

# Evolution of cooperation among individuals with limited payoff memory

Christian Hilbe, Nikoleta E. Glynatsi, Alex McAvoy

## Abstract

## 1 Introduction

One of the most important applications of evolutionary game theory is the evolution of cooperation. Why is it that some individuals choose to help others, increasing their payoff, at the expense of decreasing one's own payoff? In evolutionary game theory individuals are not required to be rational, instead they adapt strategies based on mutation and exploration. Strategies are more likely to spread if they have a high fitness either because the individuals who adopt them have more offsprings, or because they are imitated more often. The fitness of a strategy is not constant but depends on the composition of the population. Individuals interact based on their strategies with other members of the population and the payoffs they yield are translated into fitness.

It is commonly assumed that fitness is equivalent to the mean payoff an individual achieves over the different types in the population, over multiple interactions. These payoffs assume that individuals can interact with the entire population several times and remember each and every outcome. Thus, they imply that individuals have a perfect memory. However, in their pairwise interactions when they make decisions in each turn they are assumed to have very limited memory. To be precise most of the works in the literature, with a few notable exceptions [2, 3], focus on naive subjects who can only choose from a restricted set of strategies [4], or who do not remember anything beyond the outcome of the very last round [5].

The perfect memory assumption is not only unrealistic but it also creates this curious inconsistency. This has lead us to question how robust is our understanding of cooperation. In this work we explore whether direct reciprocity can evolve if individuals only remember a minimum of social information. Though we are not the first to question the assumptions of estimating fitness [6], we are the first to explore the effect of payoff memory.

Initially, we consider two extreme scenarios. The first is the classical scenario of the expected payoffs and the alternative scenario where individuals update their strategies only based on the very last payoff they

obtained. In the later sections, we allow individuals more memory. More specifically, they can remember up to two turns and up to two interactions. We present results on several social dilemmas which include the prisoner’s dilemma and the donation game, the snowdrift game, the stag hunt game and the harmony game.

The remainder of the paper is organized as follows. Section 2 describes the model. Section 3 presents the results of the simulations. Finally, section 4 outlines the main conclusions.

## 2 Model Setup

We consider a population of  $N$  players ( $N$  is even) where mutations are sufficiently rare. Thus, at any point in time there are at most two different strategies present in the population; a *resident* strategy and a *mutant* strategy. We assume a pairwise process where strategies spread because they are imitated more often.

Each step of the evolutionary process consists of two stages, a game stage and an updating stage. In the game stage each individual is randomly matched with some other individual in the population to interact for a number of turns, where subsequent turns occur with a fixed probability  $\delta$ . At each turn they can choose to either cooperate ( $C$ ) or to defect ( $D$ ), and thus, at each turn the possible outcomes are  $CC$ ,  $CD$ ,  $DC$  and  $DD$ . The payoffs depend on the outcome. If both cooperate they receive the reward payoff  $R$ , whereas if both defect they receive the punishment payoff  $P$ . If one cooperates but the other defects, the defector receives the temptation to defect,  $T$ , whereas the cooperator receives the sucker’s payoff,  $S$ . We denote the payoffs of an individual as  $\mathcal{U} = (R, S, T, P)$ .

We assume herein that individuals use *reactive strategies* to make decisions in each turn. Reactive strategies are a set of memory-one strategies that only take into account the previous action of the opponent. They can be written explicitly as a vector  $\in \mathbb{R}_3$ , more specifically, a reactive strategy  $s$  is given by  $s = (y, p, q)$  where  $y$  is the probability that the strategy opens with a cooperation and  $p, q$  are the probabilities that the strategy cooperates given that the opponent cooperated and defected equivalently.

In the updating stage, two players are randomly drawn from the population, a ‘learner’ and a ‘role model’. Given that the learner’s payoff  $u_L \in \mathcal{U}$  and that the role model’s payoff  $u_{RM} \in \mathcal{U}$ , we assume the learner adopts the role model’s strategy based on the Fermi distribution function,

$$\rho(u_L, u_{RM}) = \frac{1}{1 + \exp^{-\beta(u_{RM} - u_L)}}. \quad (1)$$

where  $\beta \geq 0$  is the relative influence of the payoffs on adopting the strategy of the other. We refer to  $\beta$  as the intensity of selection.

This basic evolutionary step is repeated until either the mutant strategy goes extinct, or until it fixes in the population. If the mutant fixes in the population then the mutant strategy becomes the new resident strategy. After either outcome we introduce a new mutant strategy uniformly chosen from all reactive strategies at random, and we set the number of mutants to 1. This process of mutation and fixation/extinction is then

iterated many times.

The perfect memory assumption occurs at the updating stage. The learner and the role model are assumed to interact with a representative sample of the population, and they remember all interactions they participate in. Thus, their updating payoffs are based on the mean payoff they achieved over all the interactions. These payoffs are referred to as the expected payoffs. We will compare the expected payoffs to payoffs that are calculated when the role model and learner do not remember all their interactions. In order to account for the effect of these different methods we explore the cooperation rate within the resident population over multiple generations. More details on our methodology are found in Appendix A.

### 3 Results

#### 3.1 Updating payoffs based on the last round with another member of the population

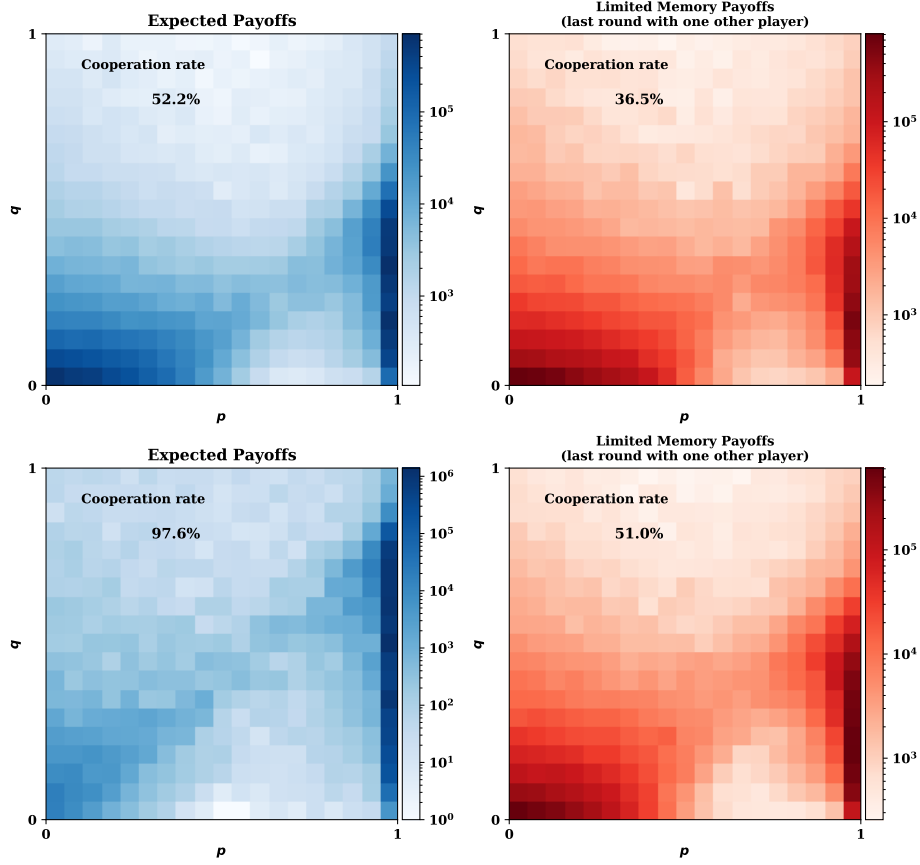
In this section we explore the case where the updating payoffs are based on the last round payoff achieved against another member of the population. We compare this to the expected payoffs. We assume that individuals interact in a donation game where each can cooperate by providing a benefit  $b$  to the other player at their cost  $c$ , with  $0 < c < b$ . Thus,  $T = b$ ,  $R = b - c$ ,  $S = -c$ ,  $P = 0$ .

Figure 1 shows simulation results for the described process of section 2. Figure 1 depicts the evolving conditional cooperation probabilities  $p$  and  $q$ . The discount factor  $\delta$  is comparably high, thus the opening move  $y$  is a transient effect and has no effect on the outcome. The left panels correspond to the standard scenario considered in the literature. It considers players who use expected payoffs to update their strategies. The right panel shows the scenario considered herein, in which players update their strategies based on their last round's payoff. The top panels assume a benefit  $b$  of 3, whereas the bottom panels assume a benefit of 10.

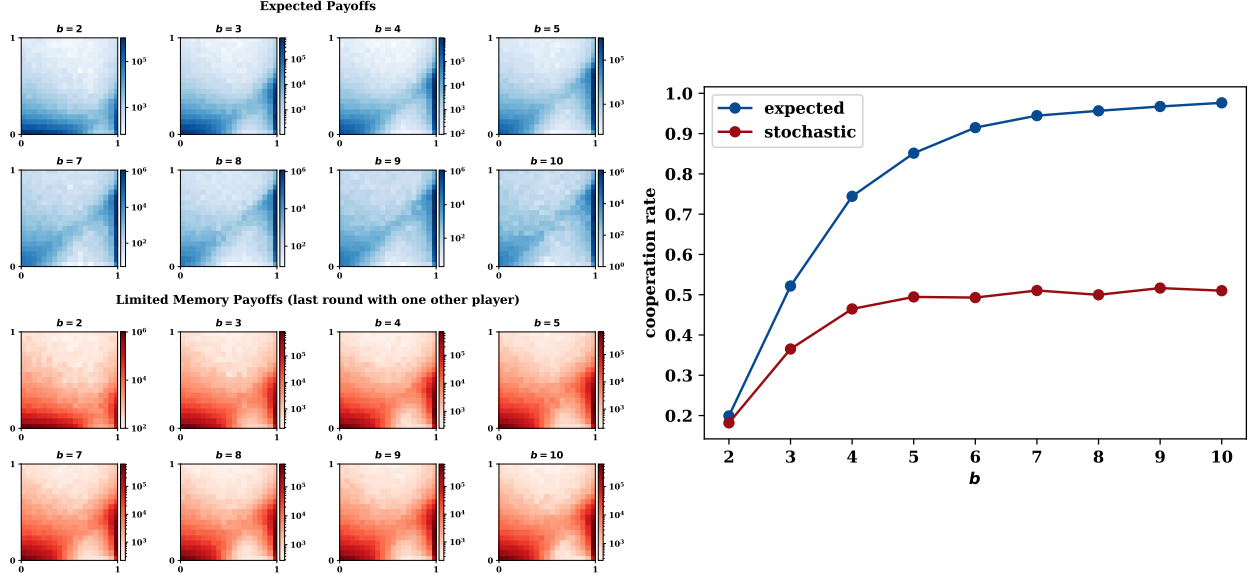
The figure suggests that when updating is based on expected payoffs, players tend to be more generous and more cooperative. The  $q$ -values are higher on average which suggests that individuals tend to be more generous when it comes to cooperating after being at the receiving end of a defection. The average cooperation rate within the resident population is strictly higher when expected payoffs are considered. The difference between the two methods for both the values of  $b$  are statistically significant. This contrast becomes more obvious for  $b = 10$ . More specifically the average cooperation drops from 97% to 51%. The residents of the population cooperated on average 97% of the times in one case and only 57% in the other.

We further explore the effect of the benefit, Figure 2. The figure suggests that the expected payoffs always overestimate cooperation. For the limited memory payoffs the cooperation rate remains unchanged at approximately 50% once  $b = 5$ . In specific simulations we have performed, if only the last round payoff is considered then the highest average cooperation rate that was achieved was 0.51. In comparison, when the expected payoffs were used the highest average cooperation rate was 0.97.

Figure 3 illustrates results for various runs of the evolutionary process where we vary the strength of

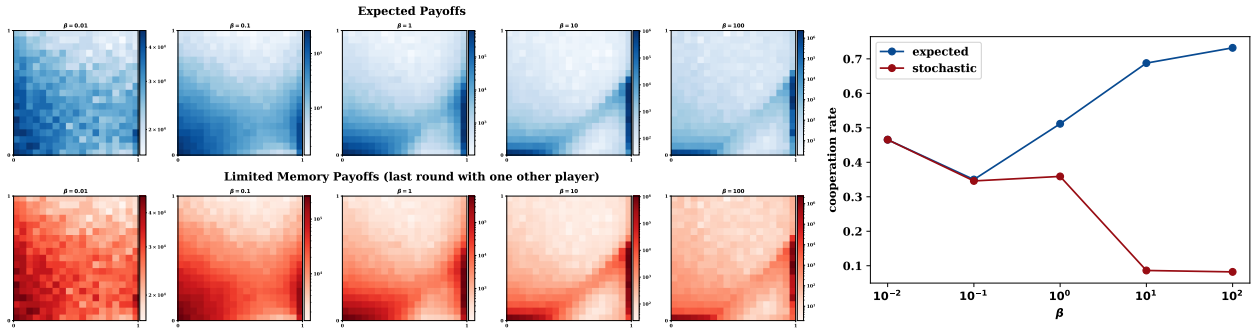


**Figure 1: Evolutionary dynamics under expected payoffs and last round with one interaction payoffs.** We have run two simulations of the evolutionary process described in section 2 for  $T = 10^7$  time steps. For each time step, we have recorded the current resident population  $(y, p, q)$ . Since simulations are run for a relatively high continuation probability of  $\delta = 0.999$ , we do not report the players' initial cooperation probability  $y$ . The graphs show how often the resident population chooses each combination  $(p, q)$  of conditional cooperation probabilities in the subsequent rounds. **(A)** If players update based on their expected payoffs, the resident population typically applies a strategy for which  $p \approx 1$  and  $q \leq 1 - c/b = 0.9$ . **(B)** When players update their strategies based on their realized payoffs in the last round, there are two different predominant behaviors. The resident population either consists of defectors (with  $p \approx q \approx 0$ ) or of conditional cooperators. In the latter case, the maximum level of  $q$  consistent with stable cooperation is somewhat smaller compared to the expected-payoff setting,  $q < 0.5$ . The cooperation rate within the resident population (averaged over all games and over all time steps) is close to 100%. Parameters:  $N = 100$ ,  $c = 1$ ,  $\beta = 1$ ,  $\delta = 0.999$ .



**Figure 2: The evolution of cooperation for different benefit values.** We vary the benefit of defection  $b$ . In all cases, expected payoffs appear to overestimate the average cooperation rate the population achieves. (A) the probabilities  $p, q$  for resident population over  $10^7$  time steps for each benefit value. (B) The cooperation rate within the resident population (averaged over all games and over all time steps) over the benefit. Unless explicitly varied, the parameters of the simulation are  $N = 100$ ,  $c = 1$ ,  $\beta = 1$ ,  $\delta = 0.99$ . Simulations are run for  $T = 5 \times 10^6$  time steps for each parameter combination.

selection. For weak selection,  $\beta < 1$ , the updating payoffs have no effect. The evolved population and the average cooperating rate are the same for both approaches. However for strong selection, it can be seen that the cooperating rate increases with  $\beta$ , when expected payoffs are used. However, when limited memory payoffs are used, the cooperating rate decreases.



**Figure 3: The evolution of cooperation for different selection strength values.** We vary the selection strength  $\beta$ . In all cases, stochastic payoff evaluation tends to reduce the evolving cooperation rates. (A) the probabilities  $p, q$  for resident population over  $10^7$  time steps for each  $\beta$  value. (B) The cooperation rate within the resident population (averaged over all games and over all time steps) over  $\beta$ . Unless explicitly varied, the parameters of the simulation are  $N = 100$ ,  $b = 3$ ,  $c = 1$ ,  $\beta = 1$ ,  $\delta = 0.99$ . Simulations are run for  $T = 5 \times 10^6$  time steps for each parameter combination.

### 3.2 Expected and stochastic payoffs in $2 \times 2$ games

## 4 Conclusions

### A Model Setup

We consider a pairwise comparison process. We assume a population of  $N$  individuals ( $N$  is even) where at any point in time there are at most two different strategies in present. There are  $k$  individuals who play the mutant strategy and  $N - k$  individuals who play the resident strategy. Each step of the evolutionary process consists of two stages; a game stage and an update stage.

In the game stage, each individual is randomly matched with some other individual in the population. Their interaction lasts for a number of turns which is not fixed but depends on the continuation probability  $\delta$ . At each turn the individuals have choose between cooperation (C) and defection (D). If both players cooperate they receive the reward payoff  $R$ , whereas if both players defect they receive the punishment payoff  $P$ . If one cooperates but the other defects, the defector receives the temptation to defect,  $T$ , whereas the cooperator receives the sucker's payoff,  $S$ .

In this work we consider different social dilemmas represented by the harmony game, the stag hunt game, the snowdrift or hawk-dove game, the prisoner's dilemma, and a special case of the prisoner's dilemma, the donation game (Table 1).

social dilemmas		payoffs' constrains
(i)	stage hunt	$R > T > P > S$
(ii)	snowdrift	$T > R > S > P$
(iii)	harmony	$R > T > S > P$
(iv)	prisoner dilemma	$T > R > P > S$
(v)	donation game	$T > R > P > S$ & $T = b, R = b - c, S = -c, P = 0$

**Table 1:** We present results on the all of the listed social dilemmas. The payoffs' constrains are different for each game.

Results for cases (i) - (iv) were presented in section 3.2 and results for case (v) were presented in section 3.1.

A further assumption of our model is that individuals make use of reactive strategies when they make decisions in each round. Reactive strategy are a set of strategies that take into account only the previous action of the opponent. A reactive strategy can be written explicitly as a vector,

$$s = (y, p, q)$$

where  $y$  is the probability that the strategy opens with a cooperation and  $p, q$  are the probabilities that the

strategy cooperates given that the opponent cooperated and defected equivalently.

In the updating stage, two players are randomly drawn from the population, a ‘learner’ and a ‘role model’. The learner adopts the role model’s strategy based on the Fermi distribution function,

$$\rho(u_L, u_{RM}) = \frac{1}{1 + \exp^{-\beta(u_{RM} - u_L)}}. \quad (2)$$

where  $u_L \in \mathcal{U}$  is the learner’s payoff,  $u_{RM} \in \mathcal{U}$  is the role model’s payoff, and  $\beta \geq 0$  is the intensity of selection.

We iterate this basic evolutionary step until either the mutant strategy goes extinct, or until it fixes in the population and becomes the new resident strategy. After either outcome, we set  $k$  to 1 and we introduce a new mutant strategy which is uniformly chosen from all reactive strategies at random.

This process of mutation and fixation/extinction is iterated many times. Instead of simulating each step of the evolutionary process, we estimate the probability that a newly introduced mutant fixes [7]. This is defined as the fixation probability of the mutant, and the standard form is the following,

$$\varphi = \frac{1}{1 + \sum_{i=1}^{N-1} \prod_k \frac{\lambda_k^-}{\lambda_k^+}}, \quad (3)$$

where  $\lambda_k^-$ ,  $\lambda_k^+$  are the probabilities that the number of mutants decreases and increases respectively.

This evolutionary process is summarized by Algorithm 1.

---

**Algorithm 1:** Pairwise comparison process

---

```

 $N \leftarrow$  population size;
 $k \leftarrow 1$ ;
resident  $\leftarrow (0, 0, 0)$ ;
while  $step < \text{maximum number of steps}$  do
    mutant  $\leftarrow$  random:  $\{\emptyset\} \rightarrow R^3$ ;
    fixation probability  $\leftarrow \varphi$ ;
    if  $\varphi > \text{random: } i \rightarrow [0, 1]$  then
        | resident  $\leftarrow$  mutant;
    end
end

```

---

The aim of this work is to explore the effect of updating memory on the cooperation rate of the evolved population. For this reason we consider two different approaches when estimating the payoffs at the updating stage. The two approaches we consider are those of (i) the expected payoffs and (ii) the stochastic payoffs.

## Expected Payoffs

The expected payoffs are the conventional payoffs used in the updating stage [8]. They are defined as the mean payoff an individual in a well-mixed population that engages in repeated games with all other population members.

We first define the payoff of two reactive strategies at the game stage. Assume two reactive strategies  $s_1 = (y_1, p_1, q_1)$  and  $s_2 = (y_2, p_2, q_2)$ . It is not necessary to simulate the play move by move. Instead the play between the two strategies is defined a Markov matrix  $M$ ,

$$M = \begin{bmatrix} p_1 p_2 & p_1 (1 - p_2) & p_2 (1 - p_1) & (1 - p_1) (1 - p_2) \\ p_2 q_1 & q_1 (1 - p_2) & p_2 (1 - q_1) & (1 - p_2) (1 - q_1) \\ p_1 q_2 & p_1 (1 - q_2) & q_2 (1 - p_1) & (1 - p_1) (1 - q_2) \\ q_1 q_2 & q_1 (1 - q_2) & q_2 (1 - q_1) & (1 - q_1) (1 - q_2) \end{bmatrix}. \quad (4)$$

whose stationary vector  $\mathbf{v}$ , combined with the payoff  $U$ , yields the expected outcome for each strategy,  $\langle \mathbf{v}(s_1, s_2), \mathbf{u} \rangle$ .

Given a population of  $N$  individual with  $k$  individuals playing the mutant strategy,  $s_M = (y_M, p_M, q_M)$ , and  $N - k$  playing the resident strategy,  $s_R = (y_R, p_R, q_R)$ , the expected payoffs of a resident  $\pi_R$  and of a mutant strategy  $\pi_M$  are defined as,

$$\begin{aligned} \pi_R &= \frac{N-k-1}{N-1} \cdot \langle \mathbf{v}(s_R, s_R), \mathbf{U} \rangle + \frac{k}{N-1} \cdot \langle \mathbf{v}(s_R, s_M), \mathbf{U} \rangle, \\ \pi_M &= \frac{N-k}{N-1} \cdot \langle \mathbf{v}(s_M, s_R), \mathbf{U} \rangle + \frac{k-1}{N-1} \cdot \langle \mathbf{v}(s_M, s_M), \mathbf{U} \rangle. \end{aligned} \quad (5)$$

Note that  $N-k-1$  is the number of other residents in the population, and  $k$  is the number of residents. The payoffs are averaged by the number of individuals in the population, given that interactions are not possible. The same applies for the payoff of a mutant.

Given the expected payoff, the probabilities that the number of mutants decreases and increases (Equation (3)), are given by:

$$\begin{aligned} \lambda_k^- &= \rho(\pi_R, \pi_M) \\ \lambda_k^+ &= \rho(\pi_M, \pi_R). \end{aligned}$$

## Stochastic Payoffs

We compare the expected payoffs with the finite memory payoffs which we refer to as the stochastic payoffs. Initially, we consider the case of where the payoffs correspond to the last round payoff an individual achieved



after interacting with one other individual.

The probability of being in any of the outcomes  $R, S, T, P$  in the last round is given by Equation (6) (Proposition 1).

**Proposition 1.** *Assume a play between the reactive strategies  $s_1$  and  $s_2$  with a continuation probability  $\delta$ . Then the probability of being in any of the four outcomes  $R, S, T, P$  are given by,*

$$\begin{aligned}
\mathbf{v}_R(S_1, S_2) &= (1-\delta) \frac{y_1 y_2}{1-\delta^2 r_1 r_2} + \delta \frac{\left( q_1 + r_1 ((1-\delta)y_2 + \delta q_2) \right) \left( q_2 + r_2 ((1-\delta)y_1 + \delta q_1) \right)}{(1-\delta r_1 r_2)(1-\delta^2 r_1 r_2)}, \\
\mathbf{v}_S(S_1, S_2) &= (1-\delta) \frac{y_1 \bar{y}_2}{1-\delta^2 r_1 r_2} + \delta \frac{\left( q_1 + r_1 ((1-\delta)y_2 + \delta q_2) \right) \left( \bar{q}_2 + \bar{r}_2 ((1-\delta)y_1 + \delta p_1) \right)}{(1-\delta r_1 r_2)(1-\delta^2 r_1 r_2)}, \\
\mathbf{v}_T(S_1, S_2) &= (1-\delta) \frac{\bar{y}_1 y_2}{1-\delta^2 r_1 r_2} + \delta \frac{\left( \bar{q}_1 + \bar{r}_1 ((1-\delta)y_2 + \delta p_2) \right) \left( q_2 + r_2 ((1-\delta)y_1 + \delta q_1) \right)}{(1-\delta r_1 r_2)(1-\delta^2 r_1 r_2)}, \\
\mathbf{v}_P(S_1, S_2) &= (1-\delta) \frac{\bar{y}_1 \bar{y}_2}{1-\delta^2 r_1 r_2} + \delta \frac{\left( \bar{q}_1 + \bar{r}_1 ((1-\delta)y_2 + \delta p_2) \right) \left( \bar{q}_2 + \bar{r}_2 ((1-\delta)y_1 + \delta p_1) \right)}{(1-\delta r_1 r_2)(1-\delta^2 r_1 r_2)}.
\end{aligned} \tag{6}$$

In these expressions, we have used the notation  $r_i := p_i - q_i$ ,  $\bar{y}_i = 1 - y_i$ ,  $\bar{q}_i := 1 - q_i$ , and  $\bar{r}_i := \bar{p}_i - \bar{q}_i = -r_i$  for  $i \in \{1, 2\}$ . Let  $\mathcal{U} = \{R, S, T, P\}$  denote the set of feasible payoffs in each round, and let  $\mathbf{u} = (R, S, T, P)$  be the corresponding payoff vector.

*Proof.* Given a play between two reactive strategies with continuation probability  $\delta$ . The probability that the game ends on round  $t$  is given by,

$$(1-\delta) \mathbf{v}_0 \sum \delta^t M^{(t)}. \tag{7}$$

Here,  $\mathbf{v}_0$  denotes the expected distribution of the four outcomes in the very first round, and  $1-\delta$  the probability that the game ends. It can be shown that,

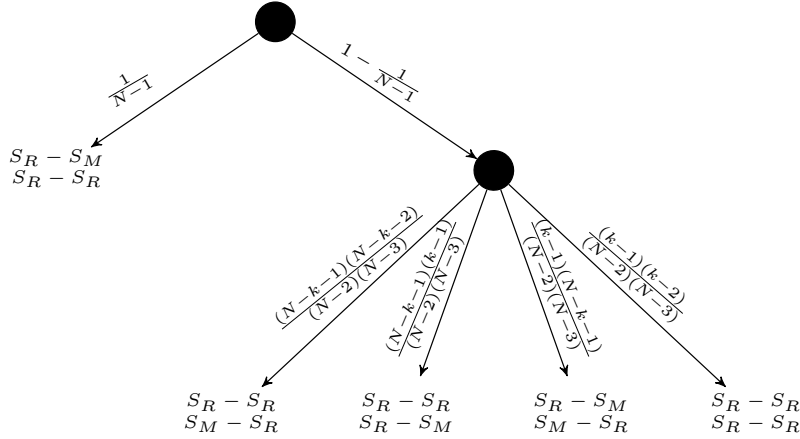
$$\begin{aligned}
(1-\delta) \mathbf{v}_0 \sum \delta^t M^{(t)} &= (1-\delta) (\mathbf{v}_0 + \delta \mathbf{v}_0 M + \delta^2 \mathbf{v}_0 M^2 + \dots) \\
&= (1-\delta) \mathbf{v}_0 (1 + \delta M + \delta^2 M^2 + \dots) \text{ using standard formula for geometric series} \\
&= (1-\delta) \mathbf{v}_0 (I_4 - \delta M)^{-1}
\end{aligned}$$

and with some algebraic manipulation we derive to Equation 6.  $\square$

Considering that individuals can only interact with one other individual, at that in each step a mutant and a resident are selected there are only five possible pairings:

- the selected resident is paired with the selected mutant
- the selected resident is paired with a resident and the selected mutant is paired with a resident
- the selected resident is paired with a resident and the selected mutant is paired with a mutant
- the selected resident is paired with a mutant and the selected mutant is paired with a resident
- the selected resident is paired with a mutant and the selected mutant is paired with a mutant

Each of the pairings happens with a given probability and that is illustrated by Figure 4.



**Figure 4:** The tree

The probability that the respective payoffs of the players are given by  $u_1$  and  $u_2$  can be calculated as

$$\begin{aligned}
 x(u_1, u_2) = & \frac{1}{N-1} \cdot v_{u_1}(S_1, S_2) \cdot 1_{(u_1, u_2) \in \mathcal{U}_F^2} \\
 & + \left(1 - \frac{1}{N-1}\right) \left[ \frac{k-1}{N-2} \frac{k-2}{N-3} v_{u_1}(S_1, S_2) v_{u_2}(S_2, S_2) + \frac{k-1}{N-2} \frac{N-k-1}{N-3} v_{u_1}(S_1, S_2) v_{u_2}(S_2, S_1) \right. \\
 & \quad \left. + \frac{N-k-1}{N-2} \frac{k-1}{N-3} v_{u_1}(S_1, S_1) v_{u_2}(S_2, S_2) + \frac{N-k-1}{N-2} \frac{N-k-2}{N-3} v_{u_1}(S_1, S_1) v_{u_2}(S_2, S_1) \right]. \quad (8)
 \end{aligned}$$

The first term on the right side corresponds to the case that the learner and the role model happened to be matched during the game stage, which happens with probability  $1/(N-1)$ . In that case, we note that only

those payoff pairs can occur that are feasible in a direct interaction,  $(u_1, u_2) \in \mathcal{U}_F^2 := \{(R, R), (S, T), (T, S), (P, P)\}$ , as represented by the respective indicator function. Otherwise, if the learner and the role model did not interact directly, we need to distinguish four different cases, depending on whether the learner was matched with a resident or a mutant, and depending on whether the role model was matched with a resident or a mutant.

Given that  $N - k$  players use the resident strategy  $S_1 = (y_1, p_1, q_1)$  and that the remaining  $k$  players use the mutant strategy  $S_2 = (y_2, p_2, q_2)$ , the probability that the number of mutants increases by one in one step of the evolutionary process can be written as

$$\lambda_k^+ = \frac{N-k}{N} \cdot \frac{k}{N} \cdot \sum_{u_1, u_2 \in \mathcal{U}} x(u_1, u_2) \cdot \rho(u_1, u_2), \quad (9)$$

$$\lambda_k^- = \frac{N-k}{N} \cdot \frac{k}{N} \cdot \sum_{u_1, u_2 \in \mathcal{U}} x(u_1, u_2) \cdot \rho(u_2, u_1). \quad (10)$$

In this expression,  $(N-k)/N$  is the probability that the randomly chosen learner is a resident, and  $k/N$  is the probability that the role model is a mutant. The sum corresponds to the total probability that the learner adopts the role model's strategy over all possible payoffs  $u_1$  and  $u_2$  that the two player may have received in their respective last rounds. We use  $x(u_1, u_2)$  to denote the probability that the randomly chosen resident obtained a payoff of  $u_1$  in the last round of his respective game, and that the mutant obtained a payoff of  $u_2$ .

This framework can be extended to consider the case of where the payoffs correspond to the last  $n$  rounds payoff an individual achieved after interacting with  $m$  other individuals. For the case  $n = 2$  the payoffs at the game stage are,

**Proposition 2.** *Assume a play between the reactive strategies  $s_1$  and  $s_2$  with a continuation probability  $\delta$ . Then the probability of being in any of the sixteen outcomes  $RR, RR, RR, RR, RR, RR, RR, RR, RR, RR, RR, RR, RR, RR, RR, RR$  on the last two rounds are given by,*

$$\mathbf{v}_{\mathbf{a}_1, \mathbf{a}_2} = (1 - \delta) m_{a_1, a_2} \delta^2 [\mathbf{v}_0 (I_4 - \delta M)^{-1}]_{a_1, a_2}, \quad \text{for } m_{a_1, a_2} \in M \text{ \& } a_1, a_2 \in \{R, S, T, P\} \quad (11)$$

Proposition 2 can be extended to the last  $n$  rounds.

**Proposition 3.** *Assume a play between the reactive strategies  $s_1$  and  $s_2$  with a continuation probability  $\delta$ . Then the probability of being in any of the sixteen outcomes  $RR, RR, RR, RR, RR, RR, RR, RR, RR, RR, RR, RR, RR, RR, RR, RR$  on the last two rounds are given by,*

$$\mathbf{v}_{\mathbf{a}_1, \mathbf{a}_2} = (1 - \delta) \prod m_{a_1, a_2} \delta^2 [\mathbf{v}_0 (I_4 - \delta M)^{-1}]_{a_1, a_2} \quad (12)$$

for  $m_{a_1, a_2} \in M$  and  $a_1, a_2 \in [1, 4]$ .

Equation 8 can also be extended to include interactions with two other individuals. The possible pairings are illustrated by Figure ??.

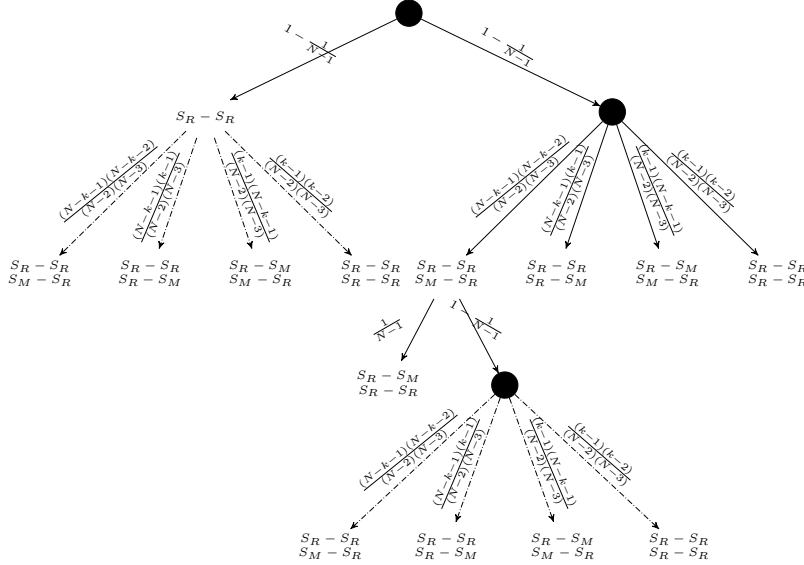


Figure 5: The tree

## B Verifying analytical results with simulations

The analytical results presented in this work have been verified with simulations. More specifically the probabilities of Equation (6),

Proposition 2,

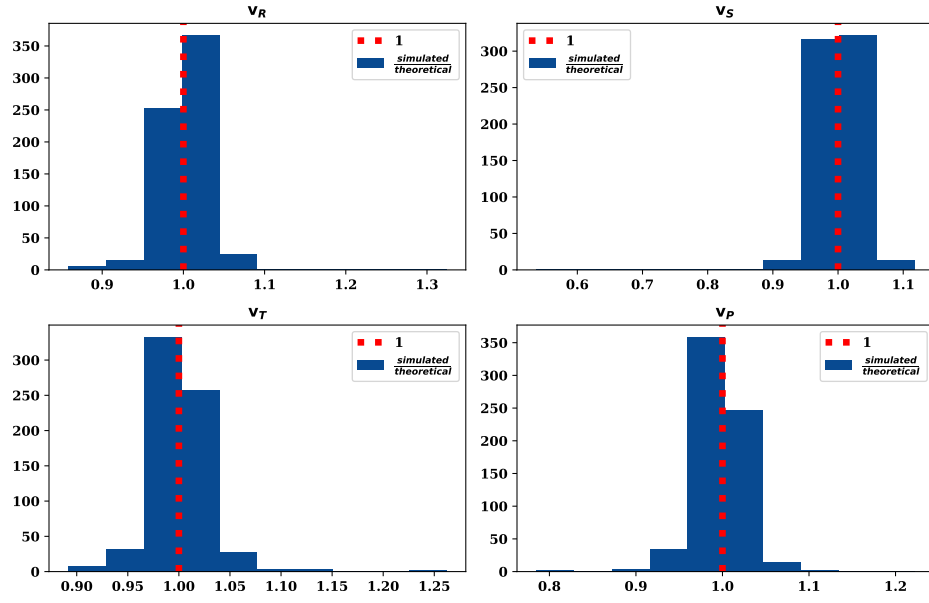


Figure 6

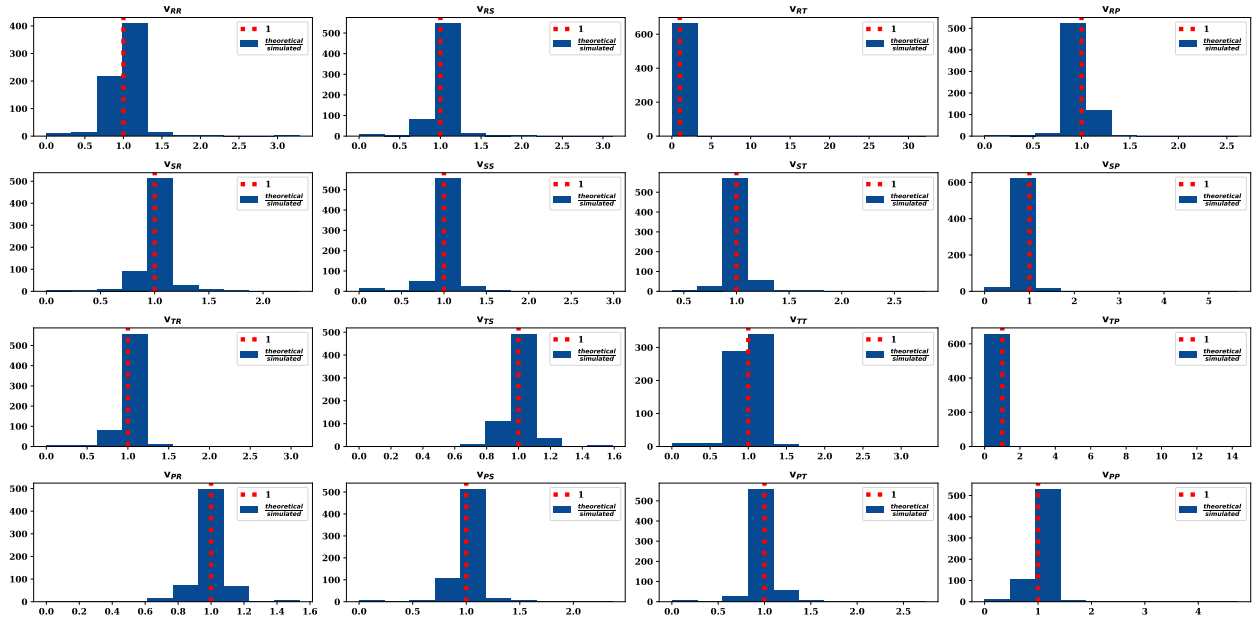


Figure 7

## References

- [1] Bin Wu, Benedikt Bauer, Tobias Galla, and Arne Traulsen. Fitness-based models and pairwise comparison models of evolutionary games are typically different—even in unstructured populations. *New Journal of Physics*, 17(2):023043, 2015.
- [2] Ch Hauert and Heinz Georg Schuster. Effects of increasing the number of players and memory size in the iterated prisoner’s dilemma: a numerical approach. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 264(1381):513–519, 1997.
- [3] Alexander J Stewart and Joshua B Plotkin. Small groups and long memories promote cooperation. *Scientific reports*, 6(1):1–11, 2016.
- [4] Martin A Nowak and Karl Sigmund. Tit for tat in heterogeneous populations. *Nature*, 355(6357):250–253, 1992.
- [5] Seung Ki Baek, Hyeong-Chai Jeong, Christian Hilbe, and Martin A Nowak. Comparing reactive and memory-one strategies of direct reciprocity. *Scientific reports*, 6(1):1–13, 2016.
- [6] Carlos P. Roca, José A. Cuesta, and Angel Sánchez. Time scales in evolutionary dynamics. *Phys. Rev. Lett.*, 97:158701, Oct 2006.
- [7] Martin A Nowak, Akira Sasaki, Christine Taylor, and Drew Fudenberg. Emergence of cooperation and evolutionary stability in finite populations. *Nature*, 428(6983):646–650, 2004.
- [8] Lorens A Imhof and Martin A Nowak. Stochastic evolutionary dynamics of direct reciprocity. *Proceedings of the Royal Society B: Biological Sciences*, 277(1680):463–468, 2010.