$n-$bits reactive strategies in repeated games

Nikoleta E. Glynatsi, Christian Hilbe, Martin Nowak

1 Introduction

In this work we explore *reactive strategies* in the infinitely repeated prisoner's dilemma. The prisoner's dilemma is a two person symmetric game that provides a simple model of cooperation. Each of the two players, p and q, simultaneously and independently decide to cooperate (C) or to defect (D). A player who cooperates pays a cost $c > 0$ to provide a benefit $b > c$ for the co-player. A cooperator either gets $b - c$ (if the co-player also cooperates) or $-c$ (if the co-player defects). A defector either gets b or 0, and thus, the payoffs of player p take the form,

$$\begin{array}{c c} & \begin{array}{cc} \text{cooperate} & \text{defect} \end{array} \\ \begin{array}{c} \text{cooperate} \\ \text{defect} \end{array} & \begin{pmatrix} b-c & -c \\ c & 0 \end{pmatrix} \end{array} \tag{1}$$

The transpose of the matrix (1) are the payoffs of co-player q. Alternatively, we can define each player's payoffs as vectors by,

$$\mathbf{S}_p = (b-c, -c, b, 0) \quad \text{and} \quad \mathbf{S}_q = (b-c, b, -c, 0). \tag{2}$$

We study the infinitely repeated prisoner's dilemma. The infinitely repeated prisoner's dilemma consists of an infinite number of repetitions of the prisoner's dilemma (called the stage game).

1.1 Strategies

A strategy is a mapping from the entire history of play to an action of the stage game, and for the repeated prisoner's dilemma there are infinitely many strategies. Here we focus on *$n-$bit reactive strategies*; a special case of *memory-n strategies*. $n-$bit reactive strategies only respond to the co-player's previous n moves, whereas memory-n strategies take into account their own and the co-player's previous n moves.

Memory-n strategies are very well studied in the literature [Baek et al., 2016, Hilbe et al., 2017, Glynatsi and Knight, 2020, Press and Dyson, 2012, Stewart and Plotkin, 2016] with a major focus on memory-one strategies. Memory-one strategies consider only the outcome of the previous stage game to decide on an action, and have gained their attention mainly due to the mathematical tractability.

There are four possible outcomes to a one stage prisoner's dilemma, and with the outcomes listed in order as CC, CD, DC, DD, a memory-one strategy for p is a vector $\mathbf{p} = (p_1, p_2, p_3, p_4)$ where p_i is the probability of playing C when the i^{th} outcome occurred in the previous round. A play between two memory-one strategies,

$\mathbf{p} = (p_1, p_2, p_3, p_4)$ and $\mathbf{q} = (q_1, q_2, q_3, q_4)$, follows a Markov chain with four states, corresponding to the four possible outcomes, and with the transition matrix M. The invariant distribution $\mathbf{v}$ is the solution to $\mathbf{v}M = \mathbf{v}$, and it gives the probabilities that the players are in any of the states in the long run of the game.

$$M_1 = \begin{bmatrix} p_1 q_1 & p_1 (1 - q_1) & q_1 (1 - p_1) & (1 - p_1)(1 - q_1) \\ p_2 q_3 & p_2 (1 - q_3) & q_3 (1 - p_2) & (1 - p_2)(1 - q_3) \\ p_3 q_2 & p_3 (1 - q_2) & q_2 (1 - p_3) & (1 - p_3)(1 - q_2) \\ p_4 q_4 & p_4 (1 - q_4) & q_4 (1 - p_4) & (1 - p_4)(1 - q_4) \end{bmatrix}. \tag{3}$$

It is a known result that given the invariant distribution we can calculate the expected payoffs for each player, $s_{\mathbf{p}}$ and $s_{\mathbf{q}}$, as follows

$$\pi(\mathbf{p}, \mathbf{q}) = \mathbf{v} \cdot \mathbf{S}_p \text{ and } \pi(\mathbf{q}, \mathbf{p}) = \mathbf{v} \cdot \mathbf{S}_q.$$

As previously mentioned, reactive strategies are a subset of memory-n strategies, and consequently, one-bit reactive strategies are a subset of memory-one strategies. One-bit reactive strategies consider only the co-player's last action. Thus, a one-bit strategy treats the outcomes CC and CD as the same; the co-player cooperated in the last turn. In this case $p_1 = p_3$. Similarly, the probabilities of cooperating given that the co-player defected in the last turn are the same regardless of the strategy's own action; $p_2 = p_4$, and so $\mathbf{p} = (p_1, p_2, p_1, p_2)$. The methodology we have outlined here also applies to one-bit reactive strategies.

Reactive strategies have some attention in the literature [Sigmund, 1989, Wahl and Nowak, 1999], however, the majority of work focuses on one-bit reactive strategies. Even though for memory-n strategies previous work have shown that we can characterize Nash equilibria for an arbitrary memory size [Hilbe et al., 2017, Stewart and Plotkin, 2016] no similar work has been done on reactive strategies. The work of [Hilbe et al., 2017] have shown that in the case of memory-n strategies, more memory allows for more cooperation to evolve. We test this hypothesis in the case of reactive strategies. Previous work [Baek et al., 2016] have compared memory-n to reactive strategies, however, this was done for low memory strategies. The aim of this work is to extensively study reactive strategies of higher memory. We want to contribute to the ongoing discussions on (1) the effects of more memory on cooperation (2) the difference between memory-n strategies and reactive ones.

In section 2.1 we analytically characterize reactive strategies that are of Nash type and cooperative. In section 2.2 we show that pure $n-$bit reactive strategies can not can not sustain a cooperative Nash equilibrium. Finally, in section 2.3 we perform an evolutionary analysis, and investigate which strategies evolve.

2 Results

2.1 Cooperative Nash for Two-bit reactive strategies

2.1.1 Good n-bit reactive strategies

In [Akin, 2016], Akin gives the following definitions for memory-one strategies.

Definition 2.1. A memory-one strategy is **agreeable** if it always cooperates following a mutual cooperation, thus $p_1 = 1$.

Definition 2.2. A strategy for p is called **good** if (i) it is agreeable, and (ii) if for any general strategy chosen by q against it the expected payoffs satisfy:

$$s_{\mathbf{q}} \geq (b-c) \Rightarrow s_{\mathbf{q}} = s_{\mathbf{p}} = (b-c). \tag{4}$$

The strategy is of **Nash type** if (i) it is agreeable and (ii) if the expected payoffs against any q general strategy satisfy:

$$s_{\mathbf{q}} \geq R \Rightarrow s_{\mathbf{q}} = (b-c). \tag{5}$$

Hence, a strategy is good if the co-player achieves the reward payoff if and only if the focal player does as well. A Nash type strategy reassures that the co-player can never receive a payoff higher than $b-c$.

Notice that the definitions of good and Nash make no assumptions regarding the type of strategies the players need to play, and thus, these are extendable to $n-$ bit reactive strategies. The notion of agreeable strategies is generalized in the case of reactive strategies as follows.

Definition 2.3. A $n-$bit reactive strategy is agreeable if it cooperates with a probability one given that the co-player has consecutively cooperated in that last n rounds.

Following the introduction of these concepts, Akin proves Theorem 2.1.1 which he uses to characterize all memory-one strategies that are of *Nash type* and *good*.

Theorem 2.1. Akin's Theorem. Assume that player p uses the memory-one strategy $\tilde{\mathbf{p}} = \mathbf{p} - \mathbf{e}_{12}$ where $\mathbf{e}_{12} = (1,1,0,0)$, and q uses a strategy that leads to a sequence of distributions $\{\mathbf{v}^{(n)}, n = 1,2,...\}$ with $\mathbf{v}^{(k)}$ representing the distribution over the states in the k^{th} round of the game. Let $\mathbf{v}$ be an associated stationary distribution. Then,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbf{v}^{(n)} \cdot \tilde{\mathbf{p}} = 0, \text{ and therefore } \mathbf{v} \cdot \tilde{\mathbf{p}} = 0. \tag{6}$$

Akin's theorem is extendable to higher memory strategies; we demonstrate this for the special case of two-bit reactive strategies.

2.1.2 Two-bit reactive strategies

In the case of *two-bit reactive strategies*, players consider the last two actions of the co-player. Since for a single round there are 4 possible outcomes, for two rounds there will be 16 (4×4). We denote the possible outcomes as $E_pE_q|F_pF_q$ ($E_p, E_q, F_p, F_q \in \{C, D\}$) where the outcome of the previous round is E_pE_q and the outcome of the current round is F_pF_q. With the outcomes listed in order as $CC|CC, CC|CD, \ldots, DD|DC, DD|DD$ a two-bit reactive strategy for p is a vector $\mathbf{p} = (p_1, p_2, p_1, p_2, p_3, p_4, p_3, p_4, p_1, p_2, p_1, p_2, p_3, p_4, p_3, p_4)$ where

- p_1 is the probability of playing C when the last two actions of the co-player were CC,
- p_2 is the probability of playing C when the last two actions of the co-player were CD,

- p_3 is the probability of playing C when the last two actions of the co-player were DC,
- p_4 is the probability of playing C when the last two actions of the co-player were DD.

For simplicity, we denote a two-bit reactive strategy for p as $\hat{\mathbf{p}} = (p_1, p_2, p_3, p_4)$.

The play between two two-bits reactive strategies can be described by a Markov process with the transition matrix $\tilde{M}$.

$$\tilde{M} = \begin{pmatrix}
p_1q_1 & p_1(1-q_1) & (1-p_1)q_1 & (1-p_1)(1-q_1) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & p_2q_1 & p_2(1-q_1) & (1-p_2)q_1 & (1-p_2)(1-q_1) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_1q_2 & p_1(1-q_2) & (1-p_1)q_2 & (1-p_1)(1-q_2) & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_2q_2 & p_2(1-q_2) & (1-p_2)q_2 & (1-p_2)(1-q_2) \\
p_3q_1 & p_3(1-q_1) & (1-p_3)q_1 & (1-p_3)(1-q_1) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & p_4q_1 & p_4(1-q_1) & (1-p_4)q_1 & (1-p_4)(1-q_1) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_3q_2 & p_3(1-q_2) & (1-p_3)q_2 & (1-p_3)(1-q_2) & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_4q_2 & p_4(1-q_2) & (1-p_4)q_2 & (1-p_4)(1-q_2) \\
p_1q_3 & p_1(1-q_3) & (1-p_1)q_3 & (1-p_1)(1-q_3) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & p_2q_3 & p_2(1-q_3) & (1-p_2)q_3 & (1-p_2)(1-q_3) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_1q_4 & p_1(1-q_4) & (1-p_1)q_4 & (1-p_1)(1-q_4) & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_2q_4 & p_2(1-q_4) & (1-p_2)q_4 & (1-p_2)(1-q_4) \\
p_3q_3 & p_3(1-q_3) & (1-p_3)q_3 & (1-p_3)(1-q_3) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & p_4q_3 & p_4(1-q_3) & (1-p_4)q_3 & (1-p_4)(1-q_3) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_3q_4 & p_3(1-q_4) & (1-p_3)q_4 & (1-p_3)(1-q_4) & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_4q_4 & p_4(1-q_4) & (1-p_4)q_4 & (1-p_4)(1-q_4)
\end{pmatrix}.$$

Note that from state $CC|CC$ only the states $CC|CC, CC|CD, CC|DC, CC|DD$ are reachable. That is because the current outcome F_pF_q in this state has to match the previous outcome E_pE_q in the "next" state. Thus, in each row of the matrix there will be at most four non-zero elements. The invariant distribution $\tilde{\mathbf{v}}$ is the solution to $\tilde{\mathbf{v}}\tilde{M} = \tilde{\mathbf{v}}$.

In the infinitely repeated prisoner's dilemma, the probability that two players are in a CC state in the last round is the same as the probability of them being in a CC in the second to last round, thus, the following holds for $\tilde{\mathbf{v}}$,

$$\sum_{i,j \in \{C,D\}} \tilde{v}_{i,j|CD} = \sum_{i,j \in \{C,D\}} \tilde{v}_{CD|ij}. \tag{7}$$

The extension to Akin's Theorem (Theorem 2.1.1) is give by Lemma 2.2.

Lemma 2.2. Assume that player p uses a two-bit reactive strategy $\tilde{\mathbf{p}} = \mathbf{p} - \mathbf{e}_{i \in \{1,2,5,6,9,10,13,14\}}$, and q uses a strategy that leads to a sequence of distributions $\{\tilde{\mathbf{v}}^{(n)}, n = 1, 2, ...\}$ with $\tilde{\mathbf{v}}^{(k)}$ representing the distribution over the states in the k^{th} round of the game. Let $\tilde{\mathbf{v}}$ be an associated stationary distribution. Then

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \tilde{\mathbf{v}}^{(n)} \cdot \tilde{\mathbf{p}} = 0, \text{ and therefore } \tilde{\mathbf{v}} \cdot \tilde{\mathbf{p}} = 0.$$

$$\begin{aligned}
\tilde{\mathbf{v}}^{(n)} \cdot \tilde{\mathbf{p}} = 0 \Rightarrow & \\
& (\tilde{v}_1 + \tilde{v}_9)(1 - p_1) + (\tilde{v}_2 + \tilde{v}_{10})(1 - p_2) + (\tilde{v}_5 + \tilde{v}_{13})(1 - p_3) + (\tilde{v}_6 + \tilde{v}_{14})(1 - p_4) \\
& + (\tilde{v}_3 + \tilde{v}_{11})p_1 + (\tilde{v}_4 + \tilde{v}_{12})p_2 + (\tilde{v}_7 + \tilde{v}_{15})p_3 + (\tilde{v}_8 + \tilde{v}_{16})p_4 = 0. \qquad (8)
\end{aligned}$$

Proof. The probability that p cooperates in the n^{th} round, denoted by $\tilde{v}_{\text{C}}^{(n)}$, is $\tilde{v}_{\text{C}}^{(n)} = \tilde{v}_1^{(n)} + \tilde{v}_2^{(n)} + \tilde{v}_5^{(n)} + \tilde{v}_6^{(n)} + \tilde{v}_9^{(n)} + \tilde{v}_{10}^{(n)} + \tilde{v}_{13}^{(n)} + \tilde{v}_{14}^{(n)} = \tilde{\mathbf{v}} \cdot \mathbf{e}_{i \in \{1,2,5,6,9,10,13,14,\}}$. The probability that p cooperates in the $(n+1)^{th}$ round, denoted by $\tilde{v}_{\text{C}}^{(n+1)} = \tilde{v}^{(n)} \cdot \mathbf{p}$. Thus,

$$\tilde{v}_{\mathrm{C}}^{(n+1)} - \tilde{v}_{\mathrm{C}}^{(n)} = \tilde{\mathbf{v}}^{(\mathbf{n})} \cdot \mathbf{p} - \tilde{\mathbf{v}} \cdot \mathbf{e}_{i\in\{1,2,5,6,9,10,13,14\}} = \tilde{\mathbf{v}}^{(\mathbf{n})} \cdot (\mathbf{p} - \mathbf{e}_{i\in\{1,2,5,6,9,10,13,14\}}) = \tilde{\mathbf{v}}^{(n)} \cdot \tilde{\mathbf{p}}.$$

This implies $\tilde{v}_{\mathrm{C}}^{(n+1)} - \tilde{v}_{\mathrm{C}}^{(n)} = \sum_{k=1}^{n}(\tilde{v}_{\mathrm{C}}^{(k+1)} - \tilde{v}_{\mathrm{C}}^{(k)}) = \sum_{k=1}^{n}(\tilde{\mathbf{v}}^{(k)} \cdot \tilde{\mathbf{p}})$. Since $0 \leq \tilde{v}_{\mathrm{C}}^{(k)} \geq 1$ for any k,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \tilde{\mathbf{v}}^{(k)} \cdot \tilde{\mathbf{p}} = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} (\tilde{v}_{\mathrm{C}}^{(n+1)} - \tilde{v}_{\mathrm{C}}^{(1)}) = 0.$$

For the stationary distribution $\tilde{\mathbf{v}}$ that is the limit of some subsequence of the Cesaro averages $\{\frac{1}{n}\sum_{k=1}^{n}\tilde{\mathbf{v}}^{(k)}\}$, the continuity of the dot product implies $\tilde{\mathbf{v}} \cdot \tilde{\mathbf{p}} = 0$

□

We know that the invariant distribution combined with the payoff vectors give the expected payoffs for each player. In the case of the two-bit reactive strategies there can be two set of payoff vectors; (1) the payoffs are defined based on the outcome of the last round,

$$\begin{array}{l} \mathbf{S}_p = (b-c, \quad -c, \quad b, \quad 0, \quad b-c, \quad -c, \quad b, \quad 0, \quad b-c, \quad -c, \quad b, \quad 0, \quad b-c, \quad -c, \quad b, \quad 0)\,, \\ \mathbf{S}_q = (b-c, \quad b, \quad -c, \quad 0, \quad b-c, \quad b, \quad -c, \quad 0, \quad b-c, \quad b, \quad -c, \quad 0, \quad b-c, \quad b, \quad -c, \quad 0). \end{array} \tag{9}$$

(2) the payoffs are defined based on the outcome of the second to last round,

$$\begin{array}{l} \mathbf{S}'_p = (b-c, \quad b-c, \quad b-c, \quad b-c, \quad -c, \quad -c, \quad -c, \quad -c, \quad b, \quad b, \quad b, \quad b, \quad 0, \quad 0, \quad 0, \quad 0)\,, \\ \mathbf{S}'_q = (b-c, \quad b-c, \quad b-c, \quad b-c, \quad b, \quad b, \quad b, \quad b, \quad -c, \quad -c, \quad -c, \quad -c, \quad 0, \quad 0, \quad 0, \quad 0). \end{array} \tag{10}$$

Note that $\mathbf{s_p} = \tilde{\mathbf{v}} \times \mathbf{S}_p = \tilde{\mathbf{v}} \times \mathbf{S}'_p$ and $\mathbf{s_q} = \tilde{\mathbf{v}} \times \mathbf{S}_q = \tilde{\mathbf{v}} \times \mathbf{S}'_q$. From hereupon we consider the payoff vectors $\mathbf{S}_p$ and $\mathbf{S}_q$ unless stated otherwise.

2.1.3 Good and Nash type two-bit reactive strategies

We are interested in which two-bit reactive strategies can sustain a Nash equilibrium, and more specifically, a good/cooperative one. An agreeable strategy in the case of two-bit reactive strategies is a play that always cooperates after two consecutive cooperations of the co-player, thus $p_1 = 1$.

Theorem 2.3. Let the two-bit reactive strategy $\hat{\mathbf{p}} = (p_1, p_2, p_3, p_4)$ be an **agreeable strategy**; that is $p_1 = 1$. Strategy $\hat{\mathbf{p}}$ is **Nash** if the following inequalities hold:

$$p_4 \leq 1 - \frac{c}{b} \qquad p_2 \leq p_4 \qquad p_3 \leq 1 \qquad (1 + p_2) \leq \frac{b}{c} - \frac{p_4(b-c)}{c}$$

The agreeable strategy $\hat{\mathbf{p}}$ is good if and only if both inequalities above are strict.

Proof. We first eliminate the possibility $p_4 = 1$. If $p_4 = 1$, then $\hat{\mathbf{p}} = (1, p_2, p_3, 1)$. If against this q plays AllD $= (0, 0, 0, 0)$, then $\{CD\}$ is a terminal set and so with $s_{\mathbf{q}} = b$ and $s_{\mathbf{p}} = -c$. Hence, $\hat{\mathbf{p}}$ is not of Nash type.

We now assume $1 - p_4 > 0$. Observe that

$$
\begin{aligned}
s_{\mathbf{q}} - (b-c) &= \tilde{\mathbf{v}} \times \mathbf{S}_q - (b-c)\sum_{i=1}^{16} \tilde{v}_i \qquad (11)\\
&= (\tilde{v}_2 + \tilde{v}_6 + \tilde{v}_{10} + \tilde{v}_{14})c + (c-b)(\tilde{v}_4 + \tilde{v}_8 + \tilde{v}_{12} + \tilde{v}_{16}) - b(\tilde{v}_3 + \tilde{v}_7 + \tilde{v}_{11} + \tilde{v}_{15}).
\end{aligned}
$$

Multiplying by the positive quantity $(1 - p_4)$ and collecting terms, we have

$$
s_{\mathbf{q}} \geq (b-c) \Rightarrow \qquad (12)
$$

$$
(1-p_4)(\tilde{v}_6 + \tilde{v}_{14})c \geq -c(1-p_4)(\tilde{v}_2 + \tilde{v}_{10}) + (1-p_4)(-c+b)(\tilde{v}_4 + \tilde{v}_8 + \tilde{v}_{12} + \tilde{v}_{16}) + b(1-p_4)(\tilde{v}_3 + \tilde{v}_7 + \tilde{v}_{11} + \tilde{v}_{15}).
$$

Since $\tilde{p_1} = 0$, equation (8) implies

$$
(1-p_2)(\tilde{v}_{10}+\tilde{v}_2)+(1-p_3)(\tilde{v}_{13}+\tilde{v}_5)+(1-p_4)(\tilde{v}_{14}+\tilde{v}_6)-p_2(\tilde{v}_{12}+\tilde{v}_4)-p_3(\tilde{v}_{15}+\tilde{v}_7)-p_4(\tilde{v}_{16}+\tilde{v}_8)-\tilde{v}_{11}-\tilde{v}_3 = 0,
$$

and so,

$$
(1-p_4)(\tilde{v}_{14}+\tilde{v}_6) = -((1-p_2)(\tilde{v}_{10}+\tilde{v}_2)+(1-p_3)(\tilde{v}_{13}+\tilde{v}_5)-p_2(\tilde{v}_{12}+\tilde{v}_4)-p_3(\tilde{v}_{15}+\tilde{v}_7)-p_4(\tilde{v}_{16}+\tilde{v}_8)-\tilde{v}_{11}-\tilde{v}_3).
$$

Substituting in the above inequality and collecting terms we get

$$
A(\tilde{v}_{10} + \tilde{v}_2) + B(\tilde{v}_{12} + \tilde{v}_4) + C(\tilde{v}_{13} + \tilde{v}_5) + D(\tilde{v}_{15} + \tilde{v}_7) + E(\tilde{v}_{11} + \tilde{v}_{16} + \tilde{v}_3 + \tilde{v}_8) \geq 0 \qquad (13)
$$

with

$$
A = (c(p_2 - p_4)), \qquad B = (c(1 + p_2 - p_4) + b(-1 + p_4)), \qquad C = (c(-1 + p_3)),
$$
$$
D = (cp_3 + b(-1 + p_4)), \qquad E = c + b(-1 + p_4).
$$

In the case where A, B, C, D and E are smaller than 0, condition (13) holds iff $\tilde{v}_2, \tilde{v}_3, \tilde{v}_4, \tilde{v}_5, \tilde{v}_7, \tilde{v}_8, \tilde{v}_{10}, \tilde{v}_{11}, \tilde{v}_{12}, \tilde{v}_{13}, \tilde{v}_{15}, \tilde{v}_{16} = 0$. This implies, that $(\tilde{v}_1 + \tilde{v}_9)(1 - p_1) + (\tilde{v}_6 + \tilde{v}_{14})(1 - p_4) = 0$. p_4 can not be 1, thus $\tilde{v}_6, \tilde{v}_{14} = 0$. This means $(\tilde{v}_1 + \tilde{v}_9) = 1$, so both players receive the reward payoff and $\hat{\mathbf{p}}$ is good.

For $A, B, C, D, E \leq 0$ we derive the following conditions,

$$
\begin{aligned}
p_4 &\leq 1 - \frac{c}{b} \qquad (14)\\
p_2 &\leq p_4 \qquad (15)\\
p_3 &\leq 1 \qquad (16)\\
(1 + p_2) &\leq \frac{b}{c} - \frac{p_4(b-c)}{c} \qquad (17)
\end{aligned}
$$

□

NG: The code for getting these conditions is in ‘src/mathemtica/Two bit reactive.nb‘.

We can also explore the space of two-bit reactive strategies numerically. For a random agreeable two-bit reactive strategy we can evaluate if it's Nash using the Algorithm 1.

Algorithm 1: Numerical evaluation for Nash.

for $i <$ *maximum number of points* **do**
  $\hat{\mathbf{p}} \leftarrow$ random: $\{\emptyset\} \;\rightarrow R^4_{[0,1]}$;
  $p_1 \leftarrow 1$;
  $L(\hat{\mathbf{p}}) = \{\hat{\mathbf{q}} \,|\, \pi(\hat{\mathbf{p}}, \hat{\mathbf{p}}) \geq \pi(\hat{\mathbf{q}}, \hat{\mathbf{p}}) \text{ for } \hat{\mathbf{q}} \in \{\hat{\mathbf{q}}_\mathbf{i} \in P([0,1]) \,|\, |\hat{\mathbf{q}}_\mathbf{i}| = 16\}\}$;
  if $L(\hat{\mathbf{p}}) = \emptyset$ **then**
    isNash ← True ;
  else
    isNash ← False ;
  end
  return ($\hat{\mathbf{p}}$, isNash) ;
end

$\{\hat{\mathbf{q}}_\mathbf{i} \in P([0,1]) \,|\, |\hat{\mathbf{q}}_\mathbf{i}| = 16\}$ is the set of all pure two-bit reactive strategies.

For a given set of parameter values ($b = 2$ and $c = 1$) we can illustrate the space of provably Nash, Figure 1A. The results of the numerical evaluation are shown in Figures 1B - C. As a proof of concept, the numerical evaluation confirms that the points which satisfy conditions (14) are indeed Nash. However, we can also observe that there are points in the two-bit space which are Nash that do not satisfy the conditions. Thus, conditions (14) are sufficient for Nash but they are not absolute. The final conclusion is that for a point to be Nash we do not need to check against all 16 pure two-bit strategies but only against two; AllD and N6= $(0, 1, 1, 0)$. Akin showed a similar result in [Akin, 2016].

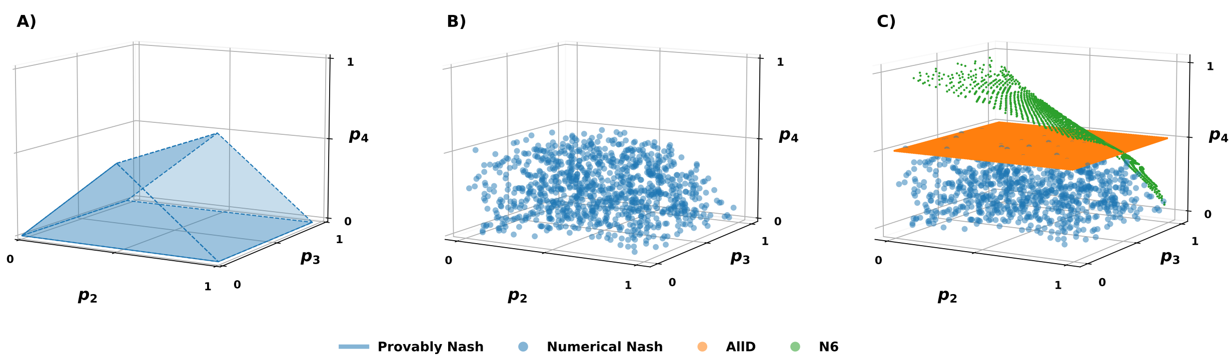

Figure 1: **Nash results for two-bit strategies. A) Provably Nash.** We have shown that if a two-bit reactive Nash is within this space, thus satisfy conditions (14), then it is Nash. **B) Numerical Nash.** The results from Algorithm 1. We evaluated 10^5 points in the space and here we are plotting 1000 randomly selected points. **C) The numerical results including the surfaces for AllD and N6.** The numerical results have shown that there are two pure strategies for which we have to check against to verify Nash. The planes are the planes for $\pi(\hat{\mathbf{p}}, \hat{\mathbf{p}}) - \pi(\hat{\mathbf{q}}_\mathbf{i}, \hat{\mathbf{p}}) = 0$ where $\hat{\mathbf{p}}$ is a generic agreeable two-bit reactive strategy and $\hat{\mathbf{q}}_\mathbf{i} \in \{\text{AllD, N6}\}$. Note that we do not plot p_1, as $p_1 = 1$. Parameters: $c = 1, b = 2$.

NG: The code for getting the plots is in ‘nbs/Two bit reactive Nash‘.

2.1.4 Some further thoughts

We have also performed the numerical evaluation for the prisoner's dilemma using the payoff matrix,

$$\begin{array}{cc} & \begin{array}{cc} \text{cooperate} & \text{defect} \end{array} \\ \begin{array}{c} \text{cooperate} \\ \text{defect} \end{array} & \left(\begin{array}{cc} R & 0 \\ 1 & P \end{array} \right) \end{array} \tag{18}$$

with $R = 0.6$ and $P = 0.1$. Though the results mainly remain similar, in the case of the prisoner's dilemma and for the given values, checking for Nash only against two strategies is not enough. In this case we also need to check for N7 which is the pure strategy given by $(0, 1, 1, 1)$.

NG: This is demonstrated in 'nbs/Two bit reactive Nash'.

Another thought. Two-bit reactive strategies are a specific case of memory-two strategies. The steps that we have taken in this section can also be applied to the case of memory-two strategies. However, now the strategies are 16 dimensional. Regardless, we can obtain some conditions.

NG: This is demonstrated in 'src/mathematica/Memory_two.nb'.

We expect that the methodology can be applied to $n-$bit reactive strategies and memory-n strategies. However, we expect that the space that we can prove is Nash is smaller and smaller in the feasible space of these strategies.

2.2 Pure $n-$bit Reactive Strategies with Errors

In this section we perform a numerical analysis to identify all strict Nash equilibria if players are restricted to pure strategies. To this end, we use a method defined by the work [Hilbe et al., 2017] which we will refer to as the Vaquero method. The method numerically identifies all (strict) Nash equilibria among a finite set of strategies, and moreover it returns for which parameter values the respective strategy is Nash. In our case we use the donation game thus the method returns which benefit-to-cost ratio $\frac{b}{c}$ is required for a strategy to be Nash.

2.2.1 The Vaquero Method

Let p and q play as $\mathbf{p}_\epsilon$ and $\mathbf{q}_\epsilon$ from a given set of strategies in a noisy environment with $\epsilon > 0$ where $\mathbf{p}_\epsilon = \epsilon(1 - \mathbf{p}) + (1 - \epsilon)\mathbf{p}$. Given the two strategies we numerically compute the three following measures:

- The fraction of rounds ρ in which player p cooperates against itself.
- The fraction of rounds $\tilde{\rho}_p$ in which player p cooperates against q.
- The fraction of rounds $\tilde{\rho}_q$ in which player q cooperates against p.

Given these measures the payoffs for p against itself can given by,

$$\pi(\mathbf{p}, \mathbf{p}) = b \cdot \rho - c \cdot \rho,$$

and the payoffs for q against p by,

$$\pi(\mathbf{q}, \mathbf{p}) = b \cdot \tilde{\rho}_p - c \cdot \tilde{\rho}_q.$$

For $\mathbf{p}_\epsilon$ to be a Nash equilibrium, it needs to be the case that $\pi(\mathbf{p}, \mathbf{p}) \geq \pi(\mathbf{q}, \mathbf{p})$, that is

$$b \cdot x_{\mathbf{p},\mathbf{q}} \geq c \cdot y_{\mathbf{q},\mathbf{p}} \tag{19}$$

where $x_{\mathbf{p},\mathbf{q}} = \rho - \tilde{\rho}_p$ and $y_{\mathbf{q},\mathbf{p}} = \rho - \tilde{\rho}_q$. For p to be a strict Nash equilibrium, the inequality (19) needs to be strict. Since $b > c > 0$, there are four possible cases

1. $x_{\mathbf{p},\mathbf{q}} > 0$ and $y_{\mathbf{q},\mathbf{p}} > 0$. In that case, $\mathbf{p}$ is stable against $\mathbf{q}$ if $b/c \geq y_{\mathbf{q},\mathbf{p}}/x_{\mathbf{p},\mathbf{q}}$ (and it is strictly stable if the inequality is strict).
2. $x_{\mathbf{p},\mathbf{q}} > 0$ and $y_{\mathbf{q},\mathbf{p}} \leq 0$. In that case, $\mathbf{p}$ is stable against $\mathbf{q}$ if $b/c \leq y_{\mathbf{q},\mathbf{p}}/x_{\mathbf{p},\mathbf{q}}$ (and it is strictly stable if the inequality is strict).
3. $x_{\mathbf{p},\mathbf{q}} \leq 0$ and $y_{\mathbf{q},\mathbf{p}} > 0$. In that case, $\mathbf{p}$ is never stable against $\mathbf{q}$, for no b/c ratio.
4. $x_{\mathbf{p},\mathbf{q}} \geq 0$ and $y_{\mathbf{q},\mathbf{p}} \leq 0$. In that case, $\mathbf{p}$ is stable against $\mathbf{q}$ for any b/c ratio.

Given the above we can define four sets:

$$\begin{align}
Q_1(p) &= \{q \mid x_{\mathbf{p},\mathbf{q}} > 0 \text{ and } y_{\mathbf{q},\mathbf{p}} > 0\}, \tag{20}\\
Q_2(p) &= \{q \mid x_{\mathbf{p},\mathbf{q}} < 0 \text{ and } y_{\mathbf{q},\mathbf{p}} \leq 0\}, \tag{21}\\
Q_3(p) &= \{q \mid x_{\mathbf{p},\mathbf{q}} \leq 0 \text{ and } y_{\mathbf{q},\mathbf{p}} > 0\}, \tag{22}\\
Q_4(p) &= \{q \mid x_{\mathbf{p},\mathbf{q}} = 0 \text{ and } y_{\mathbf{q},\mathbf{p}} = 0\}, \tag{23}\\
& \tag{24}
\end{align}$$

It follows that $\mathbf{p}$ is a Nash equilibrium if and only if $Q_3(p) = \emptyset$ and

$$\max\{\frac{y_{\mathbf{q},\mathbf{p}}}{x_{\mathbf{p},\mathbf{q}}} | q \in Q_1(p)\} \leq b/c \leq \min\{\frac{y_{\mathbf{q},\mathbf{p}}}{x_{\mathbf{p},\mathbf{q}}} | q \in Q_2(p)\}. \tag{25}$$

$\mathbf{p}$ is a strict Nash equilibrium if the inequalities in (25) are strict, $Q_3(p) = \emptyset$ and $Q_4(p) = \emptyset$.

2.2.2 Pure one, two and three bit(s) Reactive Strategies with Errors

The Vaquero method can be used with any set of strategies. The only constraint is that for two given strategies one should be able to calculate $\rho, \tilde{\rho}_p$ and $\tilde{\rho}_q$. For $n-$bit reactive strategies this is possible. For example consider the case of two-bit reactive cases where p plays as $\hat{\mathbf{p}}$. $\rho = \tilde{v}_1 + \tilde{v}_2 + \tilde{v}_5 + \tilde{v}_6 + \tilde{v}_9 + \tilde{v}_{10} + \tilde{v}_{13} + \tilde{v}_{14}$, given that $\tilde{\mathbf{v}}$ is the invariant distribution of the matrix $\tilde{M}_{|\hat{\mathbf{q}} = \hat{\mathbf{p}}}$.

We apply the Vaquero method and identify all the pure Nash equilibria in the case where players are allowed to choose from the sets of (i) one-bit (ii) two-bits and (ii) three-bits reactive strategies for a small error rate of $\epsilon = 0.01$. The results are given in Table 1.

	Strategy	ρ (self coop. rate)	Min. $\frac{b}{c}$ ratio	Max. $\frac{b}{c}$ ratio
One-bit reactive	$p_1 = 0, p_2 = 0$	0	0	0
Two-bit reactive	$p_1 = 0, p_2 = 0, p_3 = 0, p_4 = 0$	0.0	None	None
	$p_1 = 0, p_2 = 1, p_3 = 0, p_4 = 0$	0.255	1.04	None
	$p_1 = 0, p_2 = 0, p_3 = 1, p_4 = 0$	0.255	1.04	None
Three-bit reactive	$p_1 = 0, p_2 = 0, p_3 = 0, p_4 = 0, p_5 = 0, p_6 = 0, p_7 = 0, p_8 = 0$	0.0	None	None
	$p_1 = 0, p_2 = 0, p_3 = 0, p_4 = 0, p_5 = 0, p_6 = 1, p_7 = 0, p_8 = 0$	0.182	1.0590	1.0592
	$p_1 = 0, p_2 = 0, p_3 = 1, p_4 = 0, p_5 = 0, p_6 = 0, p_7 = 1, p_8 = 0$	0.255	1.041	1.042

Table 1: **Pure one, two and three bit(s) reactive strategies.** The Vaquero method allows us to numerically evaluate if pure strategies are Nash given that errors can occur. We performed the algorithm for a small percentage of error $\epsilon = 0.01$. The table shows all pure reactive strategies that are Nash, the $\frac{b}{c}$ ratio for which they are Nash and for each strategy the cooperating rate against itself. Overall, there are only a few reactive strategies that are Nash. In the case of two-bit reactive strategies, only three are Nash. In Hilbe et al. [2017] The method is applied to memory-two strategies and they show that there are 27 strategies that are Nash. This includes cooperative strategies ($\rho = 1$). In the case of reactive strategies, regardless of the memory size there are no cooperative strategies that sustain an equilibrium. For all Nash in this table $\rho \leq 0.255$. 0.255 corresponds to a quarter of cooperation. AllD is the only pure strategy that is Nash regardless of the memory size. In the case of two-bit strategies the only other strategies that are Nash are strategies that defect following a defection of the co-player. In the case of the three-bit reactive strategies only 3/64 strategies that can sustain an equilibrium, and for very few values of $\frac{b}{c}$ ratio. Thus, these strategies are not too robust in the sense that a small change in the payoff ratio can result in them not being Nash.

2.2.3 No Cooperative Nash in $n-$bit Reactive Strategies

Following the work of [Fundenberg and Maskin, 1990] for checking for stable strategies when actions are taken with a vanishingly small probability of error we show that no cooperative n-bit pure reactive strategy can sustain a Nash equilibrium, Lemma 2.4.

Lemma 2.4. In the space of n-bit pure reactive strategies, no strategy can sustain a cooperative Nash equilibrium.

Proof. Consider an agreeable $n-$bit reactive strategy $\mathbf{p}$. We already discussed that in the case of two-bit reactive strategies for a strategy to be Nash the probability of cooperating after two consecutive defections of the co-player has not different to 1. That is because against AllD in the long run the strategies will end up in $\{CD\}$ with a probability one. More generally, a $n-$bit reactive strategy that cooperates with a probability 1 after n consecutive defections, end ups in $\{CD\}$ against AllD. Thus, for $\mathbf{p}$ we know that $p_n \neq 1$ for $\mathbf{p}$ to be Nash. In the case of pure strategies $p_n \neq 1 \Rightarrow p_n = 0$.

Given this, let's define S as the set of all agreeable $n-$bit reactive strategies with $p_n = 0$. We will show that no strategy in S is not Nash when there is a vanishingly small probability of error. We consider the following cases.

Case 1: $n = 1$. In the case of $n = 1$ there are only four possible pure strategies and only two that can sustain a cooperative Nash. Those are *AllC* $= (1, 1)$ and Tit For Tat $= (1, 0)$. We know that AllC can not be Nash because if the co-player plays as AllD $\pi(\text{AllC, AllC}) < \pi(\text{AllD, AllC})$. In the case of Tit For Tat (TFT) we can show that

$$\pi(\text{TFT}, \text{TFT}) \geq \pi(\mathbf{q}, \text{TFT}) \text{ for } \mathbf{q} \in \{\text{AllC, AllC}, (0, 1)\}.$$

In [Fundenberg and Maskin, 1990] an evolutionary stable is a strategy for which given the lexicographic preferences we have assumed, s' can invade even if it performs worse to the. However, following the work [Fundenberg and Maskin, 1990] in the probability that a single error might occur observe that,

$$\text{TFT:}\{\ldots CDCDC\ldots\}$$
$$\text{TFT:}\{\ldots CCDCD\ldots\}$$
$$\pi(\text{TFT, TFT}) = \frac{(b-c)}{2},$$

whereas

$$\text{TFT:}\{\ldots CDCDC\ldots\}$$
$$\text{AllC:}\{\ldots CCCCC\ldots\}$$
$$\pi(\text{AllC, TFT}) = (b-c).$$

Thus, TFT is not Nash because $\pi(\text{TFT, TFT}) < \pi(\text{AllC, TFT})$.

Case 2: $n > 1$. In the case $n > 1$ we will use one specific strategy from S to show that no other strategy in S is Nash. This strategy is the strategy for which all conditional probabilities are 1 except p_n which is equal to 0. We will refer to this as Almost AllC.

Note that,

$$\pi(\mathbf{q}, \mathbf{q}) = \pi(\text{Almost AllC}, \mathbf{q}) = b - c \text{ for } \mathbf{q} \in \{S\}.$$

□

2.3 Evolutionary Dynamics

The results of our equilibrium analysis have shown that there are no pure cooperative Nash in the case of small errors. However, in the case of no errors we have proven that cooperative stochastic equilibria do exist when players choose from the set of two-bit reactive strategies. In this section, we explore whether cooperative equilibria evolve. Moreover, previous studies ([Hilbe et al., 2017]) have shown that in the case of memory-n strategies for intermediate b/c ratios, cooperation should more readily evolve among strategies with more memory. Here we also test if this result holds for reactive strategies.

To examine the evolutionary properties on $n-$bit reactive strategies, we perform an evolutionary study based on the framework of Imhof and Nowak [Imhof and Nowak, 2010]. 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 $[0, 1]$. 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,

$$\phi_M = \left(1 + \sum_{i=1}^{N-1} \prod_{j=1}^{i} \exp(-\beta(\pi_M(j) - \pi_R(i)))\right)^{-1} \tag{26}$$

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.

To study the effects of memory size we perform this evolutionary process when the population draws strategies from the sets of (i) one-bit (ii) two-bits and (ii) three-bits reactive strategies. We initially test the evolving cooperation rates for different selection strengths, Figure 2. To this end, we ran simulations for different b/c ratios. As expected, higher b/c values lead to more cooperation in all three spaces, and regardless of β's value. However, the more memory a strategy has it requires a lower benefit-to-cost ratio to achieve substantial cooperation. This verifies that the results of [Hilbe et al., 2017] also hold for reactive strategies.

We then explore the type of strategies that evolve for each set of reactive strategies, Figure 2. In all cases, the most abundant strategy achieves a high cooperation rate against itself. Notice that all most abundant strategies are the harsher when the co-player defects for the first time after a series of $n-1$ cooperations. We can observe that in both the case of the two-bits and three-bits, the strategies are more forgiving towards two defections.

NG: We ran these without error. Do we want to incooperate error in the evolutionary simulations?

3 Conclusion

In this work we have studied the space of $n-$bit reactive strategies. This space was originally explored by the work of [Nowak and Sigmund, 1990]. The reactive space contains many well known strategies from the literature, such as Alternator, Grudger, Tif For Tat and Generous Tit For Tat. However, note that these are reactive strategies of memory size one. We referred to these as one-bit reactive strategies. Here we aimed to explore higher memory reactive strategies, and even though this has been done previously for memory-n strategies, many questions still remain open in the case of reactive ones.

In section 2.1 we analytically explored two-bit reactive strategies. We built on the work of [Akin, 2016] and proved that there is a set of stochastic two-bit strategies that can sustain cooperative Nash equilibria. We verified our results with numerical simulations, and showed that in the space of two-bit strategies (for the donation game) one is Nash if it's Nash against AllD and (0, 1, 1, 0).

However, in the case of pure reactive strategies we showed that when there is a vanishingly small probability of error, no cooperative Nash is possible. We built on the work of [Hilbe et al., 2017] where they numerically showed that memory-n cooperative Nash are feasible. Thus, we can see that constraining the information a strategy receives to only the co-players moves makes it harder for cooperation.

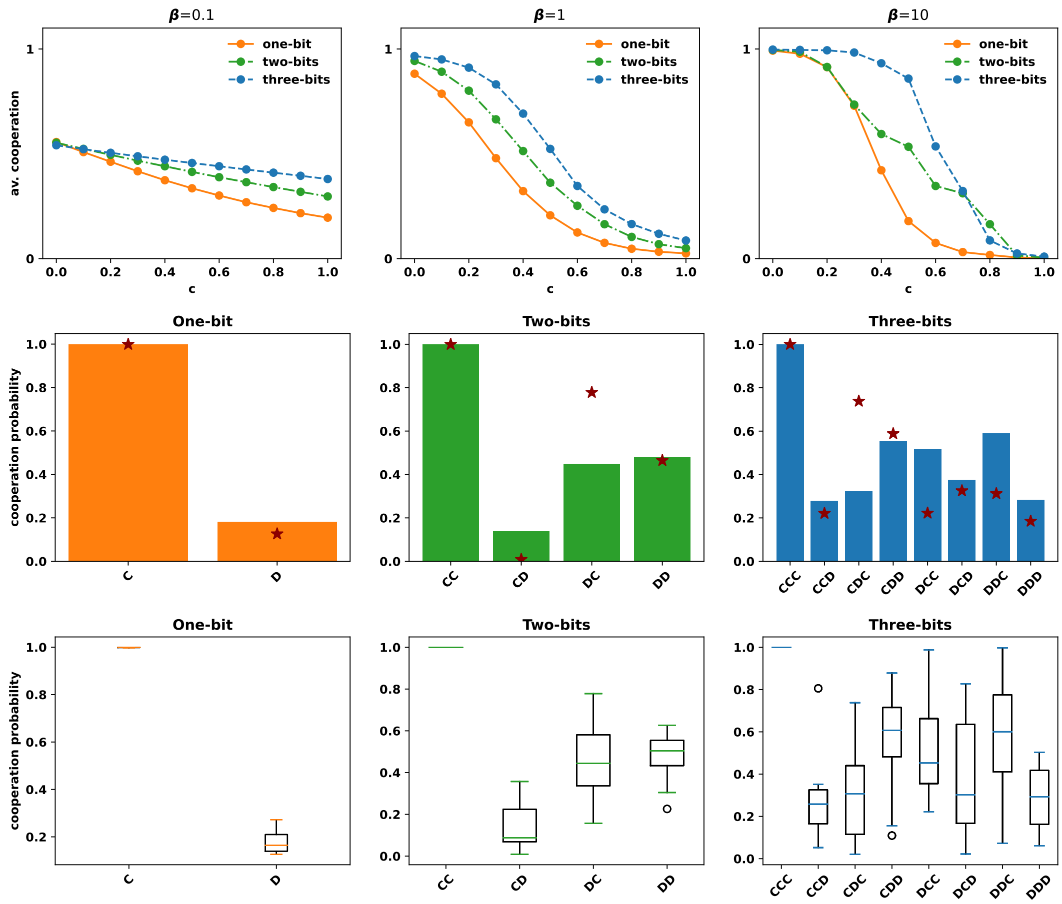

Figure 2: **Comparing the evolving cooperation rates and the most abundant strategies one-bit, two-bits, and three-bits strategies. A.** the evolving cooperation rates for different selection strengths. To assess the impact of memory on the evolution of cooperation, we ran simulations based on Imhof and Nowak for different benefit-to-cost ratios and different selection strengths. The average cooperation is calculated by considering the cooperation rate within the resident population. For a single run of the evolutionary process, we record the cooperating probabilities of the resident at each elementary time step. For each resident we estimate the cooperation rate between two resident strategies, and we take the average of that. **B and C.** We ran 10 independent simulations for each set of strategies and recorded the most abundant strategy for each run. The abundant strategy is the resident that was fixed for the most time steps. For the simulations we used $b = 3$ and $c = 1$. The colored bars show the average values of cooperation probabilities of the most abundant strategies. The stars show the cooperation probabilities of the most abundant strategy for each set. The boxplots illustrate the distributions of cooperation probabilities for the ten runs. Parameters: $N = 100$, $\beta = 1$. Each simulation was run for 10^5 mutant strategies except for the simulations where $\beta = 10$. These we run for $2 \cdot 10^5$.

In the last section 2.3, we explored the space of reactive strategies with evolutionary simulations. Though cooperative Nash can be obtained, here we asked the question: can they also evolve? In all the cases we have presented, high levels of cooperation can be achieved but larger memory allows for cooperation to emerge faster.

References

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

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

D. Fundenberg and E. Maskin. Evolution and cooperation in noisy repeated games. *The American Economic Review*, 80(2):274–279, 1990.

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

C. Hilbe, L. A. Martinez-Vaquero, K. Chatterjee, and M. A. Nowak. Memory-n strategies of direct reciprocity. *Proceedings of the National Academy of Sciences*, 114(18):4715–4720, 2017.

L. A. Imhof and M. A. Nowak. Stochastic evolutionary dynamics of direct reciprocity. *Proceedings of the Royal Society B: Biological Sciences*, 277(1680):463–468, 2010.

M. Nowak and K. Sigmund. The evolution of stochastic strategies in the prisoner's dilemma. *Acta Applicandae Mathematicae*, 20(3):247–265, 1990.

W. H. Press and F. J. Dyson. Iterated prisoner's dilemma contains strategies that dominate any evolutionary opponent. *Proceedings of the National Academy of Sciences*, 109(26):10409–10413, 2012.

K. Sigmund. Oscillations in the evolution of reciprocity. *J. theor. Biol*, 137:21–26, 1989.

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

L. M. Wahl and M. A. Nowak. The continuous prisoner's dilemma: I. linear reactive strategies. *Journal of Theoretical Biology*, 200(3):307–321, 1999.