Reactive strategies with longer memory

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

Abstract

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-counting-n 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-counting-n 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.

1 Introduction

The emergence of cooperation in social interactions can be explained by the concept of direct reciprocity, where individuals assist each other through repeated encounters (1–3). The framework commonly used to model these interactions is the Prisoner's Dilemma. In this model, two participants, referred to as players, engage in a series of repeated interactions. During each interaction, they must decide whether to cooperate or defect. Cooperation typically results in more favorable outcomes for both players, but the self-interest of each individual often leads to a temptation to defect.

Strategies in the repeated Prisoner's Dilemma can vary in complexity. Some strategies are straightforward, including zero-memory strategies like ALLC and ALLD, as well as memory-1 strategies such as Tit For Tat (1) and Win Stay Lose Shift (4). On the other hand, some strategies are more sophisticated, considering multiple past interactions or additional information, such as the history of defections by both players (5–7). Empirical evidence supports the idea that humans do not typically employ zero-memory strategies; instead, they tend to favor conditional cooperation strategies (8–10). However, when it comes to the memory size of strategies used by humans, there is conflicting evidence. Some studies suggest that many human strategies align with those based solely on the most recent interaction (11–14). Nonetheless, there is also evidence of strategies that take into account more than just the previous round (15, 16).

Theoretical models have primarily focused on memory-1 strategies (4, 17–27) due to their mathematical tractability. This is because memory-1 strategies can be described by four parameters, specifically, the probabilities of cooperating after each possible outcome of the last round. Previous work has explored the entire space of memory-1 strategies and characterized when strategies are Nash equilibria. Furthermore, previous work has uncovered other interesting properties of these Nash strategies. Examples include zero-determinant strategies, which constitute a set of strategies capable of enforcing a linear relationship between the payoffs of the two players (19), equalizers, a set of strategies that equalize the co-player's score, assigning it a predetermined value independent of the co-player's strategy (24), and partner strategies, which ensure mutual cooperation without exploitation (26).

There have been efforts to expand these findings into strategy sets with greater memory. In the study by (28), they explore zero-determinant strategies within the context of memory-two strategies. Meanwhile, (29) delve into cooperative strategies applicable to memory-2 and memory-3. Despite the complexity, they successfully demonstrate that a specific set of strategies can be classified as Nash, effectively pinpointing locations within the strategy space. Working with higher memory strategies is not a trivial matter. Memory-1 strategies are defined by four parameters, but in the case of memory-2, one must consider 16 parameters, and 256 parameters in the case of memory-three. Thus, the space of strategies expands exponentially, making it challenging to derive analytical results.

Herein, we approach higher memory strategies by focusing on a specific set of memory$-n$ strategies that respond exclusively to one player's actions. Two such sets exist. The first comprises reactive strategies, which solely take into account the co-player's actions in previous turns. The second is the set of self-reactive strategies, which consider the focal player's actions. It can be easily demonstrated that in the case of self-reactive strategies, the Nash equilibrium is consistently one of always defecting. Consequently, our focus is directed toward reactive strategies.

Using reactive strategies, we can characterize partner strategies for both reactive-2 and reactive-3 strategies. For a particular class of reactive counting strategies, which involve counting cooperations rather than remembering the actual actions, we can identify partner strategies across all memory lengths. We establish this series of results by relying on a central finding: 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. To assess the evolutionary properties of partner strategies, we conduct simulations of an evolutionary process. Our findings indicate that partner strategies are prevalent within the population, and they exhibit a higher likelihood of evolution compared to other strategies. Moreover, we observe that cooperation rates increase as the memory size grows. When it comes to counting strategies, the advantages of increased memory are rather limited. Partner strategies do not see a higher selection frequency, and the overall cooperation rate within the population remains almost the same.

2 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 $h^i = (a^i_{-n}, \dots, a^i_{-1}) \in \{C, D\}^n$ 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 (23, 30, 31), $\mathbf{p} = (p_C, p_D)$. 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.

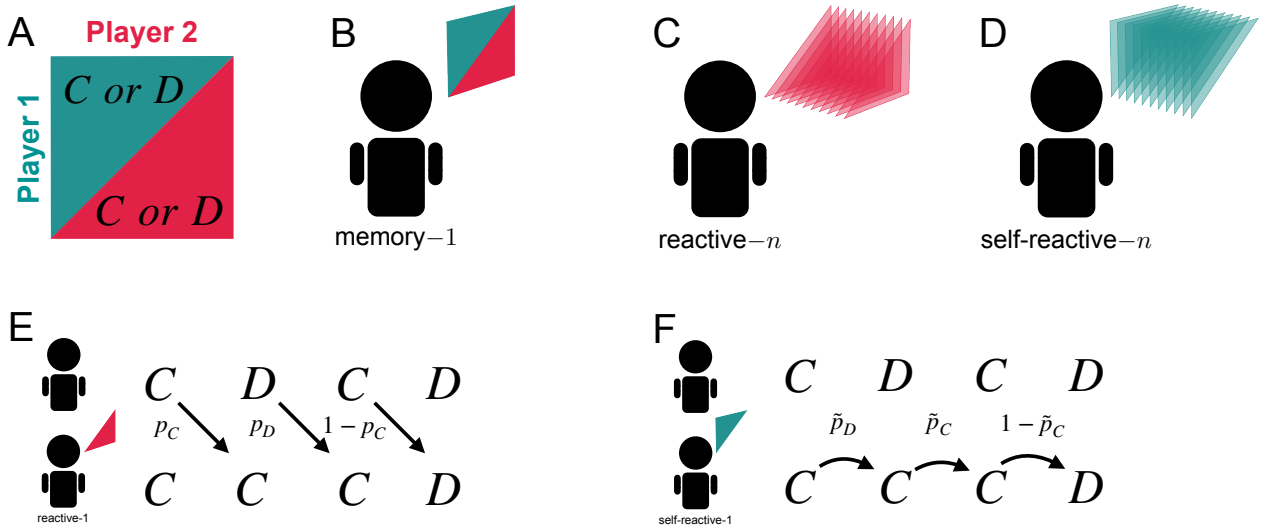

Figure 1: **A Graphical Representation of the Strategies Set. A.** In each turn of the repeated game, players 1 and 2 decide on an action, denoted as C (cooperate) or D (defect), respectively. We assume, that the information that a player can use in subsequent turns is limited to the actions taken by both players in the current turn. **B.** Memory-1 strategies, a well-studied set of strategies, utilize the actions of both players in the previous turn to make decisions. In the graphical representation of memory-1 strategies, we use a single square to illustrate this concept. **C.** This work primarily focuses on reactive-n strategies, which take into account only the actions of the co-players. **E.** For the case of $n = 1$, a reactive-1 strategy is represented as a vector $\mathbf{p} = (p_C, p_D)$, where p_C is the probability of cooperating given that the co-player cooperated, and p_D is the probability of cooperating given that the co-player defected. In the example shown, the bottom player employs a reactive-1 strategy. They cooperate with a probability p_C in the second round because the co-player cooperated in the first round. In the second round, the player cooperates with a probability p_D since the co-player previously defected. Finally, the player defects in the third round with a probability of $1 - p_C$, considering that the co-player cooperated. **D.** Another set of strategies we consider is that of self-reactive-n strategies, which rely solely on a player's own previous n actions. **F.** For the case of $n = 1$, a self-reactive-1 strategy is represented as a vector $\tilde{\mathbf{p}} = (\tilde{p}_C, \tilde{p}_D)$, where $\tilde{p}_C$ is the probability of cooperating given that the player's last action was cooperation, and $\tilde{p}_D$ is the probability of cooperating given that the player's last action was defection. In the example shown, the bottom player employs a self-reactive-1 strategy. They cooperate with a probability $\tilde{p}_D$ in the second round given that they defected in the first. In the second round, the player cooperates with a probability $\tilde{p}_C$ since they cooperated in the previous round. Finally, the player defects in the third round with a probability of $1 - \tilde{p}_C$, considering that they cooperated in the previous round.

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 (19), 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 (19). 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,

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

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 (27).

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,

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

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,

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

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,

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

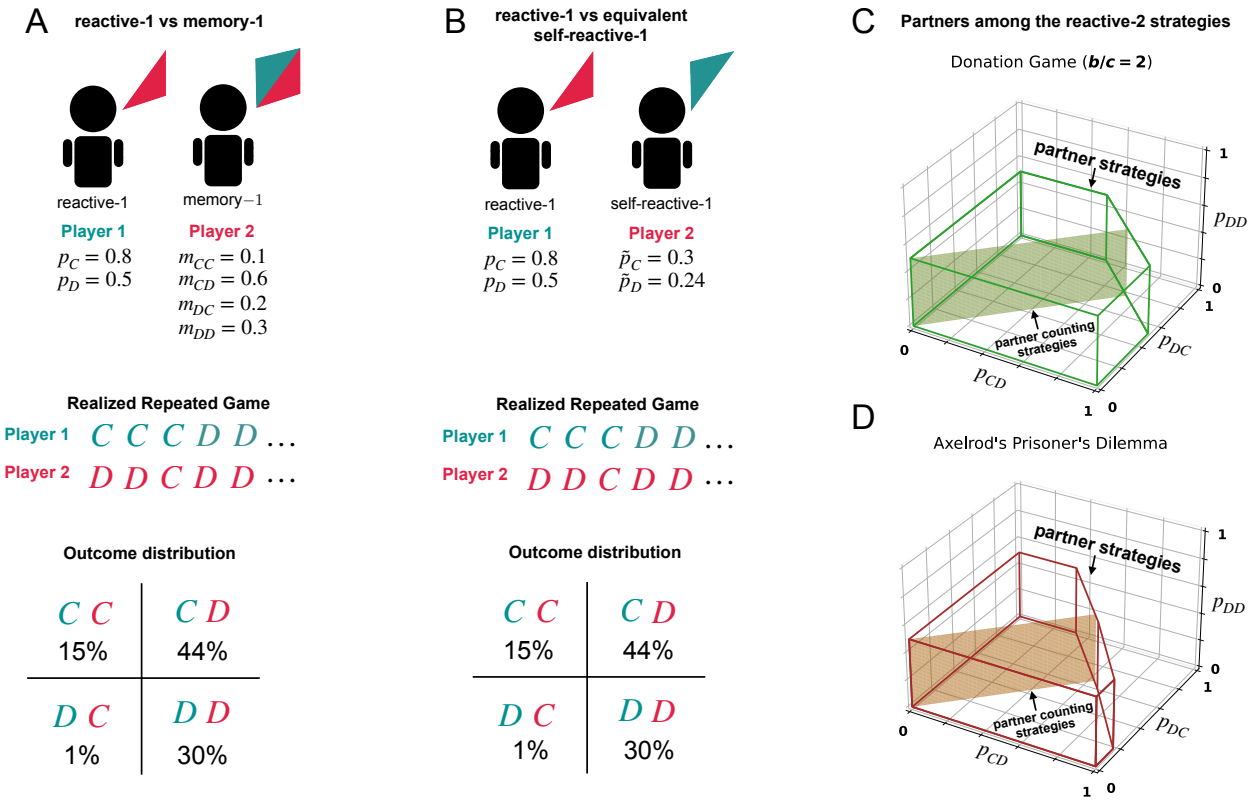

Figure 2: **Representation of Theoretical Results.** **A.** We have proven that if a player is using a reactive strategy $\mathbf{p}$, then the co-player using a memory-n strategy can switch to a self-reactive-n strategy without affecting the resulting payoffs. Namely, consider the case where player 1 uses the reactive-1 strategy $\mathbf{p} = (0.8, 0.5)$ and player 2 uses the memory-1 strategy $\mathbf{m} = (0.1, 0.6, 0.2, 0.3)$. A memory-1 strategy is defined by the vector $\mathbf{m} = (m_{CC}, m_{CD}, m_{DC}, m_{DD})$. Based on these strategies, the players interact in the repeated game. We can determine the long-term outcome, which indicates how often players are observed in any of the possible last-round outcomes. **B.** Using these outcomes and player 2's memory-1 strategy, we can compute an associated self-reactive strategy. The self-reactive strategy now consists of two cooperation probabilities, $\tilde{p}_C$ and $\tilde{p}_D$, which can be calculated as a weighted average of the respective memory-1 strategy's cooperation probabilities. The resulting reactive strategy for player 2 yields the same outcome distribution against player 2 as the original memory-1 strategy (see Supplementary Information for proof). **C.** For $c/b = 1/2$, we plot the space of reactive-2 partner strategies within the space of nice reactive-2 strategies. We can see that they form a three-dimensional polyhedron. Partner counting strategies are found in the overlapping region between the polyhedron and the counting strategies plane. **D.** For $R = 3, T = 5, P = 0, S = 1$, we explore the space of reactive-2 partner strategies. In the case of the general Prisoner's Dilemma, the entries of a nice reactive-2 strategy must satisfy a total of five conditions. However, when $R = 3, T = 5, P = 0, S = 1$, these conditions are reduced to three. Although the partner strategies also form a three-dimensional polyhedron, the shape differs from that of the donation game.

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 (32). The setup of the framework is outlined in Materials and Methods 4.

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.

3 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 (4). Experimental research has even suggested that these strategies are adopted by humans (11, 14). 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 (26), 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

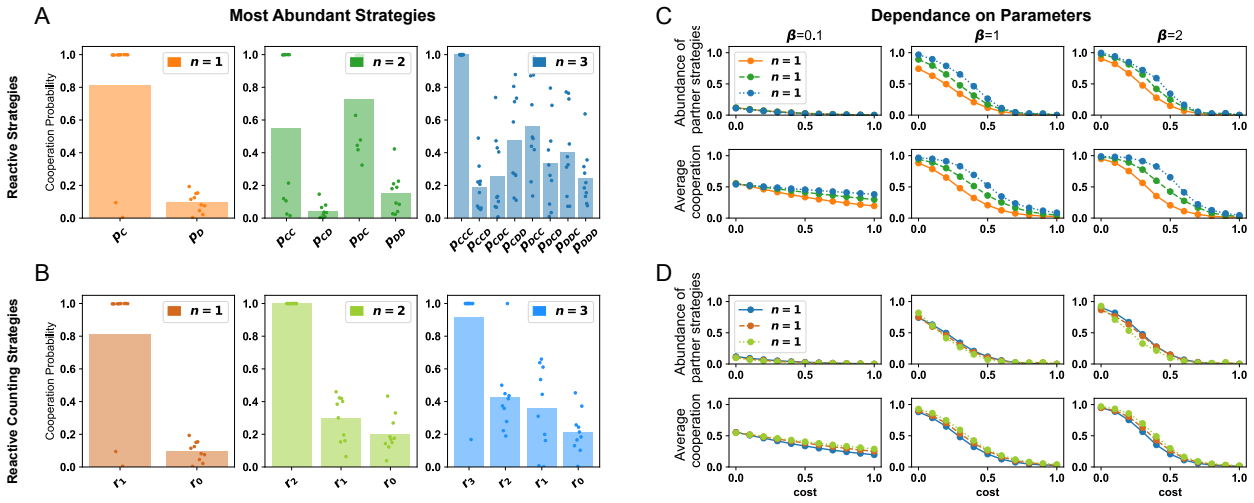

Figure 3: **Evolutionary Dynamics in the Space of Reactive Strategies.** In the previous section, we characterized partner strategies for reactive-2 and reactive-3. Additionally, we discussed the case of reactive counting strategies. Now, our focus is on assessing whether partner strategies can evolve in an evolutionary context. We ran simulations based on the methodology of Imhof and Nowak (32). In a single run of the evolutionary process, we recorded the cooperation probabilities of the resident at each elementary time step. **a-b. Most Abundant Reactive Strategies.** We performed 10 independent simulations for each set of strategies and documented the most abundant strategy for each run. The most abundant strategy is the resident that remained fixed for the most time steps. For these simulations, we used $b = 1$ and $c = .5$. For n equal to 1 and 2, $T = 10^7$ and for $n = 3$ then $T = 2 \times 10^7$. **c-d. Abundance of Partner Strategies.** We ran the evolutionary process once more, this time varying the cost (c) and the strength of selection (β). The cooperation benefit (b) remained fixed at a value of 1.

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.

4 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 $\mathbf{s}_{M,k}$, where k is the number of mutants in the population, and if residents get a payoff of $\mathbf{s}_{R,k}$, then the fixation probability ϕ_M of the mutant strategy can be calculated explicitly,

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

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.

References

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

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

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

[4] 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).

[5] Harper, M. *et al.* Reinforcement learning produces dominant strategies for the iterated prisoner's dilemma. *PloS one* **12**, e0188046 (2017).
[6] 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).
[7] Li, J. *et al.* Evolution of cooperation through cumulative reciprocity. *Nature Computational Science* **2**, 677–686 (2022).
[8] Fischbacher, U. & Gächter, S. Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *American economic review* **100**, 541–556 (2010).
[9] Rand, D. G. & Nowak, M. A. Human cooperation. *Trends in cognitive sciences* **17**, 413–425 (2013).
[10] Grujić, J. *et al.* A comparative analysis of spatial prisoner's dilemma experiments: Conditional cooperation and payoff irrelevance. *Scientific reports* **4**, 4615 (2014).
[11] Engle-Warnick, J. & Slonim, R. L. Inferring repeated-game strategies from actions: evidence from trust game experiments. *Economic theory* **28**, 603–632 (2006).
[12] Dal Bó, P. & Fréchette, G. R. The evolution of cooperation in infinitely repeated games: Experimental evidence. *American Economic Review* **101**, 411–429 (2011).
[13] Camera, G., Casari, M. & Bigoni, M. Cooperative strategies in anonymous economies: An experiment. *Games and Economic Behavior* **75**, 570–586 (2012).
[14] Bruttel, L. & Kamecke, U. Infinity in the lab. How do people play repeated games? *Theory and Decision* **72**, 205–219 (2012).
[15] 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).
[16] Romero, J. & Rosokha, Y. Constructing strategies in the indefinitely repeated prisoner's dilemma game. *European Economic Review* **104**, 185–219 (2018).
[17] Nowak, M. A. & Sigmund, K. Tit for tat in heterogeneous populations. *Nature* **355**, 250–253 (1992).
[18] 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).
[19] 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).
[20] Stewart, A. J. & Plotkin, J. B. Small groups and long memories promote cooperation. *Scientific reports* **6**, 1–11 (2016).
[21] Kraines, D. P. & Kraines, V. Y. Natural selection of memory-one strategies for the iterated prisoner's dilemma. *Journal of Theoretical Biology* **203**, 335–355 (2000).
[22] Imhof, L. A. & Nowak, M. A. Stochastic evolutionary dynamics of direct reciprocity. *Proceedings of the Royal Society B: Biological Sciences* **277**, 463–468 (2010).
[23] 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).
[24] Hilbe, C., Nowak, M. A. & Sigmund, K. Evolution of extortion in iterated prisoner's dilemma games. *Proceedings of the National Academy of Sciences* **110**, 6913–6918 (2013).
[25] Chen, X. & Fu, F. Outlearning extortioners: unbending strategies can foster reciprocal fairness and cooperation. *PNAS nexus* **2**, pgad176 (2023).
[26] Hilbe, C., Chatterjee, K. & Nowak, M. A. Partners and rivals in direct reciprocity. *Nature human behaviour* **2**, 469–477 (2018).
[27] Akin, E. The iterated prisoner's dilemma: good strategies and their dynamics. *Ergodic Theory, Advances in Dynamical Systems* 77–107 (2016).
[28] Ueda, M. Memory-two zero-determinant strategies in repeated games. *Royal Society open science* **8**, 202186 (2021).
[29] 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).
[30] Wahl, L. M. & Nowak, M. A. The continuous prisoner's dilemma: I. linear reactive strategies. *Journal of Theoretical Biology* **200**, 307–321 (1999).
[31] McAvoy, A. & Nowak, M. A. Reactive learning strategies for iterated games. *Proceedings of the Royal*

Society A **475**, 20180819 (2019).

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