$n-$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). Respectively, a defector either gets b (if the co-player cooperates) or 0 (if the co-player defects), and so, the payoffs of player p take the form,

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

The transpose of (1) gives the payoffs of co-player q. We can also define each player's payoffs as vectors,

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

In the one shot game the players' dominant choice is to defect (since $b > b - c$ and $0 > -c$), and once they reach mutual defection neither have a reason to deviate. Though the prisoner's dilemma has one nash equilibrium which is that of mutual defection, cooperative nash can subsist if we consider the iterated prisoner's dilemma [Axelrod and Hamilton, 1981, Hilbe et al., 2017]. We study cooperative nash in the infinitely repeated prisoner's dilemma when players can choose how to behave from specific sets of strategies. More specifically, we study if cooperative nash can exist when players use $n-$*bit reactive strategies* for $n > 1$.

1.1 Strategies

A strategy in the iterated prisoner's dilemma is a mapping from the entire history of play to an action of the prisoner's dilemma. For the iterated prisoner's dilemma there are infinitely many strategies, and 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 respond to the player's and co-player's 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 the case of memory-one strategies.

Memory-one strategies are attractive because they are mathematically tractable. For memory-one strategies it is possible to characterize all the cooperative nash equilibria [Akin, 2016], and furthermore, all nash equilibria that are achieved when players consider such strategies [Stewart and Plotkin, 2016]. In the case of memory-two strategies the work of [Hilbe et al., 2017] characterizes a set of cooperative nash equilibria. They show that pure strategies with three specific properties are subgame perfect equilibria. Moreover, they showed that in a setting with discounting these strategies are strict nash, and thus, evolutionary stable.

We build upon the previous work we have discussed here and aim to further extend the results to strategies with higher memory. However, as players are allowed to recall more of the previous rounds, the dimensions of the strategies space increases exponentially, such that analytical results or even simulation results become unattainable. To this end, we consider reactive strategies, a set of strategies that has not been given as much attention [Baek et al., 2016, Sigmund, 1989, Wahl and Nowak, 1999]. Reactive strategies remember the same number of rounds as memory$-n$ strategies, however, since they consider only the co-player's actions there are mathematically, and numerically more tractable.

In section 2 we introduce the methodology we will be using in this paper, and in section 3 we present the results. More specifically, in section 3.1 we analytically characterize strategies that can sustain cooperative nash in the case of two-bit reactive strategies, and memory-two. In section 3.2 we numerically characterize the full space on cooperative nash equilibria for two-bit reactive strategies. In section 3.5 we show that pure $n-$bit reactive strategies can not sustain a cooperative nash equilibrium. Finally, in section 3.6 we perform an evolutionary analysis for $n \in \{1, 2, 3\}$, and investigate which strategies evolve.

2 Methodology

In this section we describe our methodology, and we start by discussing the case of memory-one strategies.

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

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

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, and so, for a one-bit strategy the probabilities of cooperating following a CC and DC are the same since the co-player cooperated in the last turn. Similarly, for the probabilities of cooperating given that the co-player defected in the last turn. Thus, a one-bit reactive strategy for p is of the form $\mathbf{p} = (p_1, p_2, p_1, p_2)$.

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 general strategy used by q 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, and a Nash type strategy reassures that the co-player can never receive a payoff higher than $b - c$ (the payoff for mutual cooperation).

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 memory$-n$ and $n-$bit reactive strategies. The definition of agreeable strategies for a memory$-n$ strategy is given by Definition 2.3 and for a $n-$bit reactive strategy by Definition 2.4.

Definition 2.3. A memory-n strategy is **agreeable** if it always cooperates in the n opening rounds, and following n mutual cooperation(s) thus $p_1 = 1$.

Definition 2.4. A $n-$bit reactive strategy is agreeable if it cooperates with a probability one in the n opening rounds and if the co-player has consecutively cooperated in that last n rounds.

Following the introduction of these concepts, Akin proves Theorem 2 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 in section 3.1 for the special cases of two-bit reactive strategies.

In the case of $n = 2$, players consider the last two rounds. 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 memory-two strategy for p is a vector $\mathbf{p} = (p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8, p_9, p_{10}, p_{11}, p_{12}, p_{13}, p_{14}, p_{15}, p_{16})$. For 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 cooperating when the last two actions of the co-player were C and C, p_2 is the probability cooperating when the last two actions of the co-player were C and D, and so on. 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 a pair of two-bit 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 $\tilde{\mathbf{v}}$,

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

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,

$$\begin{aligned}
\mathbf{S}_p &= (b-c, \quad -c, \quad b, \quad 0, \quad b-c, \quad -c, \quad b, \quad 0, \quad b-c, \quad -c, \quad b, \quad 0, \quad b-c, \quad -c, \quad b, \quad 0)\,, \\
\mathbf{S}_q &= (b-c, \quad b, \quad -c, \quad 0, \quad b-c, \quad b, \quad -c, \quad 0, \quad b-c, \quad b, \quad -c, \quad 0, \quad b-c, \quad b, \quad -c, \quad 0).
\end{aligned} \tag{8}$$

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

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

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.

3 Results

In the following, we first characterize two-bit strategies that are of Nash type and good, and then we discuss pure reactive strategies in environments with noise. Lastly, we explore which reactive strategies evolve from an evolutionary process. We will also demonstrate how our results are applied to the case of memory-two strategies.

3.1 Good $n-$bit reactive and memory-n strategies for $n = 2$

We start by discussing the case of two-bit strategies. The extension to Akin's Theorem (Theorem 2) is given by Lemma 3.1.

Lemma 3.1. Assume that player p uses a two-bit reactive strategy $\tilde{\mathbf{p}} = \mathbf{p} - \hat{\mathbf{e}}_{12}$ (where $\hat{\mathbf{e}}_{12} = (1, 1, 1, 1, 0, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 0))$, 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

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

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

Proof. The probability that p cooperates in the n^{th} round, denoted by $\tilde{v}_{\text{C}}^{(n)}$, is $\tilde{v}_{\text{C}}^{(n)} = \tilde{v}_1^{(n)} + \tilde{v}_2^{(n)} + \tilde{v}_5^{(n)} + \tilde{v}_6^{(n)} + \tilde{v}_9^{(n)} + \tilde{v}_{10}^{(n)} + \tilde{v}_{13}^{(n)} + \tilde{v}_{14}^{(n)} = \tilde{\mathbf{v}} \cdot \hat{\mathbf{e}}_{12}$. 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,

$$\tilde{v}_{\text{C}}^{(n+1)} - \tilde{v}_{\text{C}}^{(n)} = \tilde{\mathbf{v}}^{(\mathbf{n})} \cdot \mathbf{p} - \tilde{\mathbf{v}} \cdot \hat{\mathbf{e}}_{12} = \tilde{\mathbf{v}}^{(\mathbf{n})} \cdot (\mathbf{p} - \hat{\mathbf{e}}_{12}) = \tilde{\mathbf{v}}^{(n)} \cdot \tilde{\mathbf{p}}.$$

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

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \tilde{\mathbf{v}}^{(k)} \cdot \tilde{\mathbf{p}} = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} (\tilde{v}_{\text{C}}^{(n+1)} - \tilde{v}_{\text{C}}^{(1)}) = 0. \tag{11}$$

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 can derive a further relation from Akin's Theorem using the equations (7). More specifically by substituting $v_1 + v_5 + v_9 + v_{13} = v_1 + v_2 + v_3 + v_4$ and $v_2 + v_6 + v_{10} + v_{14} = v_5 + v_6 + v_7 + v_8$ in (11) and some algebraic manipulation one can show that,

$$p_1(v_1+v_3+v_9+v_{11})+p_2(v_2+v_4+v_{10}+v_{12})+p_3(v_5+v_7+v_{13}+v_{15})+p_4(v_6+v_8+v_{14}+v_{16}) = (v_1+v_2+v_3+v_4+v_5+v_6+v_7+v_8). \tag{12}$$

Note that the right hand side of equation (12) is the focal's player's cooperation rate in the second to last round.

We are interested in which two-bit reactive strategies can sustain a Nash equilibrium, and more specifically, a good/cooperative one. We show that:

Theorem 3.2. 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 the 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

$$\begin{aligned} s_{\mathbf{q}} - (b-c) &= \tilde{\mathbf{v}} \times \mathbf{S}_q - (b-c)\sum_{i=1}^{16} \tilde{v}_i \\ &= (\tilde{v}_2 + \tilde{v}_6 + \tilde{v}_{10} + \tilde{v}_{14})c + (c-b)(\tilde{v}_4 + \tilde{v}_8 + \tilde{v}_{12} + \tilde{v}_{16}) - b(\tilde{v}_3 + \tilde{v}_7 + \tilde{v}_{11} + \tilde{v}_{15}). \end{aligned} \tag{13}$$

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

$$\begin{aligned} &s_{\mathbf{q}} \geq (b-c) \Rightarrow \\ &(1-p_4)(\tilde{v}_6 + \tilde{v}_{14})c \geq -c(1-p_4)(\tilde{v}_2 + \tilde{v}_{10}) + (1-p_4)(-c+b)(\tilde{v}_4 + \tilde{v}_8 + \tilde{v}_{12} + \tilde{v}_{16}) + b(1-p_4)(\tilde{v}_3 + \tilde{v}_7 + \tilde{v}_{11} + \tilde{v}_{15}). \end{aligned} \tag{14}$$

Since $\tilde{p_1} = 0$, equation (10) implies

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

and so,

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

Substituting (10) in the above inequality and collecting terms we get

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

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 strictly smaller than 0, condition (15) holds iff $\tilde{v}_2, \tilde{v}_3, \tilde{v}_4, \tilde{v}_5, \tilde{v}_7, \tilde{v}_8, \tilde{v}_{10}, \tilde{v}_{11}, \tilde{v}_{12}, \tilde{v}_{13}, \tilde{v}_{15}, \tilde{v}_1$ 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,

$$p_4 \leq 1 - \frac{c}{b} \tag{16}$$
$$p_2 \leq p_4 \tag{17}$$
$$p_3 \leq 1 \tag{18}$$
$$(1 + p_2) \leq \frac{b}{c} - \frac{p_4(b - c)}{c} \tag{19}$$

□

NG: The code for getting these conditions is in ‘src/mathematica/Two_bit_reactive_clean.nb‘. The mathematica file ‘Two bit reactive’ contains some of the other cases I have explored but it’s a bit of a mess.

We apply the same methodology for the case of memory-two strategies. This time we first eliminate the case of $p_6 = 1$. If $p_6 = 1$, then against AllD $\{CD\}$ is a terminal set and thus a strategy with $p_6 = 1$ can not be nash. The results are summarised by Theorem 3.3.

Theorem 3.3. Let the memory-two strategy $\mathbf{p} = (p_1, p_2, \dots, p_{16})$ be an **agreeable strategy**; that is $p_1 = 1$. Strategy $\mathbf{p}$ is **Nash** if the following inequalities hold:

$$p_2, p_{10}, p_{14} \leq p_6 \qquad p_5, p_9, p_{13} \leq 1 \qquad \frac{c}{b}p_{11}, \frac{c}{b}p_{15}, \frac{c}{b}p_3 \leq 1 - p_6$$
$$1 + \frac{c}{b - c}p_4, 1 + \frac{c}{b - c}p_{12}, 1 + \frac{c}{b - c}p_{16} \geq p_6$$
$$\frac{c}{b}p_7 \leq 1 - p_6 \qquad \frac{c}{b - c}p_8 \leq 1 - p_6$$

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

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

3.2 A numerical evaluation of good $(n = 2)$ reactive strategies

We can also explore which agreeable strategies are nash numerically. We take a random point $\hat{\mathbf{p}}$ in the space of two-bit reactive and we check if it’s nash. Thus we check if condition $\pi(\mathbf{q}, \hat{\mathbf{p}}) \leq (b - c)$ holds, against all pure memory-two strategies ($\mathbf{q} \in \binom{\{0,1\}}{16}$). It is sufficient to check only against the pure memory-two strategies based on the result of [McAvoy and Nowak, 2019] (see Lemma 2.1). We repeat this step for a large number of randomly selected strategies, and we record if the strategy is nash or not, and against which the pure strategies the condition for Nash is not satisfied. The process is described by Algorithm 1

Algorithm 1: Numerical evaluation for Nash.

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

We run the above algorithm for 10,000 random strategies for parameter values ($b = 2$ and $c = 1$). The results are shown in Figure 1. Figure 1**A)** is a visual representation of the area of strategies we can prove are good and Nash based on Theorem 3.2. Figure 1**B)** illustrates the set of random points that based on Algorithm 1 are Nash. We can observe that there are several points that the algorithm classifies as Nash which are not explained by Theorem 3.2. More specifically 70% of the points that are Nash based on the numerical method are outside the proved area. This concludes that the inequalities (16) are sufficiency for a point to be Nash but not necessary.

[Akin, 2016] showed that for a strategy to be a good Nash one does not have to check against all pure 16 memory-one strategies, but it is efficient to check against only two. These strategies are AllD and (0, 1, 1, 1). We can derive a similar result for the case of two-bit strategies, more specifically, we derive Lemma 3.4, and the illustration is given also in Figure 1**B)**.

Lemma 3.4. Let the two-bit reactive strategy $\hat{\mathbf{p}} = (p_1, p_2, p_3, p_4)$ be an **agreeable strategy**; that is $p_1 = 1$. For $\hat{\mathbf{p}}$ to be Nash the following inequalities must hold:

$$\begin{aligned} \pi(\text{AllD}, \hat{\mathbf{p}}) \leq (b-c) \quad &\text{and} \quad \pi(\text{Alternator}, \hat{\mathbf{p}}) \leq (b-c) \Rightarrow \\ bp_4 \leq (b-c) \quad &\text{and} \quad \frac{bp_2 + bp_3 - c}{2} \leq (b-c) \Rightarrow \\ p_4 \leq 1 - \frac{c}{b} \quad &\text{and} \quad p_2 + p_3 \leq 1 + \frac{b-c}{c} \end{aligned}$$

where AllD$= (0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)$ and Alternator$= (0,0,1,1,0,0,1,1,0,0,1,1,0,0,1,1)$.

We derive the conclusion that these two strategies are efficient from the numerical results. Let all the non nash points be a set of elements $U = \{\hat{\mathbf{p}}^{(1)}, \hat{\mathbf{p}}^{(2)}, \dots, \hat{\mathbf{p}}^{(m)}\}$ which we refer to as the universe. For each of the pure strategies we create a set that contains all the points for which the nash condition failed against the specific strategy. Let S be the collection of these sets such that,

$$S = \{\{\mathbf{l} | \pi(\mathbf{q}, \mathbf{l}) > (b-c) \quad \forall \quad \mathbf{l} \in U\} \quad \forall \quad \mathbf{q} \in \binom{\{0,1\}}{16}\}. \tag{20}$$

We want to identify the smallest sub-collection of S whose union equals the universe.

For example, consider the universe $Q = \{1, 2, 3, 4, 5\}$ and the collection of sets $S = \{\{1, 2, 3\}, \{2, 4\}, \{3, 4\}, \{4, 5\}\}$. Clearly the union of S is Q. However, we can cover all of the elements with the following, smaller number of sets is two: $\{\{1, 2, 3\}, \{4, 5\}\}$. This is a classical question in combinatorics, computer science, operations research, and complexity theory, and it is known as the set cover problem [Beasley, 1987].

We identified that there are several pairs of strategies whose union is equal to U. One of such pairs is AllD and Alternator.

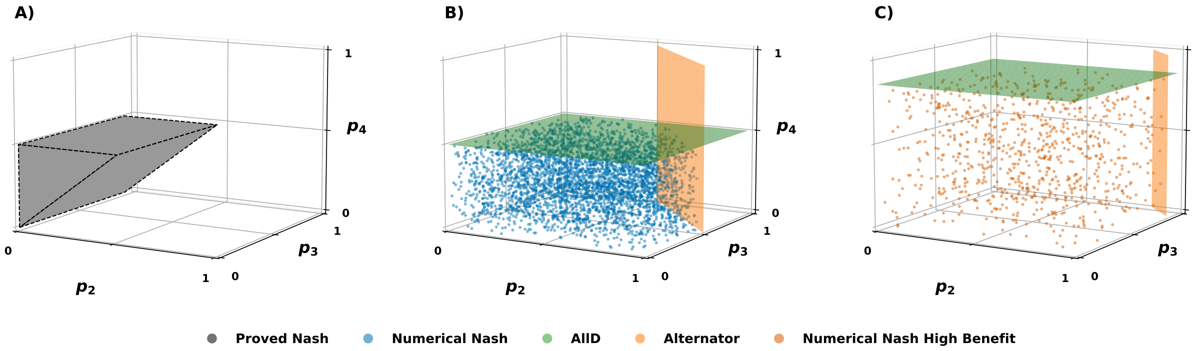

Figure 1: **Nash results for two-bit strategies. A) Proved Nash.** We have shown that if a two-bit reactive strategy is within this space, thus satisfies conditions (16), then it is good Nash. **B) Numerical nash results.** The results from Algorithm 1. We evaluated 10^4 points in the space. The numerical results have shown that there are two pure strategies that constrain the nash space; these are AllD and Alternator. The equations for the planes are obtained by solving $\pi(\mathbf{q}, \hat{\mathbf{p}}) = (b-c)$. The equations are $p_4 = 1 - \frac{c}{b}$ and $p_3 = 1 + \frac{b-c}{c} - p_2$. Parameters: $c = 1, b = 2$. **C) Numerical nash for high benefit.** We repeat Algorithm 1 for a higher value of benefit ($b = 7$) for 10^3 random points. We can see that the strategies AllD and Alternator still constrain the space of possible nash. Note that we do not plot p_1 for any of the above plots, since $p_1 = 1$.

NG: The code for checking Lemma can be found in the notebook. The code for plotting Figure 1 in.

3.3 A numerical evaluation of good ($n = 2$) reactive strategies for the prisoner's dilemma

The payoff matrix (1) is a special case of the prisoner's dilemma known as the donation game. In the general case the prisoner's dilemma the payoffs of a player p are given by the following payoff matrix,

$$\begin{array}{c} \\ \text{cooperate} \\ \text{defect} \end{array} \begin{array}{c} \begin{array}{cc} \text{cooperate} & \text{defect} \end{array} \\ \begin{pmatrix} R & T \\ S & P \end{pmatrix} \end{array} \tag{21}$$

where R is the reward for mutual cooperation, T is the temptation to defect, S is the payoff of the sucker and P is the punishment for mutual defection. We have also carried out the numerical evaluation of nash using the payoff matrix (21). Though the results remain fairly similar, there is no pair of strategies that are efficient to check for nash this time. More specifically for a given set of values, $R = 0.6, T = 1, S = 0$

and $P = 0.1$, only a triple of strategies can verify that a random point is nash. These are AllD, Delayed Alternator ((0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1)) and N8898 ((0, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 1, 0)). N8898 is a strategy that cooperates if in the last round the outcome was CD except in the case where the second to last outcome was CD. Then cooperate if cooperated in the last turn. The results are summarised in Lemma 3.5 and an illustrations of the results is shown in Figure 2.

Lemma 3.5. Let the two-bit reactive strategy $\hat{\mathbf{p}} = (p_1, p_2, p_3, p_4)$ be an **agreeable strategy**; that is $p_1 = 1$. For $\hat{\mathbf{p}}$ to be Nash the following inequalities must hold:

$$\begin{array}{rrr}
\pi(\text{AllD}, \hat{\mathbf{p}}) \leq R \text{ and} & \pi(\text{Delayed Alternator}, \hat{\mathbf{p}}) \leq R \text{ and} & \pi(\text{N8898}, \hat{\mathbf{p}}) \leq R \Rightarrow \\
P(1 - p_4) + p_4 \leq R \text{ and} & -\dfrac{P(p_2 - 1) - R(p_3 - p_4) - p_2 - 1}{4} \leq R \text{ and} & \dfrac{R(p_2 + p_3) + 1}{3} \leq R \Rightarrow \\
p_4 \leq \dfrac{P}{(P - 1)} \text{ and} & (p_3 + p_4) \geq \dfrac{P(p_2 - 1) + 4R - p_2 - 1}{R} \text{ and} & p_2 + p_3 \leq 3 - \dfrac{1}{R}
\end{array}$$

where AllD$= (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$, Delayed Alternator $= (0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1)$ and N8898 $= (0, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 1, 0)$.

To derive to this triple we consider the set cover problem approach, similar to Lemma 3.4.

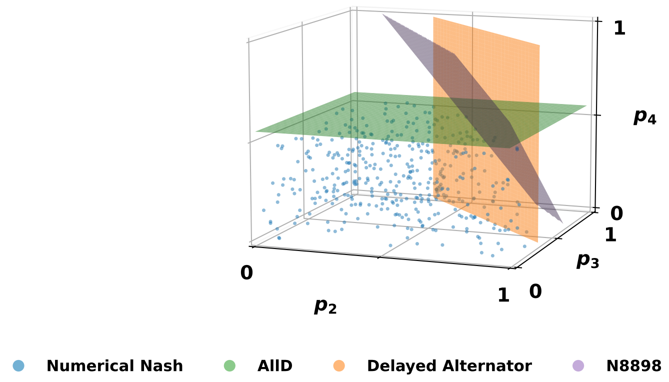

Figure 2: **Nash results for two-bit strategies for the prisoner's dilemma.** We performed Algorithm 1 for 10^3 random points. Now the payoffs are given by matrix (21) with $R = 0.6, T = 1, S = 0$ and $P = 0.1$. There are several triples of strategies that can constrain the nash space, such a triple is AllD, Delayed Alternator and N8898. To derive this triple we followed the same approach as in section 3.2.

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

3.4 A numerical evaluation of good memory-two strategies

We can repeat the same process for the case of memory-two strategies for the donation game. We evaluated 10^3 random points for the parameter values $b = 2$ and $c = 1$. The numerical results show that in the case of memory-two strategies, Theorem 3.3, is also sufficient but not necessary. We hypothesise that even though our methodology can be applied to greater memory sizes, the space that the derived conditions will explain will reflect a smaller and smaller portion of the feasible nash space.

We have also tried to obtain the smallest possible subset of strategies that can constrain the space of nash, similar to the donation game (AllD and Alternator) and the prisoner's dilemma (AllD, Delayed Alternator and N8898) for two-bit reactive strategies. We can show that a pair of such strategies does not exist. Excluding AllD, there are a total of 35,766 strategies that their sets' combination could potentially be the smallest sub-collection necessary to explain nash. However, this is a problem that we can not solve explicitly. Instead we have to consider an approximate solution using a greedy algorithm.

Initially, we remove all the elements in the elements from the universe that are covered by AllD. The algorithm takes an input a starting strategy (and subsequently its set), and thereafter at each stage it chooses the strategy that contains the largest number of uncovered elements. In the case that two or more strategies have contain the same number of uncovered elements, the algorithm randomly picks one. We run the algorithm for 3467 initial conditions, and for each we repeat the process 20 times. The initial conditions were chosen based on the largest number of uncovered elements. For each run we record the output subset and its size. The results are given by Table 1. Note that the minimum subset size includes AllD that we initially excluded.

Min subset size	Num. of times solution was reached
15	3899
14	19856
13	28385
12	14876
11	1964

Table 1: **Results of greedy algorithm.** The greedy algorithm was used to find the smallest possible subset of pure memory-two strategies that can constrain the space of nash in the case of memory-two strategies. Based on the approximate solution, the smallest subset is of size 11.

3.5 Pure $n-$bit Reactive Strategies with Errors

The aim of the section is to explore if there are pure reactive strategies that are evolutionary stable. In this section we consider the donation game with noise. Noise is a given probability ϵ that a player's action is flipped, such that if a player intends to cooperate the defect with a probability ϵ.

The work of [Hilbe et al., 2017] introduced a method that can identify all (strict) Nash equilibria among a finite set of strategies, and for which parameter values the respective strategy is stable (i.e., which benefit-to-cost ratio b/c is required in the donation game). We refer to this method as the Martinez-Vaquero method, and a description of the method can be found in Appendix A. Note that one of the assumptions of the method is that $\epsilon \neq 0$.

We apply the Martinez-Vaquero method and identify all the pure strict 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 2.

In the case of reactive strategies there are no cooperative pure strategies that are evolutionary stable in the presence of noise. This could indicate that in the case of reactive strategies, where memory is even more limited compared to memory-n strategies, stochasticity can be important.

NG: Code.

	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 2: **Pure one, two and three bit(s) reactive strategies.** The Martinez-Vaquero method allows us to numerically evaluate if pure strategies are evolutionary stable 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 strict 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 there are only three. 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 strategies in this table $\rho \leq 0.255$. 0.255 corresponds to a quarter of cooperation. AllD is the only pure strategy that is stable regardless of the memory size. In the case of two-bit strategies the only other strategies that are stable 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 results in them not being stable.

3.6 Evolutionary Dynamics

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,

$$\phi_M = \left(1 + \sum_{i=1}^{N-1} \prod_{j=1}^{i} \exp(-\beta(\pi_M(j) - \pi_R(i)))\right)^{-1} \tag{22}$$

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 3. 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 3. 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?

4 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 3.1 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.

In the last section 3.6, 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.

A The Martinez-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:

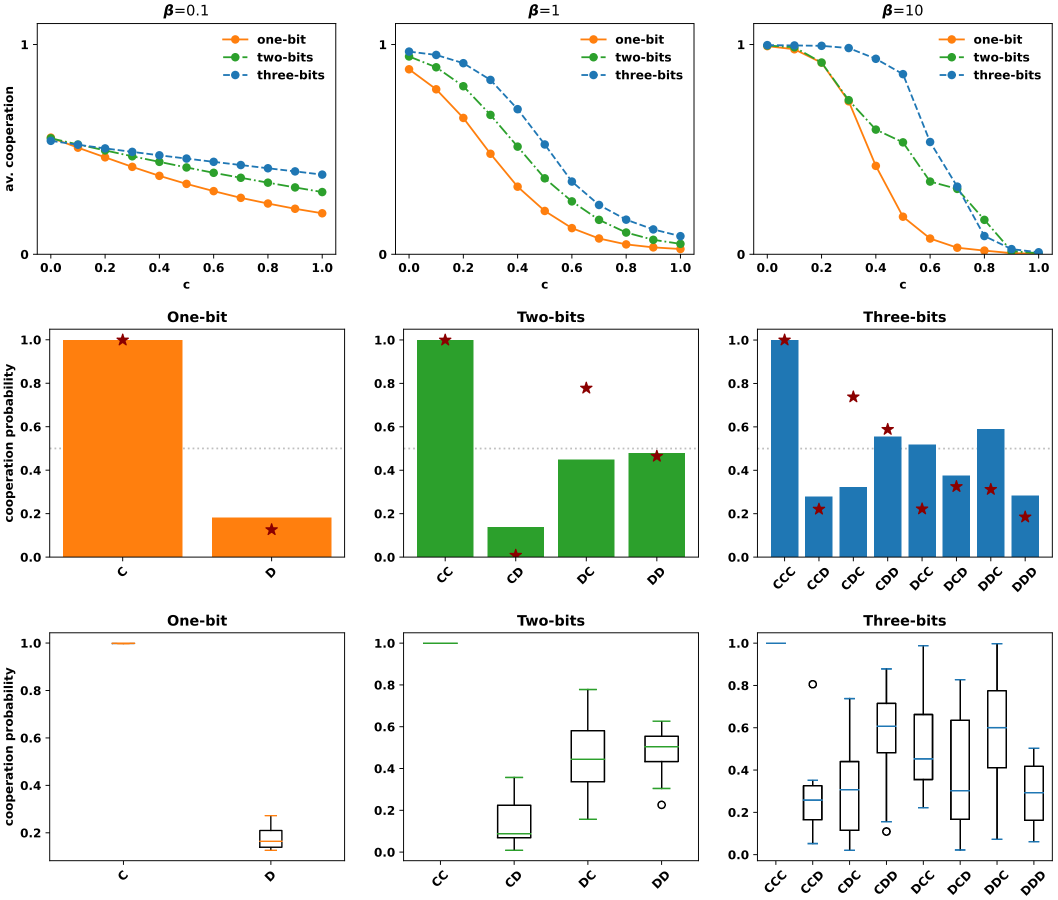

Figure 3: **Comparing the evolving cooperation rates and the most abundant strategies one-bit, two-bits, and three-bits strategies. A.** the evolving cooperation rates for different selection strengths. To assess the impact of memory on the evolution of cooperation, we ran simulations based on Imhof 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 $2 \cdot 10^5$.

- 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

$$b \cdot x_{\mathbf{p},\mathbf{q}} \geq c \cdot y_{\mathbf{q},\mathbf{p}} \tag{23}$$

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 (23) 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:

$$\begin{align} Q_1(p) &= \{q \mid x_{\mathbf{p},\mathbf{q}} > 0 \text{ and } y_{\mathbf{q},\mathbf{p}} > 0\}, \tag{24}\\ Q_2(p) &= \{q \mid x_{\mathbf{p},\mathbf{q}} < 0 \text{ and } y_{\mathbf{q},\mathbf{p}} \leq 0\}, \tag{25}\\ Q_3(p) &= \{q \mid x_{\mathbf{p},\mathbf{q}} \leq 0 \text{ and } y_{\mathbf{q},\mathbf{p}} > 0\}, \tag{26}\\ Q_4(p) &= \{q \mid x_{\mathbf{p},\mathbf{q}} = 0 \text{ and } y_{\mathbf{q},\mathbf{p}} = 0\}, \tag{27}\\ & \tag{28} \end{align}$$

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

$$\max\{\frac{y_{\mathbf{q},\mathbf{p}}}{x_{\mathbf{p},\mathbf{q}}} | q \in Q_1(p)\} \leq b/c \leq \min\{\frac{y_{\mathbf{q},\mathbf{p}}}{x_{\mathbf{p},\mathbf{q}}} | q \in Q_2(p)\}. \tag{29}$$

$\mathbf{p}$ is a strict Nash equilibrium if the inequalities in (29) are strict, $Q_3(p) = \emptyset$ and $Q_4(p) = \emptyset$.

B Anti Press and Dyson

The work of [Press and Dyson, 2012], which is the work that introduced zero determinant set of strategies, proved a further interesting result which states the following:

Theorem B.1 (Iterated Prisoner's Dilemma contains strategies that dominate any evolutionary opponent)**.** Let p play a short memory strategy, and q play with longer memory of the past outcomes. In the perspective of the forgetful strategy p, p's score is exactly the same as if q had played a certain shorter-memory strategy.

Thus, for a given memory-one strategy that interacts with a memory-two strategy, the forgetful strategy will always be able to find a memory-one representation for the memory-two co-player, such that its score remains the same.

Note that in sections 3.2- 3.3 when we explore whether a random two-bit reactive strategy is nash we check against all pure memory-two strategies, not pure two-bit reactive strategies. In fact, it does not suffice to check only pure two-bit reactive strategies. That is because a $n-$bit reactive strategy can not guarantee that it will find a reactive representation for a memory-n strategy. We refer to this as the Anti Press and Dyson result.

References

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

R. Axelrod and W. D. Hamilton. The evolution of cooperation. *science*, 211(4489):1390–1396, 1981.

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.

J. E. Beasley. An algorithm for set covering problem. *European Journal of Operational Research*, 31(1): 85–93, 1987.

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.

A. McAvoy and M. A. Nowak. Reactive learning strategies for iterated games. *Proceedings of the Royal Society A*, 475(2223):20180819, 2019.

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.