Material and Methods** section. As we explain there, the process is particularly easy to explore when mutations are rare (68–71). In that case, the population is typically homogeneous, such that all players adopt the same (resident) strategy. Once a new mutant strategy appears, this strategy fixes or goes extinct before the next mutation happens. Evolutionary processes with rare mutations can be simulated more efficiently because there is an explicit formula for the mutant's fixation probability (72).

The results of these simulations are shown in Fig. 4. First, we explore which reactive- n strategies evolve for a fixed set of game parameters. Here, we vary the strategies' memory length n , and whether mutations introduce all reactive- n strategies, or counting strategies only. For twenty independent simulations, Fig. 4A,B displays the most abundant strategy for each simulation run (those are the strategies that prevent the largest number of mutants from taking over). We note that all the shown strategies show behavior consistent with our characterization of partners: If a co-player fully cooperated in the previous n rounds, these strategies prescribe to continue with cooperation. If the co-player defected, however, they cooperate with a markedly reduced cooperation probability that satisfies the constraints in Eqs. (1) – (3).

Interestingly, however, the evolving strategies exhibit an interesting asymmetry. For example, for reactive-2 strategies, we observe that players' strategies tend to satisfy $p_{CD} < p_{DC}$. That is, they are more likely to defect if their opponent

defected last round, rather than two rounds ago. In light of our equilibrium analysis, this result is surprising. After all, according to our partner condition Eq. (1), the two cooperation probabilities are completely interchangeable. This asymmetry arises because our evolutionary process with uniform mutations does not introduce perfect partner strategies (with $p_{CC}=1$). Rather, it introduces strategies in the respective neighborhood (with, say, $p_{CC}=0.99$). Among these noisy partner strategies, we show that strategies are more resilient when they punish defection without delay (for more details, see SI and Figs. S4, S5, S6 Tab. S3).

In a next step, we systematically explore the impact of several key parameters: the cost-to-benefit ratio c/b , the selection strength β , and the memory length n (Fig. 4C,D). In addition, we vary the error rate ε in Fig. S7. In each case, we record how these parameters affect the abundance of partner strategies and the population's average cooperation rate. Overall, the effect of each parameter is as expected. In particular, interactions are most cooperative when cooperation is comparably cheap. This effect is magnified for stronger selection strengths. Two results, however, are particularly noteworthy. First, the curves representing evolving cooperation rates align with the prevalence of partner strategies. This observation suggests that partner strategies are indeed crucial for the evolution of cooperation. Second, the positive effects of larger memory are most pronounced for reactive- n strategies. In contrast, for counting strategies any positive effect of increasing n is considerably dampened.

We repeat these simulations for the more general sets of memory- n strategies and memory- n counting strategies (see SI). Again, among memory- n strategies, larger values of n lead to more cooperation. But even among counting strategies, longer memory has a positive, albeit smaller, effect (Fig. S6). We conclude for the considered strategy spaces that the timing of cooperation can be important, even in additive games such as the donation game.

Discussion

Direct reciprocity is a key mechanism for cooperation, based on the intuition that individuals are more likely to cooperate when they meet repeatedly (8). To capture the logic of reciprocity, most previous theoretical studies focus on a subset of strategies, the memory-1 strategies (21–31). This set is comparably easy to work with: the number of deterministic memory-1 strategies is manageable; most strategies are easy to interpret; and payoffs can be computed efficiently (9). At the same time, however, this strategy space leaves out many interesting reciprocal behaviors that are of theoretical or empirical relevance. For example, already simple behaviors such as Tit-for-Two-Tat (7) are not representable with one-round memory. This shortcoming is particularly consequential for noisy games, where higher-memory strategies are important (54). In such games, individuals often take into account information from previous rounds to make sense of a co-player's defection in the last round. That is, the earlier history of play provides an important context to interpret the co-player's last-round behavior.

To make progress, we consider an easily interpretable set of strategies with higher memory. These reactive- n strategies take into account a co-player's moves during the past n rounds. They capture the basic idea of conditional cooperation: people

are responsive to the previous actions of their interaction partners. For reactive- n strategies, we derive a convenient method to characterize all 'partner strategies' – strategies that sustain full cooperation in a Nash equilibrium (29, 66). We show that for a reactive- n strategy to be a Nash equilibrium, it is not necessary to check all possible deviations. It suffices to only check deviations towards (deterministic) self-reactive- n strategies. Self-reactive players are particularly simple to describe. They only take into account their own previous moves. In particular, the future behavior of a self-reactive player is independent of the co-player. We use this insight to characterize the reactive- n partner strategies (and the defector strategies) in the repeated prisoner's dilemma. But the same insights should be applicable to other contexts. For example, we expect that similar techniques can be used to characterize the equilibria of other repeated games, such as the snowdrift game (73) or the volunteer's dilemma (74). In this way, some of our technical results represent useful tools to make further progress on the theory of repeated games, similar to Press and Dyson's insight that any memory-1 strategy has a memory-1 best response (25).

Especially for small memory lengths, the conditions for partner strategies are intuitive. For example, for the donation game with $n=2$ rounds of memory, we end up with three conditions, see Eq. (1). (i) If the co-player cooperated twice, continue to cooperate; (ii) If the co-player cooperated once, cooperate with a slightly reduced probability of $1-c/(2b)$ on average. (iii) If the co-player did not cooperate at all, reduce the cooperation probability even further, to $1-c/b$. As we increase the memory length to $n \geq 3$, or as we consider more general games, there are more conditions to satisfy, and the conditions become harder to interpret. However, the three simple conditions do generalize to larger n if we focus on the set of counting strategies. These are the reactive- n strategies that react to how often the co-player cooperated during the last n rounds. For counting strategies, we show that for each defection of the co-player in memory, a partner reduces its cooperation probability by $c/(nb)$. A partner's generosity decreases in proportion to their opponent's selfishness.

While in practice, people's cooperative decisions often depend on the outcome of their last encounter, they rarely depend on that last encounter *only*. Overall, our results suggest a way how individuals can integrate information from previous interactions to cooperate most effectively.

Materials and Methods

Our study combines two independent approaches, an equilibrium analysis and evolutionary simulations.

Equilibrium analysis. Here we only summarize our approach; all details are in the SI. There, we formally introduce the three relevant strategy spaces, memory- n strategies, reactive- n strategies, and self-reactive- n strategies. Then we provide an explicit algorithm for computing these strategies' payoffs. This algorithm uses a Markov chain approach. The states of the Markov chain are the possible combinations of n -histories of the two players. Given the players' current n -histories and their strategies, we can compute the likelihood of observing each possible state one round later.

In a second step, we explore the partner strategies among the reactive- n strategies. To this end, we first generalize

some well-known reactive-1 partner strategies: Tit-for-Tat (7) and Generous Tit-for-Tat (64, 65). In a next step, we derive a general algorithm to check whether a given reactive- n strategy is a partner. We use this algorithm to characterize all reactive- n partners for $n \in \{1, 2, 3\}$, for both the donation game and the prisoner's dilemma. For counting strategies in the donation game, we characterize partners for all n .

Evolutionary analysis. For our simulations, we consider a population of size N . Initially all members are of the same strategy (in our case, they are unconditional defectors). In each elementary time step, one individual switches to a new mutant strategy. The mutant strategy is generated by independently drawing each individual cooperation probability from the unit interval $[0, 1]$ uniformly at random. If the mutant strategy yields a payoff of $\pi_{M,k}$, where k is the number of mutants in the population, and if residents get a payoff of $\pi_{R,k}$, then the fixation probability ϕ_M of the mutant strategy can be calculated explicitly (72),

$$\phi_M = \left(1 + \sum_{i=1}^{N-1} \prod_{j=1}^i e^{-\beta(\pi_{M,j} - \pi_{R,i})}\right)^{-1}. \quad [4]$$

The parameter $\beta \geq 0$ reflects the strength of selection. It measures the importance of relative payoff advantages for the evolutionary success of a strategy. When β is small, $\beta \approx 0$, payoffs become irrelevant, and a strategy's fixation probability approaches $\phi_M \approx 1/N$. The larger the value of β , the more strongly the evolutionary process favors the fixation of strategies with a high payoff. Depending on ϕ_M , the mutant either fixes (becomes the new resident) or goes extinct. Afterwards, the process repeats, and another mutant strategy is introduced to the population.

We iterate this elementary population updating process for a large number of mutant strategies. At each step, we record the current resident strategy and the resulting average cooperation rate, indicating how often the resident strategy cooperates with itself. Additionally, we assess how many resident strategies qualify as partner strategies in our simulation. For a resident strategy to be classified as a partner, it must satisfy all inequalities in the respective definition of partner strategies. In addition, it must cooperate with a probability of at least 95% after full cooperation.

Data, Materials, and Software Availability. The source code used to reproduce the results of this study is available on the online GitHub repository: [Nikoleta-v3/conditional-cooperation-with-longer-memory](https://github.com/Nikoleta-v3/conditional-cooperation-with-longer-memory). The simulation data have been archived on Zenodo and can be found at: zenodo.org/records/10605988.

Acknowledgements. N.G. and C.H. acknowledge generous support by the European Research Council Starting Grant 850529: E-DIRECT, and by the Max Planck Society.

1. AP Melis, D Semmann, How is human cooperation different? *Philos. Transactions Royal Soc. B* **365**, 2663–2674 (2010).
2. DG Rand, MA Nowak, Human cooperation. *Trends Cogn. Sci.* **117**, 413–425 (2012).
3. WS Neilson, The economics of favors. *J. Econ. Behav. & Organ.* **39**, 387–397 (1999).
4. U Fischbacher, S Gächter, Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *Am. economic review* **100**, 541–556 (2010).
5. C Hilbe, T Röhl, M Milinski, Extortion subdues human players but is finally punished in the prisoner's dilemma. *Nat. Commun.* **5**, 3976 (2014).
6. B Xu, Y Zhou, JW Lien, J Zheng, Z Wang, Extortion can outperform generosity in iterated prisoner's dilemma. *Nat. Commun.* **7**, 11125 (2016).

7. R Axelrod, WD Hamilton, The evolution of cooperation. *Science* **211**, 1390–1396 (1981).
8. MA Nowak, Five rules for the evolution of cooperation. *Science* **314**, 1560–1563 (2006).
9. K Sigmund, *The calculus of selfishness*. (Princeton University Press), (2010).
10. J García, M van Veelen, No strategy can win in the repeated prisoner's dilemma: Linking game theory and computer simulations. *Front. Robotics AI* **5**, 102 (2018).
11. C Hilbe, K Chatterjee, MA Nowak, Partners and rivals in direct reciprocity. *Nat. Hum. Behav.* **2**, 469–477 (2018).
12. C Rossetti, C Hilbe, Direct reciprocity among humans. *Ethology* **130**, e13407 (2023).
13. MR Frean, The prisoner's dilemma without synchrony. *Proc. Royal Soc. B* **257**, 75–79 (1994).
14. T Killingback, M Doebeli, N Knowlton, Variable investment, the continuous prisoner's dilemma, and the origin of cooperation. *Proc. Royal Soc. B* **266**, 1723–1728 (1999).
15. C Hauert, O Stenull, Simple adaptive strategy wins the prisoner's dilemma. *J. Theor. Biol.* **218**, 261–72 (2002).
16. S Kurokawa, Y Ihara, Emergence of cooperation in public goods games. *Proc. Royal Soc. B* **276**, 1379–1384 (2009).
17. FL Pinheiro, VV Vasconcelos, FC Santos, JM Pacheco, Evolution of all-or-none strategies in repeated public goods dilemmas. *PLoS Comput. Biol.* **10**, e1003945 (2014).
18. 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).
19. A McAvoy, MA Nowak, Reactive learning strategies for iterated games. *Proc. Royal Soc. A* **475**, 20180819 (2019).
20. DP Kraines, VY Kraines, Pavlov and the prisoner's dilemma. *Theory Decis.* **26**, 47–79 (1989).
21. 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).
22. LA Imhof, D Fudenberg, MA Nowak, Evolutionary cycles of cooperation and defection. *Proc. Natl. Acad. Sci. USA* **102**, 10797–10800 (2005).
23. J Grujic, JA Cuesta, A Sanchez, On the coexistence of cooperators, defectors and conditional cooperators in the multiplayer iterated prisoner's dilemma. *J. Theor. Biol.* **300**, 299–308 (2012).
24. S van Segbroeck, JM Pacheco, T Lenaerts, FC Santos, Emergence of fairness in repeated group interactions. *Phys. Rev. Lett.* **108**, 158104 (2012).
25. WH Press, FJ Dyson, Iterated prisoner's dilemma contains strategies that dominate any evolutionary opponent. *Proc. Natl. Acad. Sci.* **109**, 10409–10413 (2012).
26. AJ Stewart, JB Plotkin, From extortion to generosity, evolution in the iterated prisoner's dilemma. *Proc. Natl. Acad. Sci. USA* **110**, 15348–15353 (2013).
27. DFP Toupou, DG Rand, SH Strogatz, Limit cycles sparked by mutation in the repeated prisoner's dilemma. *Int. J. Bifurc. Chaos* **24**, 2430035 (2014).
28. AJ Stewart, JB Plotkin, Collapse of cooperation in evolving games. *Proc. Natl. Acad. Sci. USA* **111**, 17558–17563 (2014).
29. 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).
30. NE Glynnatsi, VA Knight, Using a theory of mind to find best responses to memory-one strategies. *Sci. Reports* **10**, 1–9 (2020).
31. X Chen, F Fu, Outlearning extortions: Unbending strategies can foster reciprocal fairness and cooperation. *PNAS nexus* **2**, pgad176 (2023).
32. M Kleiman-Weiner, MK Ho, JL Austerweil, ML Littman, JB Tenenbaum, Coordinate to cooperate or compete: abstract goals and joint intentions in social interaction in *CogSci*. (2016).
33. R Boyd, Mistakes allow evolutionary stability in the repeated Prisoner's Dilemma game. *J. Theor. Biol.* **136**, 47–56 (1989).
34. D Hao, Z Rong, T Zhou, Extortion under uncertainty: Zero-determinant strategies in noisy games. *Phys. Rev. E* **91**, 052803 (2015).
35. H Zhang, Errors can increase cooperation in finite populations. *Games Econ. Behav.* **107**, 203–219 (2018).
36. A Mamiya, G Ichinose, Zero-determinant strategies under observation errors in repeated games. *Phys. Rev. E* **102**, 032115 (2020).
37. AJ Stewart, JB Plotkin, The evolvability of cooperation under local and non-local mutations. *Games* **6**, 231–250 (2015).
38. A McAvoy, J Kates-Harbeck, K Chatterjee, C Hilbe, Evolutionary instability of selfish learning in repeated games. *PNAS nexus* **1**, pgac141 (2022).
39. K Brauchli, T Killingback, M Doebeli, Evolution of cooperation in spatially structured populations. *J. Theor. Biol.* **200**, 405–417 (1999).
40. G Szabó, T Antal, P Szabó, M Droz, Spatial evolutionary prisoner's dilemma game with three strategies and external constraints. *Phys. Rev. E* **62**, 1095–1103 (2000).
41. B Allen, MA Nowak, U Dieckmann, Adaptive dynamics with interaction structure. *Am. Nat.* **181**, E139–E163 (2013).
42. A Szolnoki, M Perc, Defection and extortion as unexpected catalysts of unconditional cooperation in structured populations. *Sci. Reports* **4**, 5496 (2014).
43. SK Baek, HC Jeong, C Hilbe, MA Nowak, Comparing reactive and memory-one strategies of direct reciprocity. *Sci. Reports* **6**, 1–13 (2016).
44. M Harper, et al., Reinforcement learning produces dominant strategies for the iterated prisoner's dilemma. *PLoS one* **12**, e0188046 (2017).
45. V Knight, M Harper, NE Glynnatsi, O Campbell, Evolution reinforces cooperation with the emergence of self-recognition mechanisms: An empirical study of strategies in the moran process for the iterated prisoner's dilemma. *PLoS one* **13**, e0204981 (2018).
46. P Duersch, J Oechssler, B Schipper, When is tit-for-tat unbeatable? *Int. J. Game Theory* **43**, 25–36 (2013).
47. J Engle-Warnick, RL Slonim, Inferring repeated-game strategies from actions: evidence from trust game experiments. *Econ. Theory* **28**, 603–632 (2006).
48. P Dal Bó, GR Fréchette, The evolution of cooperation in infinitely repeated games: Experimental evidence. *Am. Econ. Rev.* **101**, 411–429 (2011).

993	49. G Camera, M Casari, M Bigoni, Cooperative strategies in anonymous economies: An experiment. <i>Games Econ. Behav.</i> 75 , 570–586 (2012).	1055
994	50. L Bruttel, U Kamecke, Infinity in the lab. How do people play repeated games? <i>Theory Decis.</i> 72 , 205–219 (2012).	1056
995	51. E Montero-Porras, J Grujić, E Fernández Domingos, T Lenaerts, Inferring strategies from observations in long iterated prisoner's dilemma experiments. <i>Sci. Reports</i> 12 , 7589 (2022).	1057
996	52. J Romero, Y Rosokha, Constructing strategies in the indefinitely repeated prisoner's dilemma game. <i>Eur. Econ. Rev.</i> 104 , 185–219 (2018).	1058
997	53. M Kleiman-Weiner, JB Tenenbaum, P Zhou, Non-parametric bayesian inference of strategies in repeated games. <i>The Econom. J.</i> 21 , 298–315 (2018).	1059
998	54. D Fudenberg, DG Rand, A Dreber, Slow to anger and fast to forgive: Cooperation in an uncertain world. <i>Am. Econ. Rev.</i> 102 , 720–749 (2012).	1060
999	55. C Hauert, HG Schuster, Effects of increasing the number of players and memory size in the iterated prisoner's dilemma: a numerical approach. <i>Proc. Royal Soc. B</i> 264 , 513–519 (1997).	1061
1000	56. Y Murase, SK Baek, Grouping promotes both partnership and rivalry with long memory in direct reciprocity. <i>PLoS Comput. Biol.</i> 19 , e1011228 (2023).	1062
1001	57. S Do Yi, SK Baek, JK Choi, Combination with anti-tit-for-tat remedies problems of tit-for-tat. <i>J. Theor. Biol.</i> 412 , 1–7 (2017).	1063
1002	58. C Hilbe, LA Martinez-Vaquero, K Chatterjee, MA Nowak, Memory-n strategies of direct reciprocity. <i>Proc. Natl. Acad. Sci.</i> 114 , 4715–4720 (2017).	1064
1003	59. Y Murase, SK Baek, Five rules for friendly rivalry in direct reciprocity. <i>Sci. Reports</i> 10 , 16904 (2020).	1065
1004	60. J Li, et al., Evolution of cooperation through cumulative reciprocity. <i>Nat. Comput. Sci.</i> 2 , 677–686 (2022).	1066
1005	61. M Ueda, Memory-two zero-determinant strategies in repeated games. <i>Royal Soc. Open Sci.</i> 8 , 202186 (2021).	1067
1006	62. M Ueda, Controlling conditional expectations by zero-determinant strategies. <i>Oper. Res. Forum</i> 3 , 48 (2022).	1068
1007	63. AJ Stewart, JB Plotkin, Small groups and long memories promote cooperation. <i>Sci. Reports</i> 6 , 1–11 (2016).	1069
1008	64. MA Nowak, K Sigmund, Tit for tat in heterogeneous populations. <i>Nature</i> 355 , 250–253 (1992).	1070
1009	65. P Molander, The optimal level of generosity in a selfish, uncertain environment. <i>J. Confl. Resolut.</i> 29 , 611–618 (1985).	1071
1010	66. C Hilbe, A Traulsen, K Sigmund, Partners or rivals? Strategies for the iterated prisoner's dilemma. <i>Games economic behavior</i> 92 , 41–52 (2015).	1072
1011	67. MC Boerlijst, MA Nowak, K Sigmund, Equal pay for all prisoners. <i>Am. Math. Mon.</i> 104 , 303–307 (1997).	1073
1012	68. LA Imhof, MA Nowak, Stochastic evolutionary dynamics of direct reciprocity. <i>Proc. Royal Soc. B: Biol. Sci.</i> 277 , 463–468 (2010).	1074
1013	69. D Fudenberg, LA Imhof, Imitation processes with small mutations. <i>J. Econ. Theory</i> 131 , 251–262 (2006).	1075
1014	70. B Wu, CS Gokhale, L Wang, A Traulsen, How small are small mutation rates? <i>J. Math. Biol.</i> 64 , 803–827 (2012).	1076
1015	71. A McAvoy, Comment on "imitation processes with small mutations". <i>J. Econ. Theory</i> 159 , 66–69 (2015).	1077
1016	72. MA Nowak, A Sasaki, C Taylor, D Fudenberg, Emergence of cooperation and evolutionary stability in finite populations. <i>Nature</i> 428 , 646–650 (2004).	1078
1017	73. M Doebeli, C Hauert, Models of cooperation based on the prisoner's dilemma and the snowdrift game. <i>Ecol. Lett.</i> 8 , 748–766 (2005).	1079
1018	74. A Diekmann, Volunteer's dilemma. <i>J. Confl. Resolut.</i> 29 , 605–610 (1985).	1080
1019		1081
1020		1082
1021		1083
1022		1084
1023		1085
1024		1086
1025		1087
1026		1088
1027		1089
1028		1090
1029		1091
1030		1092
1031		1093
1032		1094
1033		1095
1034		1096
1035		1097
1036		1098
1037		1099
1038		1100
1039		1101
1040		1102
1041		1103
1042		1104
1043		1105
1044		1106
1045		1107
1046		1108
1047		1109
1048		1110
1049		1111
1050		1112
1051		1113
1052		1114
1053		1115
1054		1116