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. Humans tend to cooperate conditionally. We are often influenced by how cooperative others are, and we adapt our behavior accordingly. To describe reciprocal 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 clarify which strategies sustain cooperation under 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 [12–28]. 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 [29–32]. Others describe the effect of different learning dynamics [33, 34], of population structure [35–38], or of the strategies that players are permitted to use [39].

Strategies of the repeated prisoner's dilemma can vary in their complexity. While some are straightforward to implement, like always defect, others are more sophisticated [40, 41]. 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, 42] or Win-Stay Lose-Shift [17, 18], are memory-1 (**Fig. 1B**). Similarly, one can distinguish strategies that require more than one round of memory, or rules that cannot be represented as a finite-memory strategy [10].

Traditionally, most theoretical research on the evolution of reciprocity focuses on memory-1 strategies [18–28]. Although one-round memory can explain some of the empirical regularities in human behavior [43–47], people often take into account more than the last round [49]. Longer memory seems particularly relevant for noisy games, where people might defect because of unintended errors [48]. 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 [50]. 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 payoffs, increases dramatically in n. For example, for $n=1$, there are $2^4=16$ deterministic strategies. Computing the payoffs of players with one-round memory requires the inversion of a 4×4 matrix [9]. Already for $n=2$, there are $2^{16}=64,536$ deterministic strategies, and payoffs follow from a 16×16 matrix. Probably for these reasons, previous studies were either restricted to simulations for small n [50–52], or they analyzed the properties of a few particular higher-memory strategies [53–55].

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, such strategies 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 [22]: 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 identify all partner strategies. These are the strategies that sustain full cooperation in a Nash equilibrium. We fully characterize the partners among the reactive-n strategies 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 can 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

Definitions. We consider an infinitely repeated game with two players, player 1 and player 2. In each round players can choose to cooperate (C) or to defect (D). If both players cooperate, they receive a payoff R (the reward for mutual cooperation), and if both players defect, they receive a payoff P (the punishment for mutual defection). If one player cooperates, the cooperative player receives the sucker's payoff S, and the defecting player receives the temptation payoff T. We assume that the payoff are such that $T>R>P>S$ and $2R>T+S$. This game is known as the Prisoner's Dilemma. Here, we employ a specific parametrization of the Prisoner's Dilemma, where cooperation implies incurring a cost c for the co-player to derive a benefit $b>c$. Consequently, the payoffs are defined as follows: $R=b-c, S=-c, T=b, P=0$. In the Supplementary Information, we show that our main results are applicable to the general Prisoner's Dilemma.

We assume in the following, that the players' decisions only depend on the outcome of the previous n rounds. To this end, an *n-history for player i* $\in\{1,2\}$ is a string $h^i=(a^i_{-n},\ldots,a^i_{-1})\in\{C,D\}^n$ where

an entry a^i_{-k} corresponds to player i's action k rounds ago. Let H^i denote the space of all n-histories for player i where set H^i contains $|H^i| = 2^n$ elements. A *reactive$-n$ strategy* for player 1 is a vector $\mathbf{p} = (p_h)_{h\in H^2} \in [0,1]^n$. Each entry p_h corresponds to the player's cooperation probability in the next round, based on the co-player's actions in the previous n rounds. Therefore, reactive-n strategies exclusively rely on the co-player's n-history, independent of the focal player's own actions. For $n = 1$, this definition of reactive-n strategies recovers the typical format of reactive-1 strategies [16, 39, 56], $\mathbf{p} = (p_C, p_D)$. Another class of strategies we will be discussing in this work are, *self-reactive-n* strategies which only consider the focal player's own n-history, and ignore the co-player's. Formally, a self-reactive-n strategy for player 1 is a vector $\tilde{\mathbf{p}} = (\tilde{p}_h)_{h\in H^1} \in [0,1]^n$. Each entry $\tilde{p}_h$ corresponds to the player's cooperation probability in the next, depending on the player's own actions in the previous n rounds. In Fig. 1, we summarize and provide a graphical representation of reactive and self-reactive strategies, as well as examples of these classes for $n = 1$. Lastly, note that we refer to a reactive or self-reactive strategy as pure if all the entries of the strategy are either 0 or 1.

In a repeated game, a strategy is considered a *Nash strategy* if and only if the payoff when playing against itself is greater than or equal to any payoff that any other strategy can achieve against it. In this work, we focus on a set of Nash strategies called partner strategies. To define partner strategies, we first need to introduce the notion of a nice strategy. A strategy is considered *nice* if the player is never the first to defect. A nice strategy, when played against itself, receives the mutual cooperation payoff. Now, a *partner strategy* is a nice strategy that also satisfies the Nash equilibrium condition.

Self-Reactive Sufficiency. To predict which reactive-n strategies are partner strategies, we must characterize which nice reactive-n strategies are Nash equilibria. Determining whether a given strategy, $\mathbf{p}$, is a Nash equilibrium is not straightforward. In principle, this would involve comparing the payoff of $\mathbf{p}$ to the payoff of all possible other strategies; however, due to the result of [22], we know that we only have to compare against memory-n strategies.

There can still be infinitely many memory-n strategies one would have to check against. However, we restrict the search space even further. Namely, we have shown that if a player adopts a reactive strategy, it is only necessary to consider mutant strategies that are self-reactive-n (Fig. 2A-B). Our result aligns with the findings of [22]. They explored a scenario where one player uses a memory-1 strategy while the other employs a longer memory strategy. They demonstrated that the payoff of the player with the longer memory is exactly the same as if they had used a specific shorter-memory strategy, disregarding any history beyond what is shared with the short-memory player. Our result that follows a similar intuition: there is a part of history that a reactive player does not observe, the co-player gains nothing by considering the history not shared with the reactive player.

Furthermore, we have shown that we only need to consider pure self-reactive-n strategies (see Supplementary Information for proof). Thus, in the case of $n = 2$, we can check whether a given strategy $\mathbf{p}$ is Nash by comparing its payoff to $2^4 = 16$ possible self-reactive strategies, and in the case of $n = 3$, we can check against $2^8 = 256$ possible self-reactive strategies.

Partner Strategies Amongst Reactive-2 and Reactive-3 Strategies. Using the self-reactive sufficiency result we can characterize partner strategies amongst the reactive-2 and reactive-3 strategies. A reactive-2 strategy can be defined as the vector $\mathbf{p} = (p_{CC}, p_{CD}, p_{DC}, p_{DD})$, and it is a partner strategy if and only if,

the strategy entries satisfy the conditions,

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

Hence, for a strategy to be a Nash equilibrium, it must ensure that the strategy ALLD doesn't yield a higher payoff (achieved by $p_{DD} \leq 1 - \frac{c}{b}$), and the average cooperation rate after a single defection by the co-player in the last two rounds must be less than half the cost-benefit ratio (c/b). These conditions define partner strategies as a three-dimensional polyhedron within the space of all nice reactive-2 strategies (Fig. 2C).

A reactive-3 strategy is defined by the vector $\mathbf{p} = (p_{CCC}, p_{CCD}, p_{CDC}, p_{CDD}, p_{DCC}, p_{DCD}, p_{DDC}, p_{DDD})$, and it is a partner strategy, if and only if the strategy entries satisfy the conditions,

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

Inherently, these conditions still exhibit some symmetry with the case of reactive-2. Namely, for the strategy to be Nash, ALLD should not achieve a higher payoff. Additionally, the average cooperation following a single defection must be lower than 2/3 of the cost-benefit ratio, and the average cooperation following two defections must be smaller than 1/3 of the cost-benefit ratio. However, there are two further conditions that appear not to align with this intuition. We hypothesize that as the memory space we allow increases, the number of conditions will also increase, and some of these conditions will deviate from the symmetry. Note that the two additional conditions ensure that strategies playing the sequence of actions $CCDD$ and CD cannot exploit the strategy.

The proofs for the above results can be found in the Supplementary Information. In addition to demonstrating the results using the methodology we have described in the paper, we can also verify them using an independent proof. This independent proof builds upon the framework developed by Akin [26].

In the Supplementary Information we derive the conditions for partner strategies for the general Prisoner's Dilemma for reactive-2 and reactive-3 strategies. In Fig. 2D, we plot the space of partner strategies for $n = 2$ and for $R = 3, S = 0, T = 5, P = 1$.

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

indicates the probability of cooperating given that the co-player cooperated i times in the last n turns. Note that a reactive-1 strategy $\mathbf{p} = (p_C, p_D)$ and a counting strategy $\mathbf{r} = (r_1, r_0)$ are equivalent because both strategies describe the probability of cooperating after a single or no cooperation by the co-player through their respective entries.

A reactive-2 counting strategy is denoted by the vector $\mathbf{r} = (r_2, r_1, r_0)$, and we can characterise partner strategies among the reactive-2 counting strategies by simply setting $r_2 = 1$, and $p_{CD} = p_{DC} = r_1$ and $p_{DD} = r_0$ in conditions (1) which gives us the following conditions,

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

Similarly, a reactive-3 counting strategy is denoted by the vector $\mathbf{r} = (r_3, r_2, r_1, r_0)$, and we characterise partner strategies among reactive-3 counting strategies by setting $r_3 = 1$, and $p_{CCD} = p_{CDC} = p_{DCC} = r_2, p_{DCD} = p_{DDC} = p_{CDD} = r_1$ and $p_{DDD} = r_0$ in conditions (2). This gives us the following conditions,

$$r_3 = 1, \quad r_2 < 1 - \frac{1}{3} \cdot \frac{c}{b}, \quad r_1 < 1 - \frac{2}{3} \cdot \frac{c}{b} \ \textit{and} \ r_0 < 1 - \frac{c}{b}. \tag{4}$$

Counting strategies are a subset of reactive strategies, and as such, they exist within the space of reactive partner strategies. For example, in the case of $n = 2$, the counting partner strategies form a plane within the three-dimensional polyhedron of reactive-2 partners (Fig. 2B). Counting partner strategies appear to align with the intuition that the generosity (the probability of cooperating after a defection, thus being generous with your co-player) exhibited by a strategy after a k number of defections in the last n rounds must be less than $1 - k/n$ of the cost-benefit ratio. As the total number of defections increases, the strategy's generosity decreases. And, precisely, this is the result we prove (see Supplementary Information). In the case of reactive-counting strategies, we characterize partner strategies for all memory lengths. A reactive-counting strategy is a partner if and only if,

$$r_n = 1 \ \textit{and} \ r_{n-k} < 1 - \frac{k}{n} \cdot \frac{c}{b}, \text{ for } k \in \{1, 2, \ldots, n\}. \tag{5}$$

Evolutionary Dynamics. Based on our previous equilibrium analysis, we know the conditions that a reactive strategy must satisfy to be considered a partner strategy. The next step is to determine whether these strategies are likely to evolve through an evolutionary process. Additionally, what remains unclear is the impact of increased memory, as well as the consequences of limiting strategies to counting alone. Here, we will empirically explore these questions by simulating an imitation process, using the framework described by Imhof and Nowak [57]. The setup of the framework is outlined in Materials and Methods .

First, we explore which strategies evolve from the evolutionary dynamics for a fixed set of parameters. We ran 10 independent simulations for each set of strategies and recorded the resident strategy at each elementary time step. Once a strategy has become a resident we also record the number of time steps it remained a resident. Thus, the number of mutants that have unsuccessfully tried to invade the resident population. In Fig. 3A and B we represent those strategies that repelled the highest number of mutant in each run. We call these strategies the *most abundant*. Fig. 3A shows the most abundant strategies for reactive strategies and Fig. 3B shows the most abundant strategies for counting strategies. In both cases

the most abundant strategies resemble partner strategies. In the case of counting strategies, we can see the decreasing levels of forgiveness as the number of cooperations decreases.

Next, we compare the evolution of partner strategies and the changing cooperation rates for different memory sizes while varying the selection strength. To this end, we ran simulations for different b/c ratios. As we examine the impact of memory size on the evolution of partner strategies, several patterns emerge. Increasing memory size tends to result in a higher abundance of partner strategies, regardless of the selection strength. Notably, the highest abundance is observed for lower cost values. We notice that the curves representing evolving cooperation rates align with the prevalence of partner strategies. Thus, it is the presence of partner strategies that facilitates the evolution of cooperation, and as memory selects partner strategies more frequently, the cooperation rate also increases with memory. In contrast, when examining counting strategies, we notice that the abundance of partner strategies rise with the strength of selection. However, there is no corresponding increase as memory size expands. Thus, in the case of counting strategies there is no added value in increasing memory size, from an evolutionary perspective.

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 [18]. Experimental research has even suggested that these strategies are adopted by humans [43, 46]. 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,

$$\phi_M = \frac{1}{\left(1 + \sum_{i=1}^{N-1} \prod_{j=1}^{i} e^{(-\beta(\pi_{M,j} - \pi_{R,i}))}\right)} \tag{6}$$

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: Here, could we provide a link to some online repository?

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] 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).

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

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

[18] 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).

[19] 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).

[20] 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).

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

[22] 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).

[23] 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).

[24] 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).

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

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

[27] 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).

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

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

[30] Hao, D., Rong, Z. & Zhou, T. Extortion under uncertainty: Zero-determinant strategies in noisy games. *Physical Review E* **91**, 052803 (2015).

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

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

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

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

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

[36] 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).

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

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

[39] 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).

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

[41] 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).

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

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

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

[45] Camera, G., Casari, M. & Bigoni, M. Cooperative strategies in anonymous economies: An experiment.

Games and Economic Behavior **75**, 570–586 (2012).

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

[47] 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).

[48] 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).

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

[50] 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).

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

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

[53] 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).

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

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

[56] Wahl, L. M. & Nowak, M. A. The continuous prisoner's dilemma: I. linear reactive strategies. *Journal of Theoretical Biology* **200**, 307–321 (1999).

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

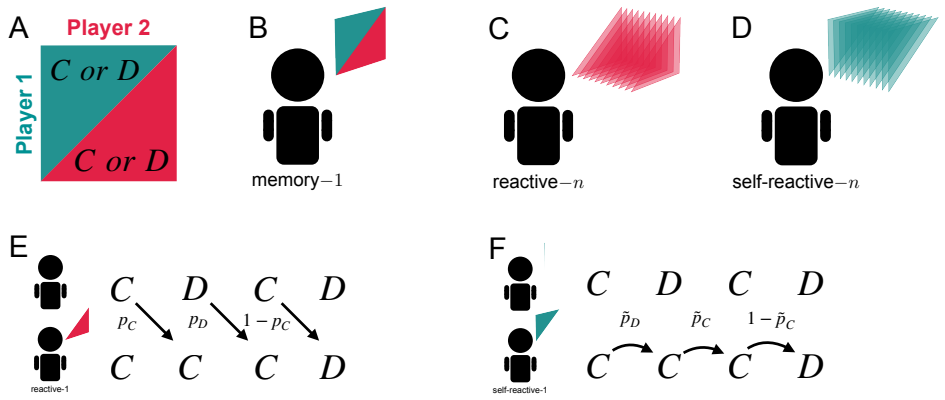

Figure 1: The repeated prisoner's dilemma among players with finite memory. **A,** In the repeated prisoner's dilemma, in each round two players independently decide whether to cooperate (C) or to defect (D). **B,** When players adopt memory-1 strategies, their decisions depend on the outcome of the 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 an arbitrary player (top) and a player with a reactive-1 strategy (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 after the co-player cooperated in the previous round. The entry p_D is the cooperation probability after the co-player defected. **E,** Now, the bottom 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.

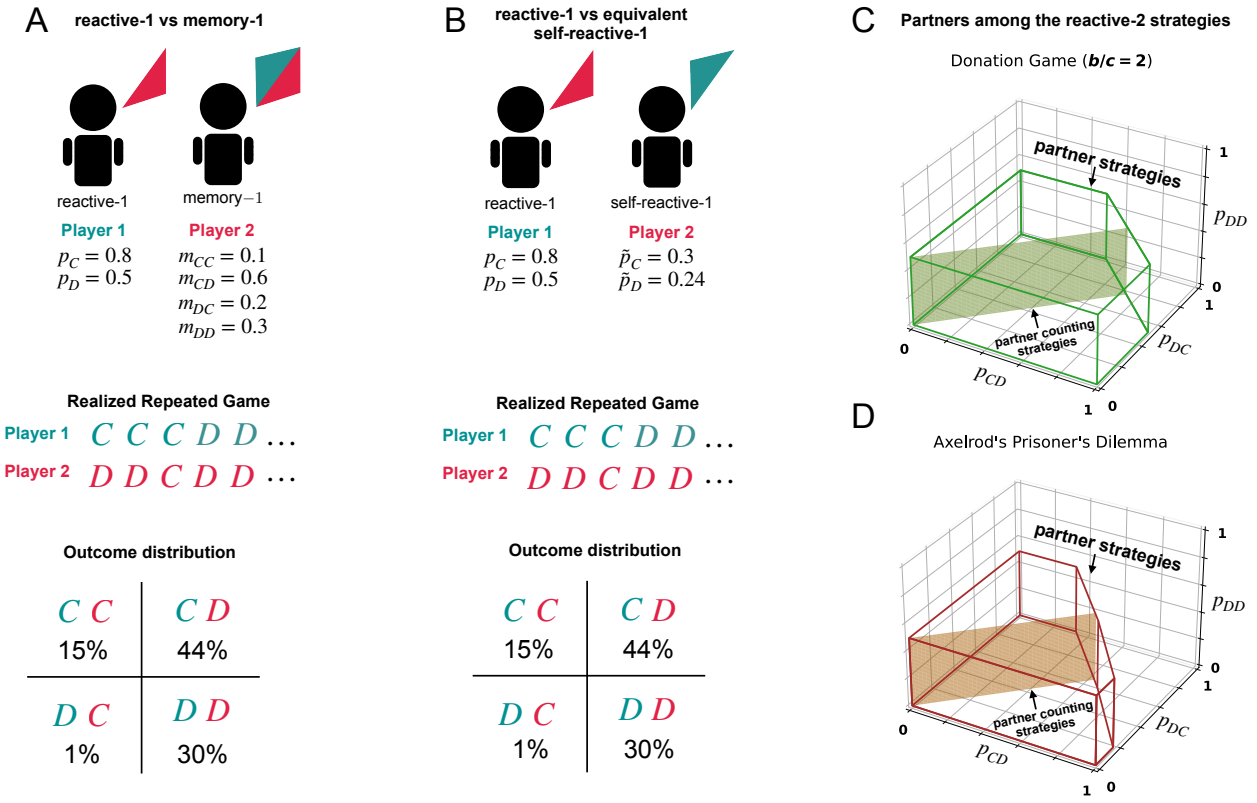

Figure 2: Characterizing the partners among the reactive-n strategies. A,B, To characterize the reactive-n partner strategies, we prove the following result. Suppose the focal player adopts a reactive-n strategy. Then, for any strategy of the opponent (with arbitrary memory), one can find an associated 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].

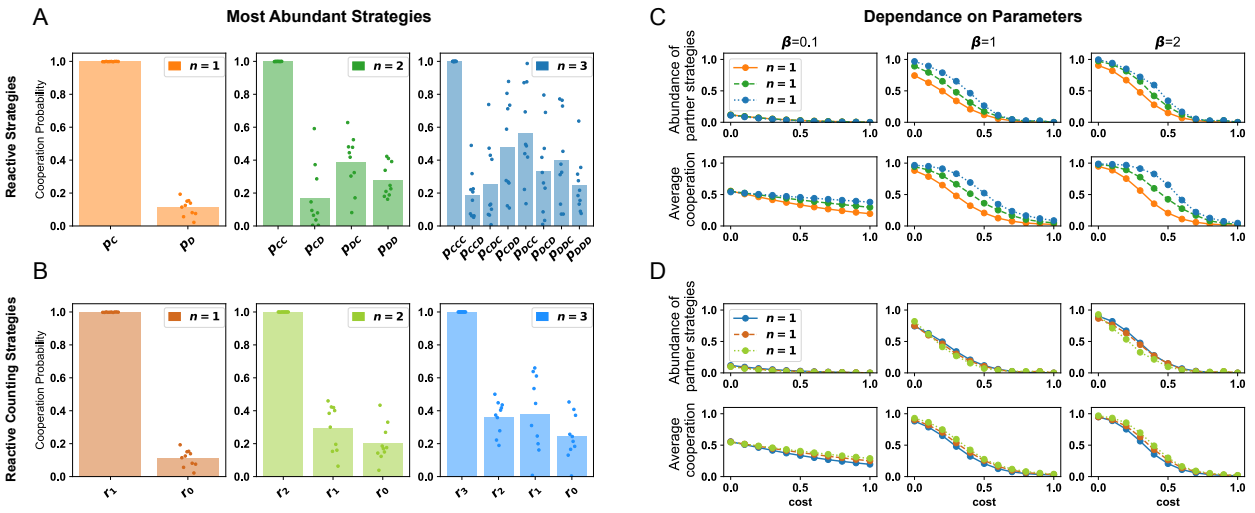

Figure 3: Evolutionary dynamics of reactive-n strategies. To explore the evolutionary dynamics among reactive-n strategies, we run simulations based on the method of Imhof and Nowak [57]. This method assumes rare mutations. Every time a mutant strategy appears, it goes extinct or fixes before the arrival of the next mutant strategy. **A,B,** We consider 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 range of different cooperation costs and selection strengths. In all cases, we only observe high cooperation rates when partner strategies evolve. Simulations are based on a donation game with $b = 1$, $c = .5$, and a selection strength parameter $\beta = 1$, 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**: Could we correct the legend in panels C and D? Also, I think the figure looks nicer if the legend does not have a grey box around it.