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? During the past decade, the literature on mechanisms that allow for cooperation, even though it's seen to be at odd has been fruitful. One such mechanism is repetition; the so called direct reciprocity. Theoretical researchers have been using evolutionary models to understand how direct reciprocity allows cooperation to evolve and which strategies are important for sustaining cooperation.

Evolutionary game theory does not require individuals to be rational, instead they adapt strategies based on mutation and exploration. Strategies with high fitness are more likely to spread, either because the individuals who adopt them have more offspring (fitness-based processes), or they are imitated more often (pairwise comparison processes) [1]. The fitness of strategies, and subsequently of the individuals, is not constant. Instead it depends on the composition of the population. Individuals interact with other members of the population according to their strategies. Their yielded payoffs, according to pre defined games, are translated into fitness.

It is commonly assumed that fitness assimilates to the *expected payoff*, which the mean payoff an individual achieved over the different types in the population after multiple intercalations. Expected payoffs imply that individuals have a perfect a memory. In order to estimate expected payoffs individuals must be able to recall several of their interactions with each player in the population. However, when modeling how they make decisions in each round they assumed to have very limited memory. To be precise, most work on this models focuses 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], with a few notable exceptions [2, 3].

This creates a curious inconsistency, which raises the following question: how robust is our understanding of cooperation? We are not the first to question the assumptions of models when estimating fitness. In [6] relax the assumption that selection occurs much more slowly than the interaction between individuals. They results shows that rapid selection affects evolutionary dynamics in such a dramatic way that for some games it even changes the stability of equilibria.

We consider two cases. Initially, we start by considering 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. We refer to these as the *stochastic* payoffs of individuals. We do this for the donation games but also the important cases of symmetric 2×2 games. These include the prisoner's dilemma, the snowdrift game, the stag hunt game and the harmony game.

In the later sections we allow individuals to use more memory. More specifically, individuals update their strategies by considering up to the last two rounds, whilst interacting with up to two different players.

2 Model Setup

In order to account for the difference in the robustness of cooperation among individuals based on their payoff memory, we consider a population of N players, where N is even, and where mutations are sufficiently rare. At any point in time the there are at most two different strategies are present in the population. A *resident* strategy and a *mutant* strategy. Suppose there are $N - k$ players who use the resident strategy whereas k players use the mutant strategy. Each step of the evolutionary process consists of two stages, a game stage and an updating stage.

In the game stage, each player is randomly matched with some other player in the population to interact in a number of turns. The number of turns is not fixed, on the contrast we consider that another interaction has a probability δ to continue following each turn. At each turn the players can choose to either cooperate (C) or to defect (D). The payoffs in a given round depend on both player's decisions and are given by:

$$U = \begin{pmatrix} R & S \\ T & P \end{pmatrix}$$

We assume herein that individuals at most make use of simple *reactive strategies* make decisions in each round. Reactive strategy are a set of memory-one strategies that only take into account the previous action of the opponent. Reactive strategies 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 $u_{RL} \in \mathcal{U}$, we assume the learner adopts the role model's strategy based on the Fermi distribution function. The relative influence of the payoffs on the adopt the strategy of the other is controlled by an external parameter. The so called intensity of selection,

$\beta \geq 0$.

This basic evolutionary step is repeated until either the mutant strategy goes extinct, or until it fixes in the population, in which case 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 $k = 1$. This process of mutation and fixation/extinction is then iterated many times.

In this work we explore the effect on the payoffs in the updating stage. The details of the process described in this section can be found in appendix A.

3 Analysis of Stochastic Payoffs in the Donation game

We first explore the effect of stochastic payoffs on the cooperative behaviour by considering the extreme case that an individual's fitness is based on their last interaction with one other individual. We compare this to the expected payoffs. In this section we consider the donation game. Each player can cooperate by providing a benefit b to the other player at their cost c, with $0 < c < b$. Then, $T = b, R = b - c, S = -c, P = 0$, and matrix (2) is give by:

$$\begin{pmatrix} b - c & -c \\ b & 0 \end{pmatrix} \tag{1}$$

Figure 1 shows simulations results for the described process of section 2. Figure 1 depicts the evolving conditional cooperation probabilities p and q. Assuming that the discount factor δ is comparably high, the opening move y is a transient effect and has not effect on the outcome. The left panels corresponds to the standard scenario considered in the previous 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 of 3, whereas the bottom assume that the benefit is 10.

The figure suggests that when updating is based on expected payoffs, players tend to be more generous. The q-values are higher on average which suggests that individuals tend to cooperating more after they are at the receiving end of a defection. In addition, for both cases of b, individuals tend to be more cooperative. The average cooperation rate is strictly higher for expected payoffs. This difference is statistically important, abd the effect is even more obvious when the benefit is higher. More specifically, for $b = 10$ the average cooperation rate drops from 97% to 51%.

Figure 2 further explores the effect of the benefit on the cooperation rate. In all cases, the stochastic payoffs evolution tends to reduce the evolving cooperation rate. For the stochastic payoffs, benefit appears to not make a difference once $b > 5$, and on average the evolving cooperation rate is at 50%. The effect of benefit on the expected payoffs smooths out after $b > 6$, and the expected payoffs estimating the evolving rate to be at 90%.

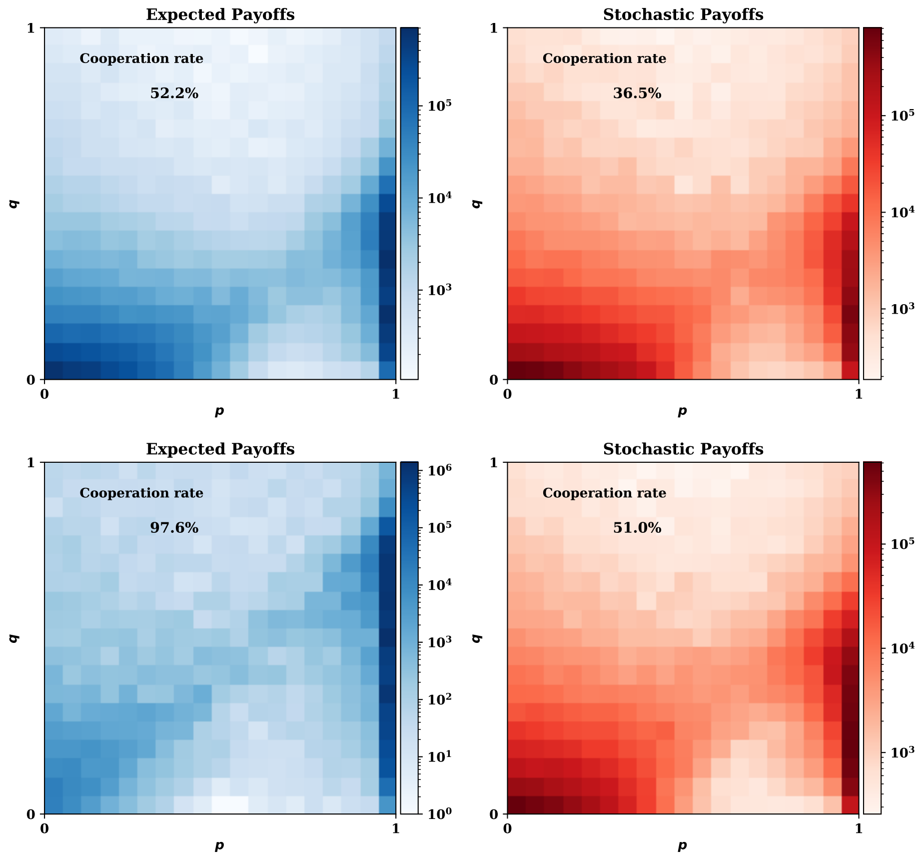

Figure 1: Evolutionary dynamics under expected payoffs and stochastic payoffs. We have run two simulations of the evolutionary process described in Section **??** for $T=10^7$ time steps. For each time step, we have recorded the current resident population (y,p,q). Since simulations are run for a relatively 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 $q\leq 1-c/b=0.9$. The cooperation rate within the resident population (averaged over all games and over all time steps) is close to 100%. (**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, $q<0.5$. Also the resulting cooperation rate is smaller. On average, players cooperate roughly in half of all rounds. Parameters: $N=100$, $b=3$, $c=1$, $\beta=1$, $\delta=0.999$.

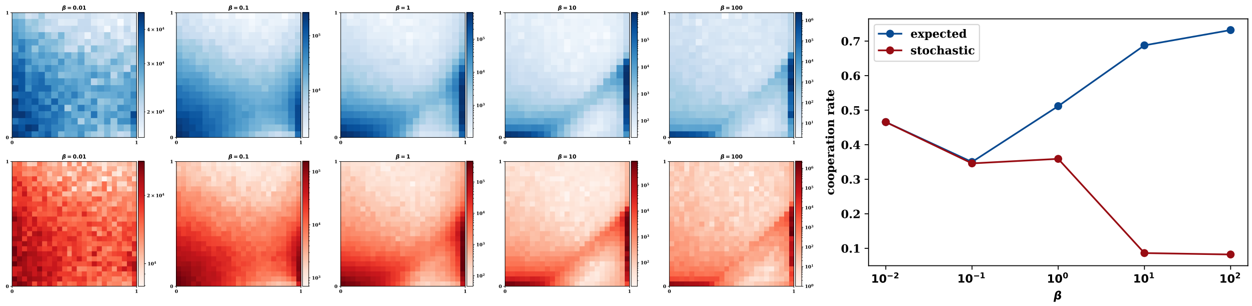

Figure 2: The evolution of cooperation for different benefit values. Here, we vary the benefit of defection b. In all cases, stochastic payoff evaluation tends to reduce the evolving cooperation rates. Unless explicitly varied, the parameters of the simulation are $N=100$, $b=3$, $c=1$, $\beta=1$, $\delta=0.99$. Simulations are run for $T=5\times10^6$ time steps for each parameter combination.

Figure 3 illustrates the results for different runs of the evolutionary process where we vary the strength of selection. In the case of a very small $\beta < \beta = 10^-1$ he process is almost random, as the effect of the payoffs is very small. Once payoff begin to matter the difference is once again evident, with the expected payoffs always overestimating the evolved cooperation rate.

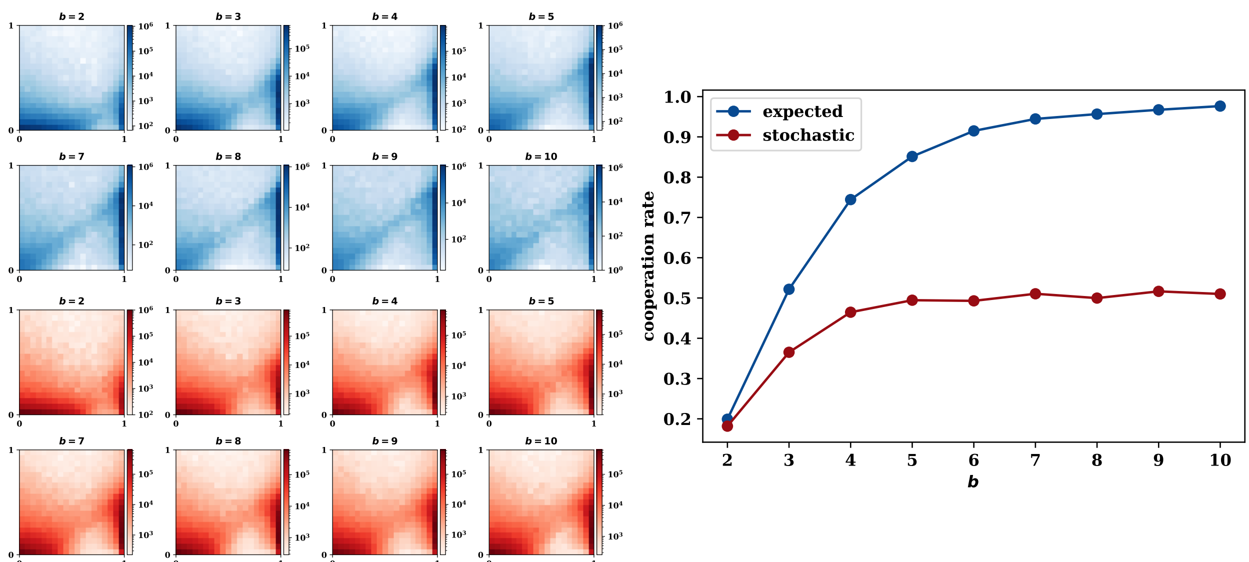

Figure 3: The evolution of cooperation for different selection strength values. Here, we vary the selection strength β. In all cases, stochastic payoff evaluation tends to reduce the evolving cooperation rates. Unless explicitly varied, the parameters of the simulation are $N = 100$, $b = 3$, $c = 1$, $\beta = 1$, $\delta = 0.99$. Simulations are run for $T = 5 \times 10^6$ time steps for each parameter combination.

4 Expected and stochastic payoffs in 2×2 games

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 δ. 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 4 and results for case (v) were presented in section 3.

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,

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

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,

$$\varphi = \frac{1}{1 + \sum\limits_{i=1}^{N-1} \prod\limits_{k}^{i} \frac{\lambda_k^-}{\lambda_k^+}}, \tag{3}$$

where λ_k^-, λ_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 ← population size;
k ← 1;
resident ← (0, 0, 0);
while step < maximum number of steps do
    mutant ← random: {∅} → R^3;
    fixation probability ← φ;
    if φ > random: i → [0, 1] then
        resident ← 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,

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

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 π_R and of a mutant strategy π_M are defined as,

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

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 δ.*

Then the probability of being in any of the four outcomes R, S, T, P are given by,

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

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 δ. 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. □

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.

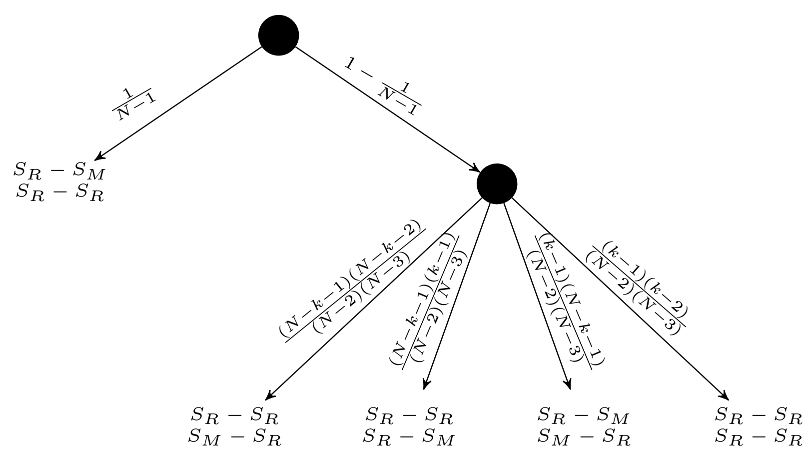

Figure 4: The tree

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

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

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

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

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

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 δ. 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,*

$$\mathbf{v_{a_1,a_2}} = (1-\delta)m_{a_1,a_2}\delta^2\left[\mathbf{v_0}(I_4-\delta M)^{-1}\right]_{a_1,a_2}, \quad \textit{for } m_{a_1,a_2}\in M \;\&\; a_1,a_2\in\{R,S,T,P\} \tag{11}$$

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 δ. 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,*

$$\mathbf{v_{a_1,a_2}} = (1-\delta)\prod m_{a_1,a_2}\delta^2\left[\mathbf{v_0}(I_4-\delta M)^{-1}\right]_{a_1,a_2} \tag{12}$$

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 **??**.

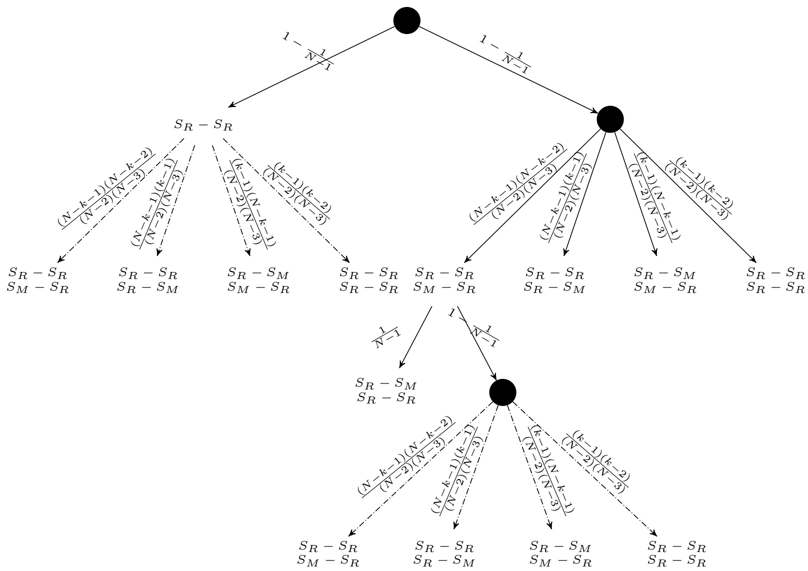

Figure 5: The tree

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),

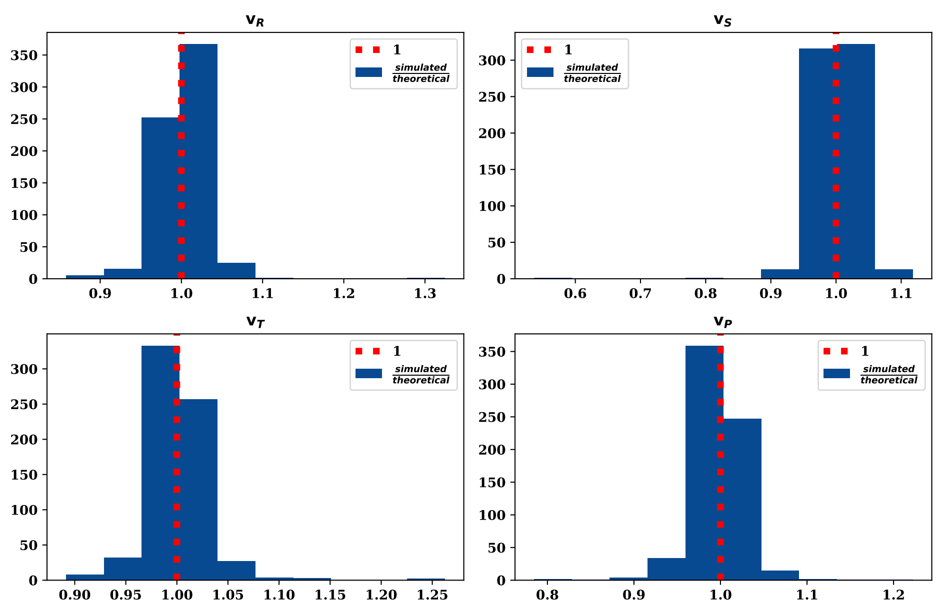

Figure 6

Proposition 2,

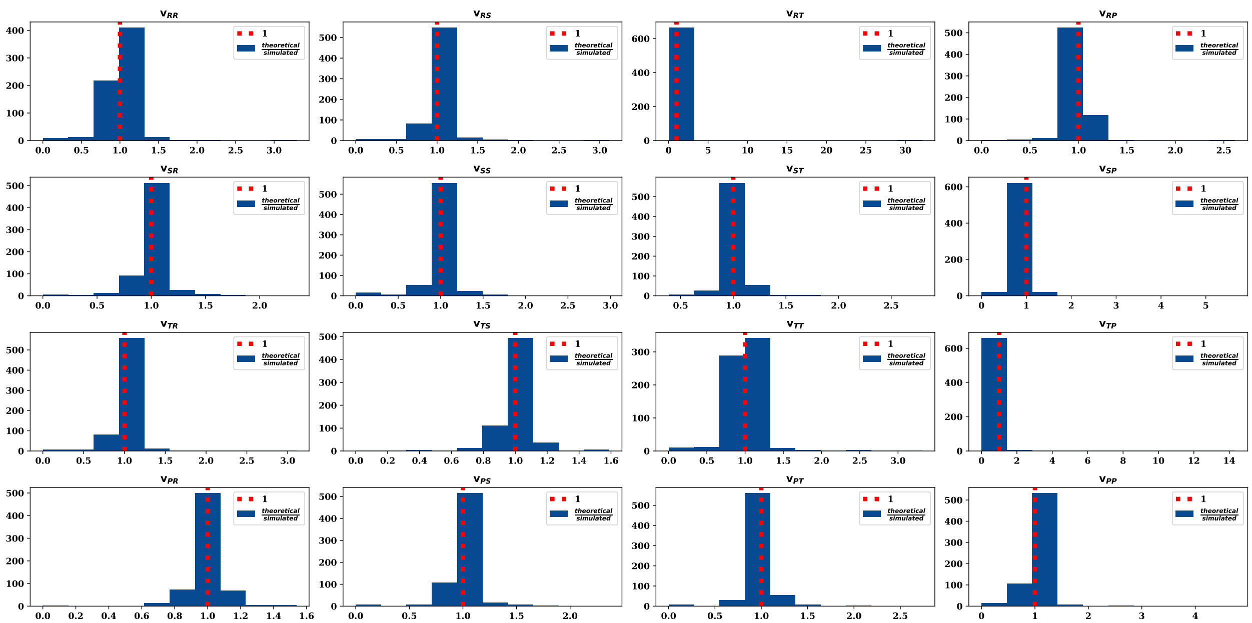
v_RR
v_RS
v_RT
v_RP
v_SR
v_SS
v_ST
v_SP
v_TR
v_TS
v_TT
v_TP
v_PR
v_PS
v_PT
v_PP
1
theoretical / simulated

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.