Conditional coperation with longer memory

Nikoleta E. Glynatsi[1,*], Martin Nowak[2], Christian Hilbe[1]

[1]Max Planck Research Group on the Dynamics of Social Behavior,
Max Planck Institute for Evolutionary Biology, Plön, Germany
[2]Department of Mathematics, Department of Organismic and Evolutionary Biology,
Harvard University, Cambridge, USA
[*]To whom correspondence should be addressed. E-mail: glynatsi@evolbio.mpg.de

Repeated games enable evolution of cooperation if players use conditional strategies that depend on previous interactions. A well known strategy set 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 n last moves of the co-player. A reactive-n counting strategy takes into account how often the co-player has cooperated during the last n round. We characterize all partner strategies among reactive-2 and reactive-3 strategies as well as among reactive-n counting strategies. Partner strategies are those that ensure mutual cooperation without exploitation. We perform evolutionary simulations and find that longer memory increases the average cooperation rate for reactive-n strategies but not for reactive counting strategies.

Keywords: Evolutionary game theory, direct reciprocity, evolution of cooperation, prisoner's dilemma

Significance statement. People tend to cooperate conditionally. We are often influenced by how cooperative others are, and we adapt our behavior accordingly. To describe conditional cooperation, theoretical models often presume that individuals only react to their last interaction. Instead, here we allow individuals to react to an opponent's n previous actions, for arbitrary n. We derive an algorithm to identify all partner strategies – strategies that sustain full cooperation in a Nash equilibrium. We give explicit conditions for $n=2$ and $n=3$. When individuals only count how often their opponent cooperated, independent of the timing of cooperation, we characterize partner strategies for all n. These results reveal which strategies sustain cooperation under more realistic assumptions on people's cognitive abilities.

Introduction

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 [7–11]. 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 that have been explored in numerous studies [12–30]. 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 [31–34]. Others describe the effect of different learning dynamics [35, 36], of population structure [37–40], or of the strategies that players are permitted to use [41].

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 [42, 43]. To quantify a strategy's complexity, it is common 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, 44] or 'Win-Stay Lose-Shift' [19, 20], 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 [20–30]. Although one-round memory can explain some of the empirical regularities in human behavior [45–49], people often take into account more than the last round [50]. Longer memory seems particularly relevant for noisy games, where people occasionally defect because of unintended errors [51]. However, a formal analysis of strategies with more than one-round memory has been difficult 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 [52]. 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

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. Probably for these reasons, previous studies considered simulations for small n [52–54], or they analyzed the properties of a few selected higher-memory strategies [55–57].

To make 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**). While it has been difficult to explicitly characterize all Nash equilibria among the memory-n strategies, we show that such a characterization is possible for reactive-n strategies. Our results rely on a central insight, motivated by previous work by Press & Dyson [24]: 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. To be stable, however, these strategies need to be sufficiently responsive to the co-player's previous actions.

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 defector receives the temptation payoff 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 $R=b-c, S=-c, T=b, P=0$. To illustrate our results, 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 **Supporting Information**).

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 $\mathbf{h}^i=(a^i_{-n},\ldots,a^i_{-1})\in\{C,D\}^n$. Each entry a^i_{-k} corresponds to player i's action k rounds ago. We use H^i for the set of all such 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 $n=1$, the above definition recovers 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, $\texttt{ALLD}=(0,0)$, and Tit-for-Tat, $\texttt{TFT}=(1,0)$. 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, $\texttt{TF2T}=(1,1,1,0)$ and Two-Tit-for-Tat, $\texttt{2TFT}=(1,0,0,0)$. 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 **Supporting Information**.

Herein, we are particularly interested in those reactive-n strategies that sustain full cooperation. Such strategies ought to have two properties. First, such a strategy ought to be *nice*, meaning that it should never be the first to defect [7]. This property ensures that two players with that strategy would 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 cooperating with certainty, $p_{\mathbf{h}_C}=1$. 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* [58] 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 $p_D\leq 1-c/b$ [28]. 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 the respective results; all proofs are in the **Supporting Information**.

An algorithm to identify partners among the 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 [24] 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 $\mathbf{p}=(p_{\mathbf{h}})_{\mathbf{h}\in H^1}$. 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, $\tilde{\mathbf{p}}=(\tilde{p}_{\mathbf{h}})_{\mathbf{h}\in H^1}$. In particular, if σ is a best response to $\mathbf{p}$, then 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 [24]: 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 reduces the search space for potential best responses further, 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 substantial number, many of the different strategies impose redundant constraints. This redundancy further reduces the number of conditions a partner strategy 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 **Supporting Information** for details). 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 conclude that $\mathbf{p}$ is a partner if and only if

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

The above conditions define a three-dimensional polyhedron within the space of all nice reactive-2 strategies (**Fig. 2C**). These conditions are straightforward to interpret. 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 by $c/(2b)$. 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*. Interestingly, 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). 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 **Supporting Information**).

Analogously, we can also characterize the partners among the reactive-3 strategies. A reactive-3 strategy is defined by the 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. However, for $n=3$, there are now more conditions to consider. 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 the case of arbitrary n if we further restrict the strategy space. In the following, we consider reactive-n *counting strategies*. These strategies take into account how often the co-player cooperated during the past n rounds. However, they do not consider in which of the past n rounds the co-player cooperated. In the following, we represent such strategies as a vector $\mathbf{r} = (r_i)_{i \in \{n, n-1, \ldots, 0\}}$. Each entry r_i indicates the player's cooperation probability if the co-player cooperated i times during the last n rounds. Note that 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 among players with counting strategies, 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

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

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.

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 can 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, however, 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 [59–62]. 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 [63].

The results of these simulations are shown in **Fig. 3**. First, we explore which reactive-n strategies evolve for a fixed set of game parameters. Here, we only vary the strategies' memory length n, and whether mutations can introduce all reactive-n strategies, or counting strategies only. For ten independent simulations, **Fig. 3A,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).

In a next step, we systematically explore the impact of three key parameters: the cost-to-benefit ratio c/b, the selection strength β, and the memory length n. In each case, we record how these parameters affect the abundance of partner strategies and the population's average cooperation rate. Overall, we find that the effect of each parameter is as expected (**Fig. 3C,D**). In particular, interactions are most cooperative when the cost-to-benefit ratio is small, such that cooperation is 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, higher memory only has a notably positive effect on cooperation for reactive-n strategies. In contrast, for counting strategies the effect of increasing n is negligible. This observation suggests that higher-memory strategies are only effective when players do not only memorize how often their co-player cooperated, but also when.

Discussion

Previous theoretical research has mainly focused on a single set of strategies in repeated games, namely, memory-1 strategies. Although several results have been proven for this class, generalizing to larger memory classes has proven to be a challenging task. We venture into the realm of higher memory strategies by concentrating on reactive strategies. Reactive strategies are a set that observes only the previous turns of the co-player. They have been studied in the past in theoretical work, with famous strategies such as Tit for Tat and Generous Tit for Tat [20]. Experimental research has even suggested that these strategies are adopted by humans [45, 48]. However, prior work on reactive strategies has also been limited to the case of memory one.

We focus on a set of Nash equilibria, which are the partner strategies. Partner strategies not only ensure that their co-player has no reason to deviate but also that as long as the co-player wants to, the payoff of mutual cooperation can be achieved. Partner strategies are a set of strategies that allow for evolution of cooperation [11], which is also verified by our own work.

We begin by proving the result that if a player employs a reactive strategy, then the co-player using a memory-n strategy can switch to a self-reactive-n strategy without altering the resulting payoffs. This result makes it easier for us to characterize Nash strategies within the reactive set. We characterize partner strategies for reactive-2 and reactive-3, both in the special case of the donation game and the general Prisoner's Dilemma. Moreover, we also demonstrate that reactive strategies such as Tit For Tat, Generous Tit For Tat, and any delayed version of them are partner strategies (see Supplementary Information).

We also focus on the set of counting strategies. In this case, we can easily derive the condition for being a partner for $n = 2$ and $n = 3$. Furthermore, counting strategies allow us to characterize all partner strategies regardless of the memory size. The conditions for being partner in the counting set are simple yet novel. The intuition of these conditions is that the generosity shown by a partner strategy after a sequence of k defections in the last n rounds must be less than $1 - k/n$ of the cost-benefit ratio. This condition ensures that as the total number of defections increases, the strategy's generosity decreases.

When testing the evolutionary properties of counting strategies, it is evident from the simulation results that cooperation cannot emerge beyond the simple case of reactive-1 strategies. Thus, we observe that within the reactive set, the evolution of cooperation relies on the sequential memory of these strategies. Overall, our study is among the first to characterize full spaces of partner strategies in higher memory spaces. Although reactive strategies are a subset of memory strategies, we have demonstrated that there are many results to explore in this case.

Materials and Methods

In the following paragraphs, we describe the framework of our evolutionary process. The framework considers a population of size N where initially all members are of the same strategy. In our case the initial population consists of unconditional defectors. In each elementary time step, one individual switches to a new mutant strategy. The mutant strategy is generated by randomly drawing cooperation probabilities from the unit interval $[0, 1]^n$. 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,

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

The parameter $\beta \geq 0$ is called the strength of selection, and it measures the importance of the relative payoff advantages for the evolutionary success of a strategy. For small values of β, $\beta \approx 0$, 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 favours the fixation of strategies that yield high payoffs. Depending on the fixation probability ϕ_M the mutant either fixes (becomes the new resident) or goes extinct. Regardless, in the elementary time step another mutant strategy is introduced to the population. We iterate this elementary population updating process for a large number of mutant strategies and we record the resident strategies at each time step.

CH: For the code, could we provide a link to some online repository? Also, it would be nice to have more information on how we classified strategies as partners in the simulations.

References

[1] Melis, A. P. & Semmann, D. How is human cooperation different? *Philosophical Transactions of the Royal Society B* **365**, 2663–2674 (2010).

[2] Rand, D. G. & Nowak, M. A. Human cooperation. *Trends in Cogn. Sciences* **117**, 413–425 (2012).

[3] Neilson, W. S. The economics of favors. *Journal of Economic Behavior & Organization* **39**, 387–397 (1999).

[4] Fischbacher, U. & Gächter, S. Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *American economic review* **100**, 541–556 (2010).

[5] Hilbe, C., Röhl, T. & Milinski, M. Extortion subdues human players but is finally punished in the prisoner's dilemma. *Nature Communications* **5**, 3976 (2014).

[6] Xu, B., Zhou, Y., Lien, J. W., Zheng, J. & Wang, Z. Extortion can outperform generosity in iterated prisoner's dilemma. *Nature Communications* **7**, 11125 (2016).

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

[8] Nowak, M. A. Five rules for the evolution of cooperation. *science* **314**, 1560–1563 (2006).

[9] Sigmund, K. *The calculus of selfishness* (Princeton University Press, 2010).

[10] García, J. & van Veelen, M. No strategy can win in the repeated prisoner's dilemma: Linking game theory and computer simulations. *Frontiers in Robotics and AI* **5**, 102 (2018).

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

[12] Frean, M. R. The prisoner's dilemma without synchrony. *Proceedings of the Royal Society B* **257**, 75–79 (1994).

[13] Killingback, T., Doebeli, M. & Knowlton, N. Variable investment, the continuous prisoner's dilemma, and the origin of cooperation. *Proceedings of the Royal Society B* **266**, 1723–1728 (1999).

[14] Hauert, C. & Stenull, O. Simple adaptive strategy wins the prisoner's dilemma. *Journal of Theoretical Biology* **218**, 261–72 (2002).

[15] Kurokawa, S. & Ihara, Y. Emergence of cooperation in public goods games. *Proceedings of the Royal Society B* **276**, 1379–1384 (2009).

[16] Pinheiro, F. L., Vasconcelos, V. V., Santos, F. C. & Pacheco, J. M. Evolution of all-or-none strategies in repeated public goods dilemmas. *PLoS Comput Biol* **10**, e1003945 (2014).

[17] García, J. & van Veelen, M. In and out of equilibrium I: Evolution of strategies in repeated games with discounting. *Journal of Economic Theory* **161**, 161–189 (2016).

[18] McAvoy, A. & Nowak, M. A. Reactive learning strategies for iterated games. *Proceedings of the Royal Society A* **475**, 20180819 (2019).

[19] Kraines, D. P. & Kraines, V. Y. Pavlov and the prisoner's dilemma. *Theory and Decision* **26**, 47–79 (1989).

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

[21] Imhof, L. A., Fudenberg, D. & Nowak, M. A. Evolutionary cycles of cooperation and defection. *Proceedings of the National Academy of Sciences USA* **102**, 10797–10800 (2005).

[22] Grujic, J., Cuesta, J. A. & Sanchez, A. On the coexistence of cooperators, defectors and conditional cooperators in the multiplayer iterated prisoner's dilemma. *Journal of Theoretical Biology* **300**, 299–308 (2012).

[23] van Segbroeck, S., Pacheco, J. M., Lenaerts, T. & Santos, F. C. Emergence of fairness in repeated group interactions. *Physical Review Letters* **108**, 158104 (2012).

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

[25] Stewart, A. J. & Plotkin, J. B. From extortion to generosity, evolution in the iterated prisoner's dilemma. *Proceedings of the National Academy of Sciences USA* **110**, 15348–15353 (2013).

[26] Toupo, D. F. P., Rand, D. G. & Strogatz, S. H. Limit cycles sparked by mutation in the repeated prisoner's dilemma. *International Journal of Bifurcation and Chaos* **24**, 2430035 (2014).

[27] Stewart, A. J. & Plotkin, J. B. Collapse of cooperation in evolving games. *Proceedings of the National Academy of Sciences USA* **111**, 17558 – 17563 (2014).

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

[29] Glynatsi, N. E. & Knight, V. A. Using a theory of mind to find best responses to memory-one strategies. *Scientific reports* **10**, 1–9 (2020).

[30] Chen, X. & Fu, F. Outlearning extortioners: unbending strategies can foster reciprocal fairness and cooperation. *PNAS nexus* **2**, pgad176 (2023).

[31] Boyd, R. Mistakes allow evolutionary stability in the repeated Prisoner's Dilemma game. *Journal of Theoretical Biology* **136**, 47–56 (1989).

[32] Hao, D., Rong, Z. & Zhou, T. Extortion under uncertainty: Zero-determinant strategies in noisy games.

Physical Review E **91**, 052803 (2015).

[33] Zhang, H. Errors can increase cooperation in finite populations. *Games and Economic Behavior* **107**, 203–219 (2018).

[34] Mamiya, A. & Ichinose, G. Zero-determinant strategies under observation errors in repeated games. *Physical Review E* **102**, 032115 (2020).

[35] Stewart, A. J. & Plotkin, J. B. The evolvability of cooperation under local and non-local mutations. *Games* **6**, 231–250 (2015).

[36] McAvoy, A., Kates-Harbeck, J., Chatterjee, K. & Hilbe, C. Evolutionary instability of selfish learning in repeated games. *PNAS nexus* **1**, pgac141 (2022).

[37] Brauchli, K., Killingback, T. & Doebeli, M. Evolution of cooperation in spatially structured populations. *Journal of Theoretical Biology* **200**, 405–417 (1999).

[38] Szabó, G., Antal, T., Szabó, P. & Droz, M. Spatial evolutionary prisoner's dilemma game with three strategies and external constraints. *Physical Review E* **62**, 1095–1103 (2000).

[39] Allen, B., Nowak, M. A. & Dieckmann, U. Adaptive dynamics with interaction structure. *American Naturalist* **181**, E139–E163 (2013).

[40] Szolnoki, A. & Perc, M. Defection and extortion as unexpected catalysts of unconditional cooperation in structured populations. *Scientific Reports* **4**, 5496 (2014).

[41] Baek, S. K., Jeong, H.-C., Hilbe, C. & Nowak, M. A. Comparing reactive and memory-one strategies of direct reciprocity. *Scientific Reports* **6**, 1–13 (2016).

[42] Harper, M. *et al.* Reinforcement learning produces dominant strategies for the iterated prisoner's dilemma. *PloS one* **12**, e0188046 (2017).

[43] Knight, V., Harper, M., Glynatsi, N. E. & Campbell, O. 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).

[44] Duersch, P., Oechssler, J. & Schipper, B. When is tit-for-tat unbeatable? *International Journal of Game Theory* **43**, 25–36 (2013).

[45] Engle-Warnick, J. & Slonim, R. L. Inferring repeated-game strategies from actions: evidence from trust game experiments. *Economic theory* **28**, 603–632 (2006).

[46] Dal Bó, P. & Fréchette, G. R. The evolution of cooperation in infinitely repeated games: Experimental evidence. *American Economic Review* **101**, 411–429 (2011).

[47] Camera, G., Casari, M. & Bigoni, M. Cooperative strategies in anonymous economies: An experiment. *Games and Economic Behavior* **75**, 570–586 (2012).

[48] Bruttel, L. & Kamecke, U. Infinity in the lab. How do people play repeated games? *Theory and Decision* **72**, 205–219 (2012).

[49] Montero-Porras, E., Grujić, J., Fernández Domingos, E. & Lenaerts, T. Inferring strategies from observations in long iterated prisoner's dilemma experiments. *Scientific Reports* **12**, 7589 (2022).

[50] Romero, J. & Rosokha, Y. Constructing strategies in the indefinitely repeated prisoner's dilemma game. *European Economic Review* **104**, 185–219 (2018).

[51] Fudenberg, D., Rand, D. G. & Dreber, A. Slow to anger and fast to forgive: Cooperation in an uncertain world. *American Economic Review* **102**, 720–749 (2012).

[52] Hauert, C. & Schuster, H. G. Effects of increasing the number of players and memory size in the iterated prisoner's dilemma: a numerical approach. *Proceedings of the Royal Society B* **264**, 513–519 (1997).

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

[54] Murase, Y. & Baek, S. K. Grouping promotes both partnership and rivalry with long memory in direct reciprocity. *PLoS Computational Biology* **19**, e1011228 (2023).

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

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

[57] Li, J. *et al.* Evolution of cooperation through cumulative reciprocity. *Nature Computational Science* **2**, 677–686 (2022).

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

[59] Fudenberg, D. & Imhof, L. A. Imitation processes with small mutations. *Journal of Economic Theory* **131**, 251–262 (2006).

[60] Wu, B., Gokhale, C. S., Wang, L. & Traulsen, A. How small are small mutation rates? *Journal of Mathematical Biology* **64**, 803–827 (2012).

[61] Imhof, L. A. & Nowak, M. A. Stochastic evolutionary dynamics of direct reciprocity. *Proceedings of the Royal Society B: Biological Sciences* **277**, 463–468 (2010).

[62] McAvoy, A. Comment on "Imitation processes with small mutations". *J. Econ. Theory* **159**, 66–69 (2015).

[63] Nowak, M. A., Sasaki, A., Taylor, C. & Fudenberg, D. Emergence of cooperation and evolutionary stability in finite populations. *Nature* **428**, 646–650 (2004).

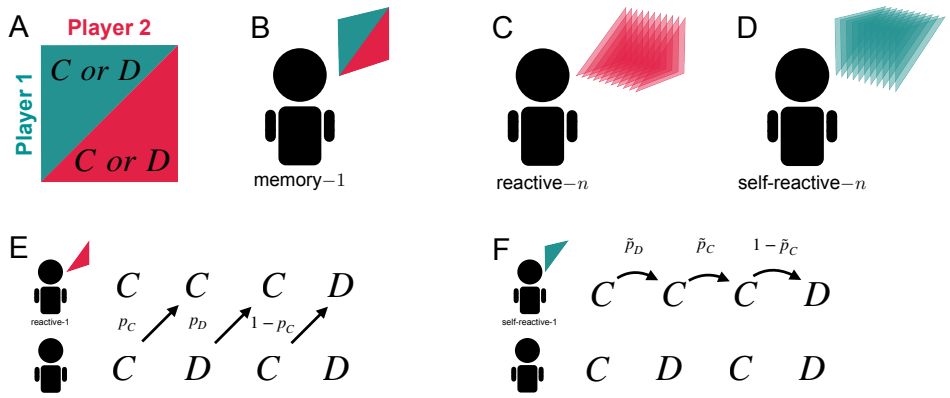

Figure 1: The repeated prisoner's dilemma among players with finite memory. **A,** In the repeated prisoner's dilemma, in each round two players independently decide whether to cooperate (C) or to defect (D). **B,** When players adopt memory-1 strategies, their decisions depend on the entire outcome 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. **E,** Now, the top player adopts a self-reactive-1 strategy, $\tilde{\mathbf{p}} = (\tilde{p}_C, \tilde{p}_D)$. Here, the bottom player's cooperation probabilities depend on their own previous action. **CH**: In panel F, I believe the first arrow should come with a $\tilde{p}_C$ instead of a $\tilde{p}_D$.

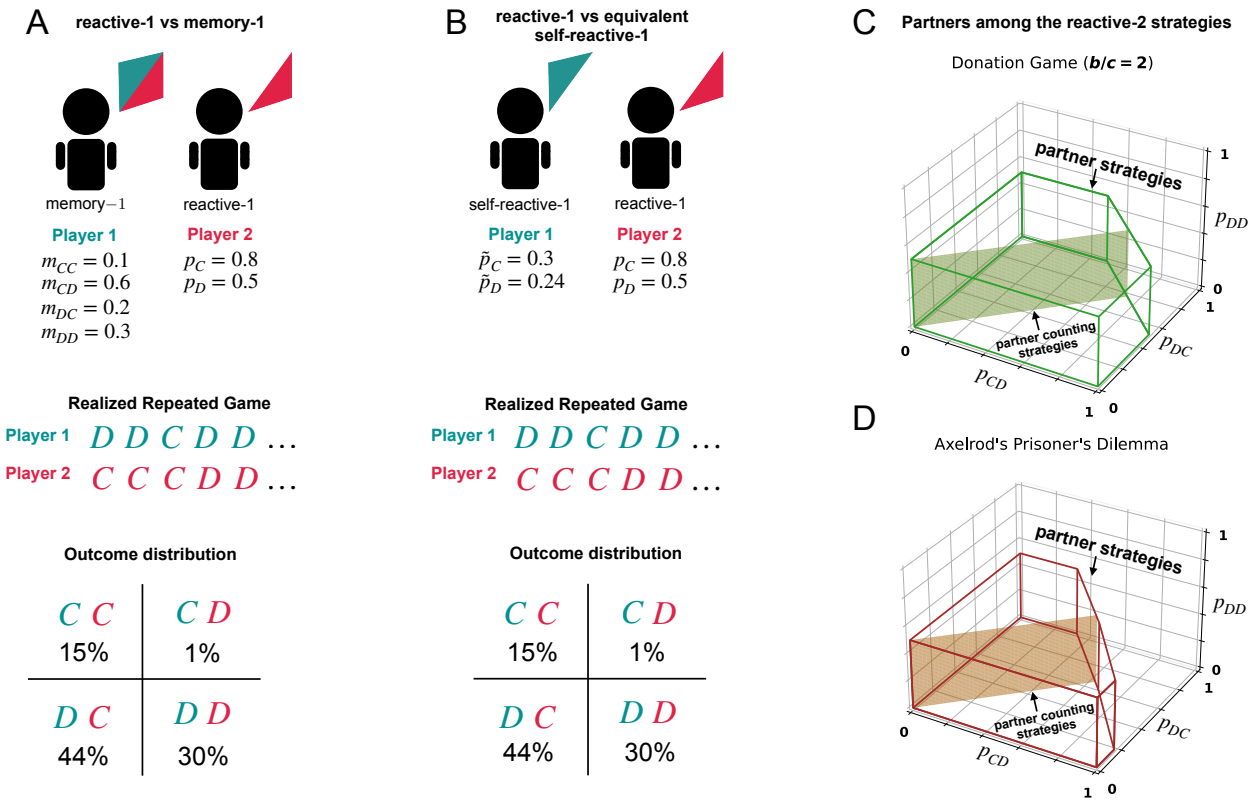

Figure 2: Characterizing the partners among the reactive-n strategies. A,B, To characterize the reactive-n partner strategies, we prove the following result. Suppose the focal player adopts a reactive-n strategy. Then, for any strategy of the opponent (with arbitrary memory), one can find an associated self-reactive-n strategy that yields the same payoffs. Here, we show an example where player 1 uses a reactive-1 strategy against player 2 with a memory-1 strategy. Our result implies that 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 (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 values of Axelrod [7]. **CH**: I believe there is an error in the displayed outcome distribution in panels A,B – the numbers don't add up to 100%. From what I can tell, the correct numbers are 15,10,43,32. Also, could we show the outcome distribution with a slightly higher accuracy – 15.3%, 10.6%, 42.5%, 31.7% – and also show the respective self-reactive-1 strategy with a higher accuracty?

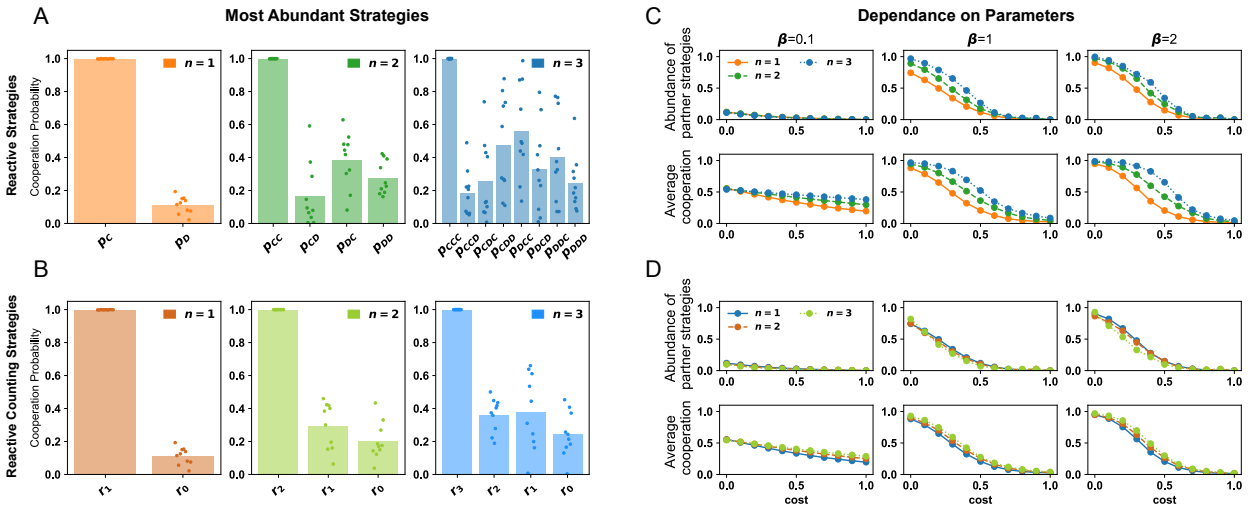

Figure 3: Evolutionary dynamics of reactive-n strategies. To explore the evolutionary dynamics among reactive-n strategies, we run simulations based on the method of Imhof and Nowak [61]. This method assumes rare mutations. Every time a mutant strategy appears, it goes extinct or fixes before the arrival of the next mutant strategy. **A,B,** We run ten 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 definition of partner strategies. 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 $b = 1$, $c = 0.5$, and a selection strength $\beta = 1$, unless noted otherwise. For n equal to 1 and 2, simulations are run for $T = 10^7$ time steps. For $n = 3$ we use $T = 2 \cdot 10^7$ time steps. **CH**: For the header on the right hand side, could we use 'dependence' instead of 'dependance'?